dc_1977_21

MTA doktori értekezés

Új dimenziók a genomszerkezet kutatásban

Dr. Székvölgyi Lóránt

Biokémiai és Molekuláris Biológiai Intézet

Általános Orvostudományi Kar

Debreceni Egyetem

2022

Tartalom

BEVEZETÉS

A természet máig megoldatlan talánya, hogy vajon hogyan fér el egy közel két méter hosszú makromolekula, az emberi örökítőanyag (DNS), egy körülbelül 5 μm átmérőjű sejtmag parányi dimenziójában? A DNS koncentrációja igen magas a sejtekben: ~100 mg/ml egy tipikus eukarióta sejtmagban, ~10 mg/ml egy átlagos baktériumban (Krebs et al. 2018); kémiai rendszerekben ez a töménység egy viszkózus gélt eredményezne, amely kizárná az élő sejtekre jellemző legtöbb sejtbiokémiai folyamatot. A DNS lánc mechanikai tulajdonságait tekintve rigid testként modellezhető, amely perzisztenciahossza (a polimerek hajlékonyságára jellemző fizikai paraméter) 0,1M-os NaCl oldatban ~50 nm (150 bp) (Manning 2006). Vagyis a DNS hosszának ($3x10^9$ bp) és perzisztenciahosszának (150 bp) ismeretében megbecsülhetjük, hogy az örökítőanyag hányféle térbeli konformációt képes spontán módon felvenni. Ez a szám az emberi genom esetén egy egyszerűsített kombinatorikai becslés alapján $n!/k!(n-k)!$ (ahol $n=3x10^9/150$, $k \leq n$) minimum 20 millió. A lehetséges térbeli szerkezetek magas száma nyilvánvalóan nincs összhangban a kromatinszerkezeti kutatások több alapvető megfigyelésével: például, azzal, hogy az eukarióta sejtek örökítő egységei, a kromoszómák, nem csupán kétszálú dezoxiribonukleinsav láncok, hanem DNS-ből, hiszton+nem-hiszton fehérjékből és ribonukleoproteinekből álló makromolekuláris komplexek, amelyek számos nyitott (vagyis rugalmas) és zárt (rigid) kromatin kompartmentet alakítanak ki az interfázisos sejtmagban. Előbbi régiókat klasszikusan eukromatinnak nevezzük, amely laza szerkezetű és elsősorban aktívan átíródó géneket tartalmaz, míg az utóbbi képlet neve heterokromatin, amely többnyire erősen kondenzált, génekben szegény, és funkcionálisan teljesen vagy részlegesen inaktív (konstitutív/fakultatív heterokromatin). Ezen felül, a kromatin szerkezeti alapegysége, a nukleoszóma, nagyszámú és kémiailag rendkívül változatos poszttranszlációs módosításokon megy keresztül, amelyek kombinatív módon, a különféle sejtbiokémiai funkcióknak megfelelően finomhangolják a nyitottabb és zártabb kromatin régiók szerkezetét és működését, tovább szűkítve a lehetséges feltekeredési útvonalak számát. További alapvető megfigyelés, hogy a kromoszómák egy sor diszkrét topológiai állapotot vesznek fel a sejtciklus során, amelyek tökéletesen reprodukálódnak minden egyes interfázis és mitózis után.

A kromatin emellett rendkívül dinamikus és adaptív, belső és külső ingerekre képes drasztikusan megváltoztatni térbeli konformációját. A genom flexibilitását növelő fontos tényező a DNS molekula egyik vagy mindkét szálát megszakító folytonossághiányok (SSB - szimplaszál törés; DSB - duplaszál törés) és speciális, ún. “nem-B DNS” struktúrák (pl. R-hurkok) jelenléte. (Ezek részletes bemutatása a következő fejezetekben történik.) DNS hibák, törések, és R-hurkok gyakorlatilag folyamatosan

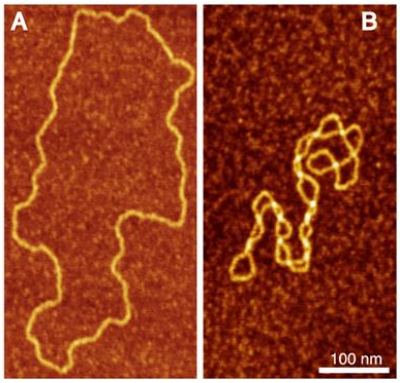
keletkeznek a sejtekben a különféle sejtélettani és/vagy patológiás folyamatok révén, ám azokat hatékony hibakorrekciós rendszerek és replikatív hibaelkerülő útvonalak javítják ki. Előbbiek közül a legismertebbek a báziskivágó, nukleotid kivágó, és hibás párok ("mismatch") hibajavítási útvonalak (amelyek felfedezését és biokémiai jellemzését kémiai Nobel-díjjal jutalmazták 2015-ben), illetve a DSB javító útvonal, amely nem-homológ végek összekapcsolása mechanizmussal (non-homologous end joining, NHEJ), vagy homológ rekombinációval (HR) történhet. A hibaelkerülő mechanizmusok három fő fajtája a templátváltás, a homológ rekombináció (ezek hibamentes, nem-mutagén útvonalak), illetve a transzléziós szintézis, amely erősen mutagén a gyakori hibás nukleotid beépítések miatt (McCulloch and Kunkel 2008). Megjegyzendő, hogy a DNS törésekre általában, mint valamilyen patológiás állapotra gondolunk (részben jogosan, hiszen azok növelik a genom instabilitását, amely pl. a rosszindulatú sejtek jellemzője), viszont az utóbbi évek kutatásai számos olyan fiziológiás folyamatot tártak fel (Niehrs and Luke 2020), amelyek genetikailag programozott kromoszómatörések vagy R-hurkok létrejötte nélkül nem játszódhatnának le (Santos-pereira and Aguilera 2015; Szekvolgyi and Nicolas 2010). A kromatinszerkezet és DNS hibajavítás zavarai azonban súlyos fejlődési rendellenességekhez és degeneratív betegségekhez társulhatnak (pl. rák, autoimmun és neurodegeneratív kórképek), amelyekben specifikus génhálózatok szabályozása sérül (García-Muse and Aguilera 2019a).

A genom e duális viselkedését (flexibilitás vs. stabilitás) az evolúció időskáláján szemlélve igen logikusnak tűnik, hogy ez a kettősség elengedhetetlenül szükséges az alapvető sejtes funkciók és a funkciókat működtető szerkezeti elemek megőrzéséhez (pl. kromatid kohézió, testvérkromatid szegregáció, DNS replikáció, transzkripció, rekombináció, hibajavítás). A genom szerkezetével kapcsolatban azonban számos megválaszolatlan (vagy részben feltárt) kérdéssel szembesülünk: mennyire specifikus a kromatin feltekeredése eltérő sejttípusokban; mekkora az egyedi sejtek közötti variabilitás; hogyan változik a 3D kromatin mintázat a szöveti differenciáció, az öregedés, és különféle patológiás állapotok során; van-e különbség a diploid testi sejtek és a haploid (meiotikus) ivarsejtek kromatinszerkezeti sajátosságai között; milyen összefüggés van a magasabbrendű kromatinszerkezet, a DNS léziók és R-hurok struktúrák kialakulása között; melyek a térbeli genomszerkezet főbb molekuláris komponensei, különös tekintettel a hiszton módosító enzimek, hiszton modifikációk és nem-kódoló RNS-ek szerepére? A következő pontokban ezekre a kérdésekre keressük a választ.

A GENOM FELÉPÍTÉSÉNEK SZERKEZETI ALAPVONÁSAI

Az örökítőanyag méretét elsősorban a DNS replikáció hibarátája és a replikációs origók száma korlátozza, így a genom maximális nagysága hibakorrekció (proofreading) nélküli replikáció esetén ~10^3-10^4 bp (pl. vírusokban), míg hibakorrekcióval rendelkező replikáció esetén ~10^6-10^{11} bp lehet (Smith and Eörs Szathmáry 1995). (A gerincesek között az Ausztrál tüdőshal (*Neoceratodus forsteri*) rendelkezik a legnagyobb genommal, amely 4.3x10^{10} bp nagyságú, 14-szer nagyobb az emberi genomnál (Meyer et al. 2021). A prokarióta DNS polimerázoknak van hibakorrekciós funkciójuk, de a genom mérete a ~10^4-10^6 bp-os tartományra korlátozódik (Sessions 2013); itt a genomméretet az limitálja, hogy csupán egy replikációs origóból indul a DNS kettőződés.

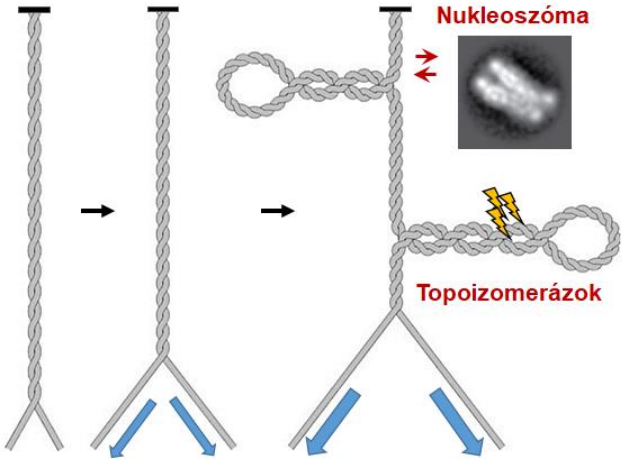
A virális genom igen szorosan pakolódik: a kapszid térfogata majdnem megegyezik a benne lévő nukleinsav térfogatával. Ezzel szemben a baktériumok genetikai állománya a sejttérfogat töredékét foglalja el. *E. coli* sejteket lizálva tömegesen izolálhatunk a széttöredezett sejtburokhoz kihorgonyzott, fehérjékkel-asszociált, hurokszerű DNS rostokat. Etídium-bromid (EtBr) titrálásos mérések alapján az izolált DNS rostok ún. zárt-cirkuláris negatív szuperhélix formában vannak, amely a DNS molekula természetes alapállapota (**1. ábra**).



1. ábra Atomerő mikroszkópos (AFM) felvétel pBR322 plazmidról (A) Relaxált állapot. (B) Szuperhelikális állapot. A DNS hélix ~10.5 bp-ként csavarodik meg a saját tengelye körül, amely mechanikai feszülés hatására (pl. EtBr interkaláció) megváltozik: alultekertből relaxált, majd túltekert állapotúvá válik. Az EtBr interkalálódik a DNS-be és megszünteti annak negatív szuperhelikális ("alultekert") állapotát (relaxáció), majd fokozatosan pozitív irányú ("túltekert") szuperhelikális állapotot idéz elő. A negatív szuperhelicitás mértéke az a EtBr koncentráció, amely a relaxált állapot eléréséhez szükséges. Forrás: (Witz and Stasiak 2009).

Az izolálás során véletlenszerű egyszáltörések (ún. nick-ek) keletkeznek a nukleoid DNS-ben, amelyek *elméletileg* megszüntetik az EtBr által létrehozott energetikailag kedvezőtlen pozitív szuperhelikális állapotot. Ám a valóságban nem ez történik: a nukleoid nem relaxálódik, hanem megtartja az EtBr okozta pozitív szupertekercs állapotát, ami azt jelenti, hogy a bakteriális genom számos egymástól független doménből áll. A genomban körülbelül 100 ilyen egység van, melyek mindegyike ~43 kb (13 μm) nagyságú autonóm hurkot alkot (Krebs et al. 2018). E topológiai domének egymástól függetlenek, mivel a szuperhelicitás változása nem propagálódik egyik hurokról a másikra; így eltérő szuperhélix állapot tartható fenn a baktérium genom különböző részein, amely alapvetően befolyásolja a különféle gének/géncsoportok koordinált expresszióját vagy csendesítését.

Ha eukarióta sejtekből sóextrakciós eljárással izolált genomon végezzük el a fenti EtBr titrálást, hasonló következtetésre jutunk: egy fehérje-mátrixhoz kötődő illetve abból kibomló kromatin hurkok tömege figyelhető meg (Berezney and Coffey 1974), melyben a negatív szuperhélix egységek relaxációjához ~85 kb-onként 1 nick szükséges (Calladine 2004). A "nettó" alultekert állapot energetikailag megkönnyíti a DNS szálak szétválását például a transzkripció iniciációja vagy a pre-replikációs komplexek összeszerelése során (Naughton et al. 2013). A fehérje-mátrixot az alkalmazott extrakciós eljárástól függően scaffold-nak, nukleoszkeletonnak, vagy "halo"-nak hívják, jellegzetes fehérje komponensei evolúciósan konzerváltak prokariótáktól az emberig (Sureka and Mishra 2021). A mátrix fehérjék közül a topoizomeráz II enzim kitüntetett szerepet játszik a DNS szuperhelicitásának szabályzásában és a transzkripció/replikáció/rekombináció során felgyülemlő torziós-feszüléssel járó többletenergia eloszlatásában (Pommier et al. 2016; Razin et al. 2014; Zhang et al. 2014; Gasser et al. 1986; Adolph et al. 1977). Fontos megjegyeznünk, hogy a transzkripció és DNS replikáció önmagában indukálja a szuperhelikális domének kialakulását, amelyek szabályozása - a topoizomerázokon kívül - a DNS-nukleoszóma interakciók kontrollált megváltoztatásával lehetséges (**2. ábra**). Nukleoszóma beépülés hatására a DNS feszített állapotba kerül, amelyet energetikailag a hiszton-DNS kölcsönhatások tartanak fenn. A nukleoszómák disszociációja viszont újrarendezi a kialakult szuperhelikális struktúrát, melynek eredményeképpen a DNS alultekertté válik.



2. ábra Topoizomeráz enzimek és nukleoszóma-DNS interakciók szabályozzák a DNS szuperhelicitását. A transzkripció és replikáció során két összecsavart kötél széthúzásához hasonlatos topológiai problémát kell megoldani (kék nyilak, alul). Ekkor a DNS mintegy önmaga körül „túltekeredve" felcsavarodik, amely önmagában indukálja a topológiai domének kialakulását. A DNS szuperhelicitását topoizomeráz enzimek képesek módosítani. A topoizomeráz I átmeneti egyszál-töréseket hoz létre, míg a topoizomeráz II tranziens duplaszáltöréseket generál, miközben a törésen átbújtat egy másik DNS szálat. A másik lehetőség a hiszton/DNS interakciók megváltoztatása, például nukleoszómák szabályozott eltávolításával vagy beépítésével. Az ábrán egy nukleoszóma cryo-EM képe látható oldalnézetből, 3.9Å felbontással. Forrás: (Chua et al. 2016, 2005).

A szuperhelicitás tehát érzékenyen reagál a nukleoszómális szerkezet megváltozására, és fordítva, a DNS hélix túltekert vagy alultekert állapota gátolja illetve serkenti a nukleoszómák összeszerelődését. A nukleoszómák DNS-el való kölcsönhatása így alapvetően befolyásolja a magasabbrendű kromatinszerkezetet és ezáltal a génexpresszió, replikáció, és rekombináció térbeli és időbeli mintázatát. A nukleoszómák biokémiai tulajdonságaival és az eukarióta génexpresszió logikájával emiatt részletesebben foglalkozunk.

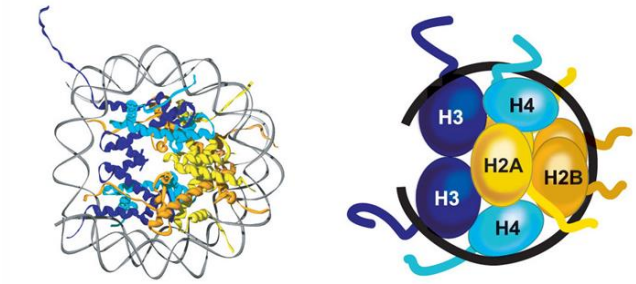
AZ EUKARIÓTA GÉNEXPRESSZIÓ LOGIKÁJA

A prokarióták és az eukarióták transzkripciós regulációjában a legfőbb különbség abból adódik, hogy az eukarióta DNS kromatinba van csomagolva. A prokarióta RNS polimerázok specifikus szekvenciáknál ismerik fel a promótereket: a transzkripció rátáját elvileg csak az korlátozza, hogy az RNS polimeráz milyen sebességgel hagyja el a promótert, hogy újabb iniciáció kezdődhessen. Valójában az aktivátorok és represszorok az RNS polimerázzal történő közvetlen kölcsönhatás útján (mintegy fizikai akadályt alkotva illetve megszüntetve) szabályozzák a prokarióta génexpressziót. A bakteriális promóterekhez könnyen hozzáférhet az RNS polimeráz, mivel nincs meg az fizikai korlát, amely az eukarióták esetén a kromoszómális szerveződés miatt alapból jelen van - a prokarióta génexpresszió alapállapota *nonrestriktív* (Udvardy 1999). Az eukarióták transzkripciós gépezete ezzel szemben egy riboszóma-méretű, két komponensből álló szerkezet: a TATA-boxhoz a TFIID fehérje kötődik (a bakteriális σ-faktorral analóg módon), míg a RNAPII holoenzimnek (amelyet az RNS polimeráz II, általános transzkripciós faktorok és egyéb fehérjék alkotnak) csak limitált szekvencia specificitása van - a TFIID fehérjén keresztül kötődik a promóterekhez. A transzkripcióhoz sokféle aktivátor és represszor kombinációjára, szinergisztikus kötődésére van szükség, s a kötőhelyek gyakran autonóm regulációs egységekbe, enhancer/silencer-ekbe tömörülnek, melyek magasan szervezett protein-DNS komplexeket alkotnak (pl: szuperenhanszerek, (Whyte et al. 2013; Bojcsuk et al. 2017). A promóterek tehát alapvetően inaktívak: az eukarióta génexpresszió alapállapota *restriktív* (Udvardy 1999). Ezt az állapotot maga a kromatin tartja fenn, mivel a nukleoszómák meggátolják a TFIID-TATA-box interakciót, viszont csak részlegesen korlátozzák az aktivátorok/represszorok kötődését a target szekvenciákhoz. E modulátor proteinek kétféle módon segítik az RNAPII és a promóterek asszociációját: közvetlen kölcsönhatásba lépnek az RNAPII holokomplexxel (a prokariótákkal analóg módon), vagy pedig közvetetten, a kromatinszerkezetet megváltoztató “remodelling” vagy hiszton módosító enzimekaktivitásokon keresztül toborozzák ada a transzkripciós apparátust. Az expresszálódó kromatin 100-1000x fogékonyabb bizonyos enzimatikus és kémiai hatásokra (pl. DNázI, MNáz, $KMnO_4$), mint a genom többi része, mivel itt a DNS nem a megszokott nukleoszómális szerkezetet veszi fel. E nukleáz szenzitív/hiperszenzitív helyek gyakran diszkrét pontokon, a gének szabályozó régióiban (promóterek, enhanszerek), replikációs origókban, rekombinációs forró pontokon és más, struktúrálisan fontos elemekben jelennek meg (Thurman et al. 2012). Ez nem fehérjementes szakaszokat jelent, hanem speciális transzkripciós regulátorok és kromatinátalakító (“remodelling”) komplexek szekvencia-specifikus kötődését, amelyek kiszorítják a nukleoszómákat. (A nukleáz szenzitív helyeket manapság Tn5 transzpozáz-alapú ATAC-seq módszerrel térképezik (Buenrostro

et al. 2013), míg a kromatinkötő fehérjéket kromatin immunprecipitáció szekvenálással (ChIP-seq) detektálják (Landt et al. 2012). A nukleáz érzékenység mellett ezen szakaszokra jellemző még a DNS alulmetiláltsága, a H1 linker hiszton hiánya, a H3/H4 hisztonok hiperacetilációja, és a H3 lizin 27-es oldallánc monometilációja. Az eukarióta génexpresszió elindítása tehát többlépcsős, egymástól különálló biokémiai részfolyamatok sorozata, amely a kromatin fellazulásával és a nukleoszómális szerkezet átrendeződésével kezdődik. De mi is az a nukleoszóma?

A NUKLEOSZÓMA

A nukleoszóma két-két kópia H2A, H2B, H3 és H4 hisztonból és ezen oktamer köré tekeredő 145-147 bp szuperhelikális DNS-ből áll (Olins and Olins 1974). Az oktamerben a hisztonok egy $(H3\text{-}H4)_2$ tetramert képeznek, amelyet H2A-H2B dimer szegélyez. Az ismétlődő nukleoszóma egységeket egy 10-90 bp-os DNS linker köti össze, amelyet a H1 hiszton stabilizál. A H1 hisztonnak fontos szerepe van a következő kromatin szerveződési szint, a 30 nm átmérőjű szolenoid kialakításában, amelyről a későbbiekben lesz szó. A nukleoszóma központi ("core") részét és a vele kölcsönható DNS-t 1997-ben röntgenkrisztallográfiás módszerrel karakterizálták (Luger et al.) (**3. ábra**). A tanulmány részletesen ismerteti a DNS kettős hélix kis árok és a hiszton-fold domén közötti kölcsönhatásokat, többek között elektrosztatikus kölcsönhatásokat, hidrogénkötéseket és apoláris kölcsönhatásokat.

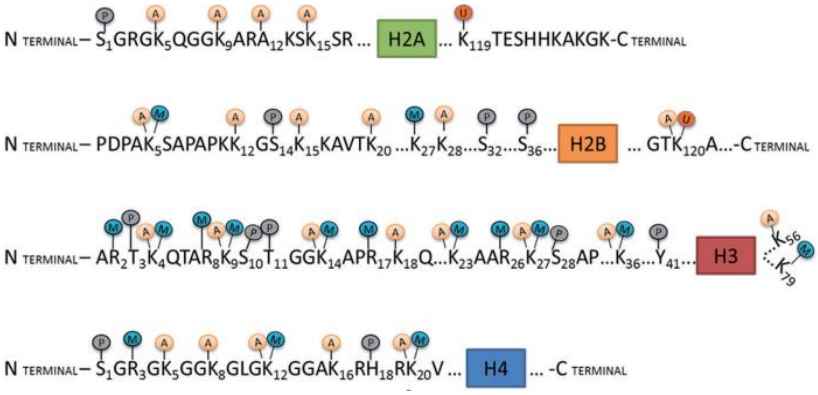


3. ábra A nukleoszóma. A nukleoszóma globuláris részét (hiszton oktamert) egy 146-147 bázispárnyi DNS veszi körül ami 1.65 fordulatnak felel meg. Egy nukleoszómában hozzávetőleg 14 kontakt pont van a DNS és a hisztonok között. E többszörös nem kovalens kölcsönhatás miatt, a DNS hiszton oktamer rendkívül stabil DNS-protein kölcsönhatást biztosít.

Proteomikai és genomikai tulajdonságaik alapján a nukleoszómális hisztonfehérjék két csoportra oszthatók: általános hisztonok (H2A, H2B, H3, H4) és hiszton-variánsok (H2A.X, H2A.Z, H3.3, CENP-A, stb.). Eukariótákban az általános hisztonokat kódoló gének esszenciálisak (deléciójuk letális) és tandem ismétlődő tömbökben csoportosulnak a genom különböző pontjain. Mivel számosságuk meghaladhatja a több ezres nagyságrendet, funkcionális vizsgálatuk rendkívül komplikált (ez alól kivételt képez néhány egyszerűbb eukarióta modellszervezet (pl. pékélesztő), amelyben speciális genetikai eljárásokkal megoldható a hiszton klaszterek kiütése és azok rekombináns hiszton génekkel történő helyettesítése (ld. később). Fontos megjegyezni, hogy az általános hisztonok transzkripciója funkcionálisan és mechanikailag kapcsolódik a DNS szintézishez, a variáns hisztonok viszont replikáció-független módon expresszálódnak a sejtciklus bármely fázisa

során. A hiszton variánsokat többnyire egyetlen vagy néhány génkópia kódolja s számos folyamatban játszanak szerepet, beleértve a DNS hibajavítást, homológ rekombinációt, kromoszóma szegregáció, és transzkripciót.

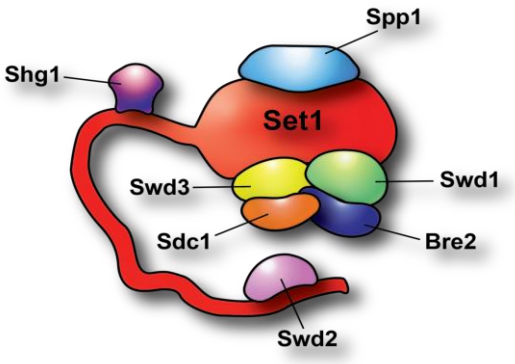
A nukleoszóma további jellemzői a flexibilis hisztonvégek, melyek áthaladnak a DNS szuperhélix felett és között. A nukleoszóma ezen N- és C-terminusai szabályozó szerepet töltenek be a különféle poszttranszlációs hiszton módosítások révén, ám mechanikai szerepük is van a kromatindinamika és a nukleoszóma-konformáció szabályozásában. A főbb hiszton módosítások közé tartozik a lizin acetiláció/metiláció/ubiquitináció/szumoiláció, arginin metiláció, és szerin foszforiláció (**4. ábra**), de ismertek egyéb, "különleges" módosítások is, például az arginin citrullináció (deimináció) (Zhu et al. 2021) vagy glutamin szerotoniláció (Farrelly et al. 2019). Ezen kémiai csoportok nem csupán a hisztonok túlnyúló N- és C-terminális oldalláncaira kerülhetnek, hanem a DNS-sel közvetlen kapcsolatban álló nukleoszóma felszínekre is, amelyre példa a DNS be- és kilépési pontjához (ún. "dyad axis"-hoz) közel eső H3 lizin 56 acetiláció (Yang et al. 2008).



4. ábra A nukleoszómális "core' hisztonok aminosav oldalláncainak főbb kémiai módosításai. Jelölések: A: acetiláció, M: metiláció, P: foszforiláció, U: ubiquitináció. (Az aminosavak standard egybetűs jelölését nem részletezzük.) Forrás: (Radzioch 2013).

A hiszton módosítások közül az egyik legabundánsabb és leginkább vizsgált aktiváló modifikáció a H3K4 trimetiláció (H3K4me3), amely tipikusan a gének 5'-végén helyezkedik el, és rendkívül konzervatív, élesztőktől az emberig az aktív génátírást stimuláló epigenetikai jel. *Saccharomyces cerevisiae*-ben a H3K4 metilációt egyetlen lizin-specifikus metiltranszferáz végzi, amelynek neve Set1C / COMPASS (Complex of Proteins Associated with Set1) (Miller et al. 2001; Dehe and Geli 2006). Gyümölcslégyben (*Drosophila melanogaster*) három, emberben hat Set1C létezik, melyet eredetileg MLL (Mixed Lineage Leukaemia) komplexeknek hívtak, utalva az akut limfoid és mieloid leukémiákban azonosított génfúziós termékekre (Takahashi et al. 2011). (Az MLL komplexek mutációit később más tumor típusokban is megtalálták, például az MLL2 és MLL3 mutációkat a gyermek medulloblasztómák 16%-ában (Weinstein et al. 2013).

Az Set1C / COMPASS önmagában katalitikusan inaktív; mono- di- és trimetiláz aktivitásaihoz további fehérjék kapcsolódása szükséges, amelyek élesztőben az Spp1, Bre2, Swd1, Swd2, Swd3, Sdc1 és Shg1 (**5. ábra**). A Swd1 és Swd3 elengedhetetlen a komplex stabilitásához és a Set1 katalitikus alegység működéséhez, ezek hiányában a Set1C destabilizálódik és a H3K4 metiláció nem megy végbe. Az Swd2 alegység a di- és trimetilációhoz szükséges (a monometilációhoz nem), amelynek viszont előfeltétele egy másik hiszton módosítás, a Rad6/Bre1 E3 ubiquitin ligáz által katalizált H2B lizin 123 mono-ubiquitináció (H2BK123ub) (Vitaliano-Prunier et al. 2008; Kim et al. 2009; Jeon et al. 2018). (Ez a mechanizmus izgalmas például szolgál a különféle hisztonmodifikációk és hisztonmódosító enzimek funkcionális interakciójára.) A Set1 és Swd2 alegységek N-terminális régiói kooperatív kölcsönhatást alakítanak ki az RNS polimeráz II C-terminális doménjével (CTD), amely a Set1C/COMPASS működését a transzkripció elongációjához kapcsolja (Bae et al. 2020). Az Swd2 ezenkívül elősegíti a transzkripció terminációjában résztvevő komplexhasítási és poliadenilációs faktor (CPF) működését, tovább erősítve a Set1C komponensek szerepét a génexpresszió szabályozásában (Cheng et al. 2004). A Bre2, Sdc1, és Spp1 alegységek a H3K4 trimetilációhoz szükségesek, lényegében ismeretlen mechanizmussal.



5. ábra Az élesztő Set1 hiszton metiláz komplex (Set1C / COMPASS) felépítése. A központi katalitikus egység az evolúciósan konzervált Set1 hisztonmetiláz domén, amelyhez 7 szabályozó alegységgel kapcsolódik. A komplex stabilitásához elengedhetetlen az Swd1 és Swd3 interakció, a többi alegység hiánya részleges funckióvesztést okoz. Az Spp1 egy H3K4me3 író-olvasó egység, amely PHD-finger doménjével a H3R2me2as módosításhoz is kötődik. A közelmúltban felfedezett interakciós partnere a Mer2 meiotikus rekombinációs fehérje, amelyről a későbbiekben lesz szó.

Az Spp1 a katalitikus (Set1) domén teljesen-aktív konformációjának kialakulásához szükséges; különlegessége, hogy H3K4 di/trimetil olvasó funkcióján kívül a H3 arginin 2 oldallánc aszimmetrikus dimetilációját (H3R2me2a) is felismeri és megköti a PHD-finger doménjén keresztül (Santos-Rosa et al. 2009; Nakanishi et al. 2008). Az Spp1-H3R2me2a interakció azonban kizárja a H3K4 metilációt (Guccione et al. 2007; Iberg et al. 2008), annak előbb le kell kerülnie a kromatinról a Set1C katalitikus aktivációjához. H3R2me2a-mentes nukleoszómákon az Spp1 stabil kölcsönhatásokat tud kialakítani a mono- di- és trimetilált H3K4 oldallánccal, amely így egyrészt védi a 4-es lizint a demetilázok hatásától, másrészt elfedi az R2 oldalláncot az arginin-metilázok elől. Nem meglepő, hogy a két hisztonmódosítás ellentétes genomi eloszlást mutat, ugyanis a H3R2me2a a heterokromatikus régiókra, míg a H3K4me az eukromatinra jellemző. Tovább árnyalja a képet, hogy a szimmetrikus arginin dimetiláció (H3R2me2s) viszont elősegíti a H3K4me olvasó egységek

bekötődését a di/trimetilált H3K4 hisztonokhoz (Migliori et al. 2012) és ezzel serkenti a transzkripció aktivációját (a szabályozás részletei azonban kevésbé ismertek). A Set1C egyéb funkcióiról és az Spp1 további interakciós partnereiről a meiotikus homológ rekombinációról szóló fejezetben lesz szó részletesebben.

A represszív hisztonmodifikációk közül kiemelendő a heterokromatin keletkezésben és epigenetikai továbbörökítésében kulcsszerepet játszó H3 lizin 27 metiláció (H3K27me), valamint a DNS metiláz rendszerrel szorosan együttműködő H3 lizin 9 metiláció (H3K9me) (Allshire and Madhani 2017). E módosítások részt vesznek a heterokromatin alapvető biológiai funkcióiban: i) a genom védelmében a kontrollálatlan transzpozonokkal és endogén retrovírusokkal szemben, ii) a pluripotencia és sejtdifferenciálódás szabályozásában a transzkripciós faktorok által közvetített kromatinszerkezeti változások és sejt-átprogramozás akadályozásával, iii) a mitotikus sejtosztódás szabályozásában a centromerikus kinetokor komplexek összeszerelődésén keresztül.

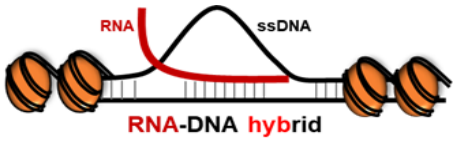
H3K9 mono- és dimetilált hisztonok (H3K9me1/2) emlősökben a SETDB1, Prdm3 és Prdm16 metilázok segítségével, ko-transzlációsan (riboszómákhoz kötődve) keletkeznek a citoplazmában, míg a H3K9 trimetilációt a SUV39H1/H2 lizin-metiltranszferázok katalizálják a sejtmagban (Gurard-levin et al. 2015). A H3K9me3 módosítást a heterokromatin fő szerkezeti fehérjéje, a HP1α/β ismeri fel specifikusan a chromodoménjei segítségével (Montavon et al. 2021), amely multimerizálódva és állványként funkcionálva további kölcsönhatásokat alakít ki transzkripciós represszorokkal, hiszton-deacetilázokkal, és a DNS metilációs rendszerrel (Torres and Fujimori 2015). Utóbbira példa a MECP2 és DNMT1 fehérjék HP1-el történő interakciója állati sejtekben (Smallwood et al. 2007; Agarwal et al. 2007), illetve a SUVH4/6 H3K9 trimetiláz és a metilált CHG/CHH szekvenciák specifikus interakciója növényekben (H=C, T, A). A SUVH4/6-mCHG/mCHH interakció pozitív visszacsatolási kört tart fenn a DNS metiláció és a hiszton H3K9 metiláció között, mivel a CMT2/3 DNS metilázok is specifikusan kötik a H3K9 trimetilált nukleoszómákat (Lindroth et al. 2001; Stroud et al. 2014). Ezek az egymást erősítő kölcsönhatások döntő szerepet játszanak a heterokromatinizáció elindításában és a génexpresszió maximális lekapcsolásában. Ezt a folyamatot „elhallgattatásnak” (vagy „silencing”-nek) nevezzük, amely egy öröklődő (tehát egymást követő sejtosztódási ciklusokon keresztül fennmaradó), illetve a kromatin mentén propagálódó represszált állapotot hoz létre, mely csak nagyon speciális körülmények között derepresszálódik.

Az elhallgattatás másik pillére a H3K27 metilációs mintázat kialakulása, amely magasabbrendű eukariótákban az ún. polycomb csoport (PcG) fehérjék aktivitásához köthető. A PcG gének mutációi a differenciálódás és egyedfejlődés súlyos zavaraihoz, homeotikus transzformációhoz és malignus sejtburjánzáshoz vezetnek (a polycomb gének emberi homológjai protoonkogének). A PcG fehérjék

funkcionálisan két külön molekuláris komplexet alkotnak (PRC1 és PRC2), amelyek eltérő, de funkcionálisan együttműködő hisztonmódosításokat katalizálnak (Blackledge and Klose 2021). A PRC1 E3 ubiquitin ligáz aktivitása révén a H2A lizin 119 oldallánc mono-ubikvitinációjáért felelős (H2AK119ub), míg a PRC2 a H3 lizin 27-et mono- és dimetilálja (H3K27me2/3). Emlősökben a PRC2-mediált H3K27me2/3 előfeltétele a H2AK119ub modifikációnak, növényekben viszont a PRC1 önmagában is kötődhet PcG target régiókhoz, amelyek H2A ubiquitinációja szükséges a PRC2 kötődéséhez és a represszív H3K27me2/3 módosítás kialakításához (Kralemann et al. 2020). A PRC2 enzimatikus aktivitásáért az evolúciósan konzervált EZH2/EZH1, SUZ12, EED egységek felelősek, azonban a komplex számos kiegészítő (opcionális) doménnel rendelkezik, amelyek finomhangolják a holoenzim funkcióját. Ugyanez érvényes a PRC1-re is, melynek RING1A/RING1B ubiquitin ligáz doménjét ún. RING-finger fehérjék (PCGF1-PCGF6) egészítik ki egységes katalitikus maggá, amelyhez különféle járulékos fehérjék kapcsolódnak (ezek egy része a PRC2-ben is megtalálható; (Blackledge and Klose 2021). Mindkét represszív komplex feladata, hogy egy stabil, zárt kromatinstruktúrát alakítson ki az enhanszer-prómoter szekvenciákon a szöveti differenciáció vagy az egyedfejlődés során. Fontos megjegyezni, hogy a PRC-kontrollált kromatin kompakció különféle intra- és interkromoszómális térbeli interakciók kialakulásával jár *Arabidopsis*-tól az emberig (Ngan et al. 2020; Alecki et al. 2020; Francis et al. 2004), amely a PRC1 esetén független a komplex katalitikus aktivitásától (Eskeland et al. 2010; Schoenfelder et al. 2015; Zhang et al. 2021). A PRC komplexekről a közelmúltban az is kiderült, hogy hatásuk számos gén esetén egy speciális háromszálú nukleinsav struktúrától függ állati és növényi sejtekben egyaránt (Skourti-Stathaki et al. 2019; Ariel et al. 2020), amelyet R-huroknak hívunk.

AZ R-HUROK

Az R-hurkok háromszálú nukleinsav struktúrák a genomban, amelyek egy RNS-DNS hibridből és egy kiszorított, szimplaszálú DNS-ből (ssDNS) állnak (**6. ábra**). 1976-os felfedezésük után (Thomas et al. 1976) *in vivo* jelenlétüket 1995-ben sikerült először kimutatni élő baktériumokban (Drolet et al. 1995). Ezt követően számos organizmusban tanulmányozták felépítésüket, illetve funkciójukat, és mára nyilvánvalóvá vált, hogy az R-hurkok az életfa minden szintjén előforduló epigenetikai elemek, élesztőtől az emberig (Sollier and Cimprich 2015). Míg korábban csupán transzkripciós melléktermékként tekintettek rájuk, az utóbbi évtizedben több fiziológiás, illetve patológiás folyamattal is kapcsolatba hozták őket (Niehrs and Luke 2020).



6. ábra Az R-hurok sematikus ábrázolása. Az R-hurkok háromszálú nukleinsav struktúrák, melyek egy DNS-RNS hibridből és egy kiszorított DNS szimplaszálból állnak.

Normál körülmények között a promóterek közelében lévő CpG szigetek metilációjának gátlásával vagy nem-kódoló RNS-eket toborozásával szabályozzák a génexpressziót (Arab et al. 2019; Ginno et al. 2012; Boque-Sastre et al. 2015; Sun et al. 2013), intermediert biztosítanak a T4 bakteriofág, az *Escherichia coli* plazmid, illetve a cirkuláris mitokondriális DNS replikációjához (Aguilera and García-Muse 2012) és a kromoszóma telomerek dinamikájához (Balk et al. 2013), továbbá B limfocitákban az R-hurkok ssDNS szála enzimatikus hatásoknak illetve mutációknak kitéve hozzájárul az immunglobulin nehézlánc izotípus váltásához (Yu et al. 2003). Patológiás körülmények között azonban az R-hurkok felhalmozódása súlyos veszélyt jelent a kromoszómák integritására, amely genom instabilitáshoz és rákos folyamatokat elindulásához vezethet (Crossley et al. 2019; Wells et al. 2019; Boque-Sastre et al. 2015). Az R-hurkokat különféle neurodegeneratív betegségekkel is összefüggésbe hozták, úgymint az Amiotrófiás Lateral Sclerosis (ALS), Aicardi-Goutieres szindróma, Friedreich ataxia, és a Fragilis X szindróma (Lim et al. 2015; Walker et al. 2017; Reddy et al. 2014; Kannan et al. 2019; Perego et al. 2018; Becherel et al. 2015; Noristani et al. 2015; Grunseich et al. 2018).

Az R-hurkok keletkezésének legelfogadottabb mechanizmusa a Lieber és Roy által javasolt modellek: a "visszafűződéses" és a "kiterjesztett hibrid" modell (Roy et al. 2010). A visszafűződéses modellben az egyszálú, naszcens RNS az RNS polimerázból kilépve azonnal visszafordul a transzkripciós buborékba és hibridizál a DNS templát szállal, kiszorítva annak kódoló szálát. A kiterjesztett hibrid modellben az RNS polimeráz elakadásával járó abortív transzkripció során alakulnak ki R-hurkok. A létrejött struktúrákat *cisz* R-hurkoknak nevezik, mivel közvetlenül társulnak a transzkripció folyamatához. (Általánosan elfogadott, hogy a legtöbb R-hurok transzkripció során jön létre.) A kotranszkripciósan kialakuló *cis* R-hurkok megjelenésének kedvez a DNS kódoló szálán lévő guanin gazdag régiók illetve DNS egyszáltörések jelenléte, az R-hurok elongációjához és stabilizálásához pedig a DNS szupertekercs szerkezete járul hozzá (Roy et al. 2010; Aguilera and García-Muse 2012). Az újabb kísérletes adatok arra utalnak, hogy az R-hurok RNS szála a genom más régiójából is származhat (Wahba et al. 2013; Ariel et al. 2020), ami lehet például hosszú nem kódoló RNS (lncRNS), körkörös RNS (circRNS) vagy ismétlődő RNS, melyek a saját transzkripciós helyüktől távoleső komplementer DNS szakaszokhoz hibridizálódnak, ún. *transz* helyzetben. Bizonyos esetekben mind a *cisz*-, mind a *transz* R-hurok jelen lehet egy adott régióban; ez az úgynevezett vegyes modell. A kialakult RNS-DNS párok termodinamikailag stabilabbak, mint a DNS-DNS duplexek (Roberts and Crothers 1992), amely valószínűleg abból adódik, hogy az RNS-DNS hibrid a duplaszálú RNS (dsRNS)

‘A’ formája és a DNS ‘B’ formája közötti intermediert képez (Shaw et al. 2008). Más adatok szerint a kiszorított ssDNS-en gyakran létrejövő G quadruplex (G4) szerkezet felelős a kialakult R-hurok stabilitásáért (Duquette et al. 2004). Más tanulmányok kimutatták, hogy a nyitott, transzkripcionálisan aktív kromatin régiók elősegítik az R-hurok képződését, illetve az R-hurkokkal rendelkező gének nagyobb mértékben expresszálódnak az R-hurok nélküli génekhez képest (Sanz et al. 2016). R-hurkok azonban represszív kromatin állapotokban is kialakulhatnak a genom kódoló és intergénikus régióiban (pl. ismétlődő elemeknél, telomerikus és pericentromerikus szakaszoknál (Nadel et al. 2015), amely részben a PRC kompexek aktivitásától függ.

Ezek a megfigyelések azt mutatják, hogy az R-hurkok kialakulása nem véletlenszerű folyamat, képződésük összetett kölcsönhatások eredménye a nukleotid-szekvencia, DNS-topológia, transzkripciós ráta, és különféle kromatin állapotok és hiszton módosító kompexek aktivítása között. E folyamatok mechanisztikus részletei még nem tisztázottak, azonban több olyan tényezőit ismerünk, amelyek segítik vagy gátolják az R-hurok kialakulását. Jelenlegi tudásunk szerint az alábbi mechanizmusok képesek szabályozni az R-hurkok mennyiségét a genomban:

(1) RNázH enzimek, amelyek szekvencia-független módon elhasítják az RNS-t az RNS-DNS hibridben (Lockhart et al. 2019; Skourti-Stathaki and Proudfoot 2014). A legtöbb organizmusban az enzim két típusa található meg, RNázH1 és RNázH2. Az egyetlen polipeptid láncból álló RNázH1 mind a sejtmagban, mind a mitokondriumban jelen van; utóbbi organellumban a DNS replikáció esszenciális eleme. Továbbá, a nukleáris RNázH1 kísérletes overexpresszió a R-hurok szint visszaszorításának leggyakrabban alkalmazott laboratóriumi módszere (Arudchandran et al. 2000). Az RNázH2 kevésbé karakterizált enzim, azonban a RNázH1-el ellentétben ez képes felismerni és kivágni a DNS-duplexbe hibásan beépülő ribonukleotidkat, akár egyetlen darabot is (Uehara et al. 2018). Míg az RNázH1 főként a transzkripció során kialakuló R-hurkok feloldásáért felelős, az RNázH2-t a DNS replikáció és hibajavítás során keletkező R-hurkok feldolgozásával hozták összefüggésbe (Chon et al. 2013).

(2) RNS-DNS hibrid helikázok, mint például a Senataxin (Skourti-Stathaki et al. 2011; Groh et al. 2017; Cohen et al. 2018), Aquarius (Sakasai et al. 2017; Sollier et al. 2014), Pif1 (Boulé and Zakian 2007; Tran et al. 2017), Srs1/BLM (Yun et al. 2017), UPF1 (Ngo et al. 2021), DHX9 (Chakraborty et al. 2018), DDX1 (Ribeiro de Almeida et al. 2018), DDX19 (Hodroj et al. 2017a, 2017b), DDX21 (Song et al. 2017), DHX30 (Hondele et al. 2019), amelyek részt vesznek az R-hurkok feloldásában, ezáltal pedig az R-hurok-mediált DNS károsodás és genomi instabilitás megelőzésében (Mischo et al. 2011). A Senataxin ezenkívül transzkripciós terminációs faktorként működik a kódoló és nem-kódoló géneknél egyaránt, továbbá részt vesz a transzkripció és replikáció koordinációjában a replikációs villa integritásának fenntartásán keresztül (Alzu et al. 2012).

(3) A topoizomerázok képesek megváltoztatni a DNS supercoiling-ot, amely egyébiránt folyamatos R-hurok keletkezéshez vezetne a transzkripció és replikáció során (Tuduri et al. 2009). A Topoizomeráz I (Top1) szerepéről tudunk többet, annak DNS relaxációs aktivitása hozzájárul a kotranszkripciós R-hurkok eltávolításához, amely a transzkripció leállását ("pausing") és a DNS repliszómával történő "frontális" ütközést okozná (Manzo et al. 2018; Cristini et al. 2019). A transzkripciós és replikációs apparátus közötti konfliktus a genom instabilitás egyik fő forrása, amely végsősoron R-hurok képződésre vezethető vissza.

(4) RNS processzáló és biogenézis faktorok, például az ASF/SF2 (Li and Manley 2005) vagy a THO/TREX (TRanscription and EXport) komplex (Domínguez-Sánchez et al. 2011), amelyek az mRNS érési/szállítási folyamatokban vesznek részt. Mindkét fehérje a frissen szintetizált RNS-t "becsomagolja", hiányukban vagy hibás működésükből adódóan a naszcens RNS visszahibridizál a DNS-duplexhez, R-hurok kialakulásához és transzkripciós-replikációs összeütközéshez vezetve (Huertas and Aguilera 2003).

(5) Rekombinázok és anti-rekombinázok, amelyek elősegítik vagy gátolják az R-hurkok keletkezését. A homológ rekombináció során a száláthelyezést katalizáló Rad51 rekombináz (és bakteriális homológja, a RecA) például segíti az R-hurok képződését *in vitro* és *in vivo* (Kasahara et al. 2000; Wahba et al. 2013). Kimutatták továbbá, hogy a Rad51 hiánya élesztőben csökkenti az R-hurkok felhalmozódását és a genomi instabilitás valószínűségét (Wahba et al. 2013). Az antirekombináz Srs2 viszont a Rad51-el ellentétes hatást fejt ki, vagyis akadályozza az R-hurkok kialakulását (Wahba et al. 2013; Yun et al. 2017). Ezenkívül az Srs2/BLM és UPF1 helikázok (Yun et al. 2017; Ngo et al. 2021), a Rad52 (Yasuhara et al. 2018), BRCA1 (Cristiano et al. 2019; Bhatia et al. 2014; Renaudin et al. 2021), BRCA2 (Hatchi et al. 2015, 2021; Zhang et al. 2017), FANCM (Liang et al. 2019; Schwab et al. 2015) rekombinációs modulátor fehérjék szintén dokumentált szerepet játszanak az R-hurkok felismerésében és a duplaszál DNS törések (DSB-k) javításában.

Az R-hurok szerkezetek keletkezésének és turnover-ének szabályozása tehát szorosan összefügg a genom integritását befolyásoló tényezőkkel, mint például a mutagenezis és homológ rekombináció. Előbbinek az alapja, hogy az R-hurok kiszorított ssDNS szála érzékeny a különféle DNS károsító (klasztogén) hatásokra, mint például a spontán vagy citidin dezamináz (APOBEC) enzimek által indukált dC → dU dezamináció, amely kettősszálú DNS törésekhez és rekombinációhoz vezet (Aguilera and García-Muse 2012; García-Muse and Aguilera 2019b). Megfigyelések szerint a humán aktiváció-indukálta citidin dezamináz (AID/APOBEC1) a B-limfociták genomjában található S (switch) régiókon kialakuló R-hurkok ssDNS-ét dezaminálja, hozzájárulva ezzel a hipermutációhoz és a immunglobulin nehézlánc izotípus váltásához (CSR: class-switch recombination) (Yu et al. 2003).

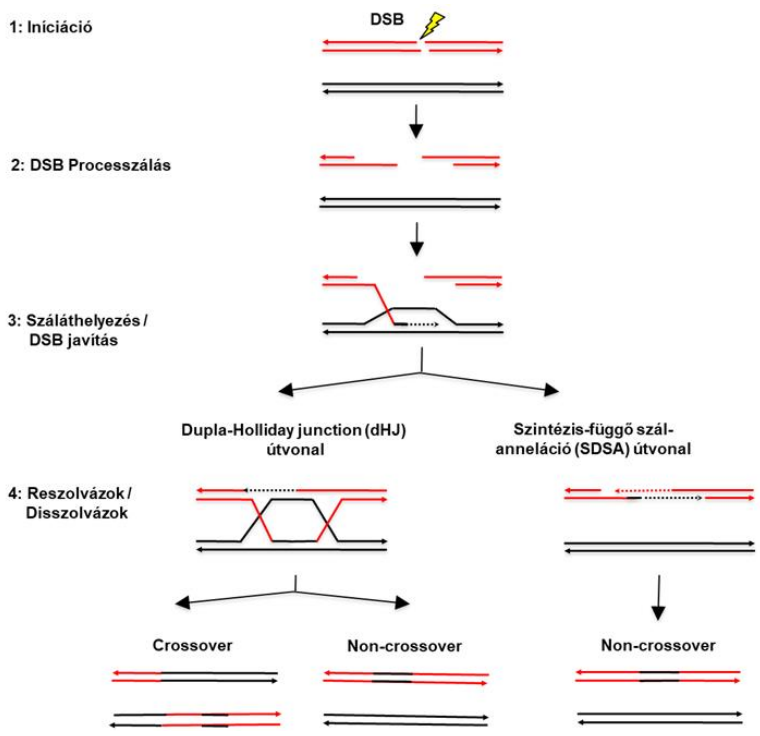
A létrehozott U:G mismatch ezekután replikálódik, a két leányszál közül az egyiken tehát egy C → T tranzíciós mutáció alakult ki. A dU azonban célpontja az uracil-DNS glikoziláznak illetve az abázikus endonukleáznak, amelyek eltávolítják az uracilt és egyszálú DNS törést (nick-et), vagy abázikus helyet hoznak létre (Sollier and Cimprich 2015). A kialakult abázikus helyre a repair DNS szintézis hibás nukleotidot építhet be, illetve a nick-ek DSB-vé konvertálódhatnak, amelyek rekombinációhoz és transzlokációkhoz vezethetnek (Skourti-Stathaki and Proudfoot 2014).

Az R-hurkok rekombinogén hatására további bizonyíték, hogy elősegítik a szálathelyezés folyamatát a DSB hibajavítás során. A DSB-k processzálását végző MRX komplex (Mre11, Rad50, Xrn2/Nbs1) kölcsönhatása az RNS polimerázzal egy R-loop szintézisét eredményezi a visszavágott DNS-vég közvetlen közelében (Ohle et al. 2016; Yasuhara et al. 2018), amely szükséges és elégséges feltétele a nem-mutagén homológia-vezérelt hibajavításnak (HDD), szemben a mutagén nem homológ vége kapcsolása (NHEJ) útvonallal. A folyamathoz Rad52 szükséges (Yasuhara et al. 2018). A Rad51 rekombináz szintén segíti az R-hurok kialakulását (Wahba et al. 2013), amely azért váratlan, mivel közismert, hogy a Rad51 kifejeződése megvédi a sejteket a külső és belső DNS károsító hatásoktól (pl. ionizáló sugárzás, a kemoterápiás szerek, spontán DNS törések), illetve a tumorsejtekben tapasztalt magas expresszió a DNS hibajavításban betöltött szerepét tükrözi; más szóval, a Rad51-ről feltételezték, hogy tumorszuppresszorként működik (Sugawara et al. 2003; Klein 2008). Koshland és munkatársai eredményei szerint viszont a Rad51 a DNS-károsodás kialakulása előtt (illetve anélkül is) jelen van az R-hurok régiókban (Wahba et al. 2013), tehát a Rad51 ezekben az esetekben onkogenikus lehet (azaz rákot okozhat) (Sue Mei Tan-Wong and Proudfoot 2013). Hétköznapi hasonlattal élve körülbelül olyan ez, mintha egy autószerelő nem megjavítaná, hanem tovább rongálná az elromlott gépjárművünket. A DNS hibajavító fehérjék R-hurkokkal történő interakciója és a kölcsönhatások specificitása intenzív kutatások tárgya, annak részletei lényegében ismeretlenek. Az R-hurkokkal összefüggésben (vagy azoktól függetlenül) keletkező DNS duplaszál törésekkel induló homológ rekombinációs útvonalakkal és azok kromatin-kontextus függő (epigenetikai) szabályozásával a következő fejezetekben foglalkozunk.

A HOMOLÓG REKOMBINÁCIÓ MECHANIZMUSA

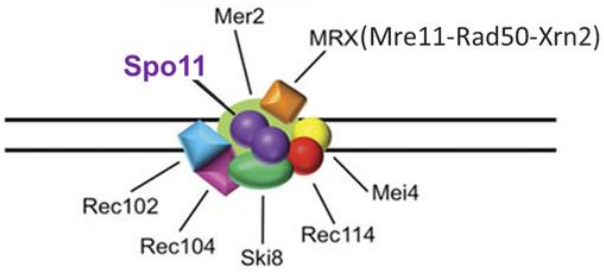
Az evolúció során a különféle DNS sérülésekre különböző hibajavító útvonalak specializálódtak, amelyek felismerik/átalakítják a léziókat és helyreállítják a genom integritását. A sejtek túlélését leginkább veszélyeztető folytonossághiányok a DNS duplaszál törések (DSB-k), amelyek szekvencia-

homológia alapján történő (hibamentes) korrekcióját a homológ rekombináció (HR) útvonala végzi (**7. ábra**).



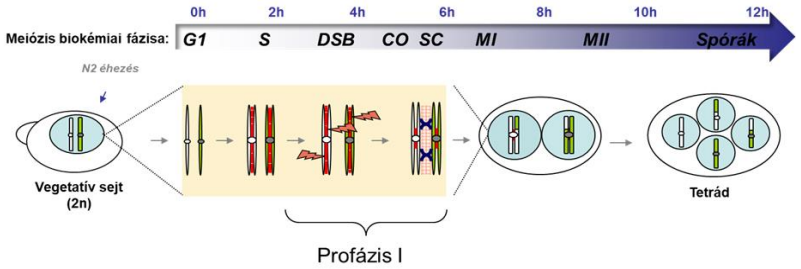
7. ábra A homológ rekombináció folyamata. 1: Iniciáció, azaz DSB-k keletkezése. 2: DSB processzálás. 5'-3' szálvisszavágás (MRX kompex), egyszálú 3'OH végű DNS keletkezése. A preszinaptikus filamentum egyike részt vesz a homológia keresésében, ún. single-end invasion (SEI) és D-hurok intermedierek keletkeznek. 3: Száláthelyezés és DSB javítás. DNS polimerázok meghosszabbítják a DSB véget, amely összekapcsolódik a második DSB véggel (second end capture) ún. dupla Holliday junction-t eredményezve, vagy visszahibridizál az eredeti szálhoz (SDSA útvonal). A dHJ útvonalon a különféle resolváz/disszolváz enzimek kétféle rekombináns molekulát hoznak létre, amelyek közül a szinte mindig reciprok crossover termék keletkezik. Non-crossover rekombináns az SDSA útvonalon keletkezik, amelyhez csak DNS polimeráz és ligáz aktivitások szükségesek. Forrás: (Székvölgyi et al. 2015a)

A hibajavítás alapját képező nem-sérült templát mitotikusan osztódó sejtekben az S fázisban szintetizálódó testvérkromatidokról származik, míg meiotikus sejtekben (ősivarsejtekben) kizárólag az apai vagy anyai homológ kromoszómákról. Ez egy döntő különbség a mitotikus és meiotikus HR között, a jelenséget interhomológ torzításnak ("interhomologue bias") nevezzük (Schwacha and Kleckner 1997). A másik döntő különbség a DNS törések keletkezési módjából és gyakoriságából adódik. Mitotikus DSB-k főleg a replikáció vagy transzkripció megakadása során, illetve a két rendszer összeütközésekor keletkeznek, amely viszonylag ritka esemény, a DSB-k eloszlása pedig sztochasztikus. A meiotikus sejtek viszont speciális DSB-, rekombinoszóma-, és szinaptonémális komplex fehérjéket expresszálnak (**8. ábra**), amelyek egyszerre nagyszámú törést generálnak térben és időben kontrollált módon (becsült számuk: 90-140 DSB / sejt; (Buhler et al. 2007), >1000-szeresére emelve a HR gyakoriságát (Székvölgyi et al. 2015b; Székvölgyi and Nicolas 2010).



8. ábra Meiotikus DSB-fehérjék *Saccharomyces cerevisiae*-ben. Az ábrán látható tíz DSB-fehérje funkcionális ortológja minden ivarosan szaporodó fajban megtalálható, bármelyikük hiánya azonnal megállítja a meiotikus rekombinációt és ivarsejtképződést (magasabbrendűekben), mert nem keletkeznek crossover rekombinánsok. Utóbbiból kromoszóma karonként legalább 1 szükséges a homológ kromoszómák szegregációjához és az aneuploidia és infertilitás elkerüléséhez (Székvölgyi et al. 2015b; Székvölgyi and Nicolas 2010). A katalitikus egység a Spo11 (amely egy topoizomeráz II-szerű transzészteráz), az MRX komplex végzi a DSB-végek visszavágását, amely a szálláthelyezés feltétele, a többi fehérje pedig szerkezeti faktorok amelyek a kromoszóma tengelyen lokalizálódnak és a Spo11 megfelelő pozícionálásához szükségesek. A Mer2 fehérje foszforilációja meiotikus S fázis kinázok által történik meg, amely elindítja a DSB komplex összeszerelését és egyben mechanikai kapcsolatot teremt a DNS replikáció és rekombináció iniciáció között (Murakami and Keeney 2008; Borde et al. 2000).

A meiotikus DSB-k és rekombináns molekulák emiatt kísérletesen viszonylag könnyen vizsgálhatók olyan modellorganizmusokban, ahol a folyamat biokémiai fázisai szinkronizáltan és rövid idő alatt zajlanak le, illetve bizonyos génmutációkkal specifikusan blokkolhatók az egyes stádiumok (**9. ábra**). Élesztőben például a *rad50S* mutáció a DSB keletkezés után / DSB processzálás előtt blokkol; a *dmc1* mutáció a processzált szimplaszálú filamentumok kialakulása után / a szálláthelyezés előtt blokkol; az *ndt80* mutáció az MI osztódás előtt blokkol, stb. Az így nyert összefüggések hasznos információkkal szolgálnak számunkra például az emberi ivarsejtképződés folyamatáról, hiszen az ivarsejtek keletkezése során a kromoszómák eltörnek, s e törések teszik alkalmassá a kromoszómákat arra, hogy rekombinációs folyamatokban vegyenek részt, amelyben DNS-szakaszok cserélődnek ki egymással. A crossovereknek két fontos élettani szerepe van. Az egyik mechanikai jellegű: az apai és anyai homológ kromoszómák fizikai kapcsolatának, tehát a törések révén történő összekapcsolódásának, összekapaszkodásának meg kell történnie ahhoz, hogy az ivarsejtképzés – amely egy számfelező, genomredukciós folyamat – végbemenjen, és a keletkező csírasejtek életképesek legyenek. A másik szerep genetikai szempontból fontos, mivel a kromoszóma egy adott helyén elhelyezkedő gén variációinak (alléleknek) kombinálódniuk kell egymással, hogy az utódok az adott környezetben életképesebbek legyenek.



9. ábra A meiózis biokémiai fázisai. S: meiotikus DNS replikáció DSB: DNS dupla-szál törés, CO: crossover rekombináció, SC: szinaptonémális komplex, MI: első meiotikus osztódás, MII: második meiotikus osztódás. Élesztő sejtekben a profázis I eseményei mintegy 6 óra alatt lezajlanak, emiatt a folyamat kísérletesen jól vizsgálható. Magasabbrendűekben ugyanezek a lépések követik egymást, csak szakaszosan, aszinkron módon, és időben rendkívül elnyújtva.

Tehát maga a DNS-törés és a törés következménye, a rekombináció a populációk szintjén ahhoz szükséges, hogy nagyobb legyen a populáció rátermettsége, alkalmazkodó képessége. Egy olyan populáció, amely genetikailag homogén, beltenyésztett, nem tud megfelelően alkalmazkodni a megváltozott környezethez. Ezzel szemben, egy heterogén (polimorf) populáció versenyelőnyt élvez, életképesebb, és éppen genetikai sokszínűsége az, ami képessé teszi arra, hogy a legkülönfélébb környezeti kihívások között is túléljen. Vagyis a folyamat genetikai jelentősége a sokszínűség megteremtése, ami a rekombináción alapul, a homológ rekombinációt pedig a DNS törések teszik lehetővé.

A Spo11-függő DSB-k keletkezését minden eddig vizsgált ivarosan szaporodó fajban visszaigazolták a legkülönfélébb metodikák felhasználásával, a '90-es évek végén elsősorban Southern blot hibridizációval majd ChIP-on-Chip microarray térképezéssel (Baudat and Nicolas 1997; Gerton et al. 2000; Buhler et al. 2007), az NGS korszakban pedig nagysűrűségű Spo11-oligo szekvenálással (Pan et al. 2011) és crossover szekvenálással (Mancera et al. 2008). Hogy az utóbbi genomikai módszereket megértsük, előbb érintenünk kell a rekombináció-térképezés klasszikus vizsgálómódszereit, amelyek elve a következő: ha sejtjük, hol keletkeznek a DNS-törések, akkor a törések következményét, vagyis a rekombinációt úgy tudjuk vizsgálni, hogy "berakunk" két jelölést (markert) a törés közvetlen környezetébe. A marker lehet két antibiotikum-rezisztencia gén, de lehet bármilyen más molekula is, amelyet detektálni és követni tudunk. Ha a két kromoszómaszakasz kicserélődik egymással, akkor a markerek is helyet cserélnek, amelyből rekombinációs gyakoriságot tudunk számolni. Ezzel a módszerrel azonban csupán egy adott szakaszt lehet vizsgálni. A teljes genom vizsgálatához több tízezer/százezer marker szükséges, amelyek elég sűrűn helyezkednek el ahhoz, hogy azok fázisát a meiózis során nyomon követve NGS szekvenálással feltérképezhessék az összes rekombinációs eseményt. A Spo11-oligo szekvenálás az eddigi legnagyobb felbontóképességű módszer, amely a Spo11 hasítási reakció során az enzimhez kovalensen kötve maradó ~12-40 nt hosszúságú, szimplaszálú DNS intermedierek biokémiai tisztításán és térképezésén alapul (Neale et al. 2005; Pan et al. 2011). E genomikai vizsgálatok alapján a meiotikus DSB-k és crossover-ek speciális kromoszómális régiókban, ún. "forrópontokban" keletkeznek, amely lehetővé teszi az alapvető sejtbiokémiai folyamatok (transzkripció, replikáció) zavartalan működését a profázis I alatt valamint a későbbi sejtdifferenciációs lépések során. A rekombinációs forrópontok többnyire nyitott (DNázI szenzitív), nukleoszóma mentes genomi régiókban találhatóak, amely szükséges, de nem elégséges feltétele a DNS törések bekövetkezésének (Lam and Keeney 2015; Singhal et al. 2015; Ohta et al. 1994). A meiotikus DSB-k pozíciója konkrét "konszenzus" DNS szekvenciához nem köthető, ám

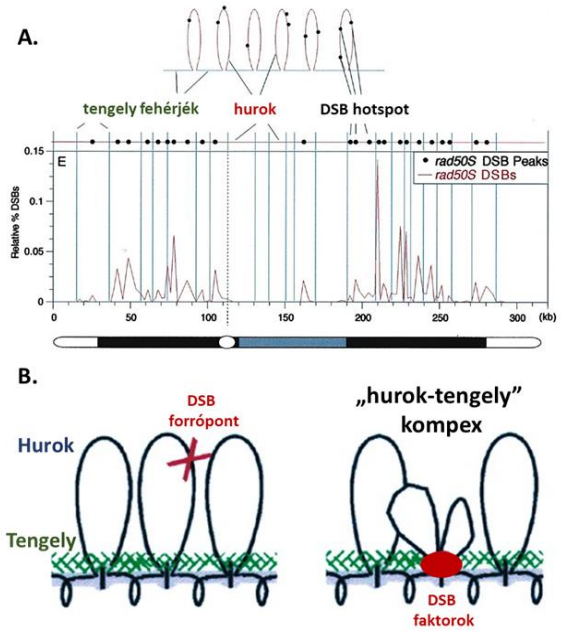
közvetlen környezetükben dúsulást mutatnak bizonyos “aktiváló” hiszton modifikációk (pl. H3K4me3, H3K36me3, H3K9ac) illetve bizonyos hiszton módosító enzimek deléciója (pl. Gcn5, Rpd3, Sir2, Set1, Set2, Dot1, Rad6, Him-17) néhány hotspot régió esetén szignifikánsan csökkentette a DSB keletkezés gyakoriságát (Székvölgyi et al. 2015b; Székvölgyi and Nicolas 2010). Ezek közül a H3K4 trimetiláció szerepe igazolódott vissza a teljes genomban élesztőtől az emberig, mivel mind a Set1 hiányos élesztő sejtek (Sollier et al. 2004; Borde et al. 2009), mind a Prdm9 (meiózis-specifikus H3K4 hiszton metiláz (Mihola et al. 2009; Berg et al. 2011) mutáns egerek, emberszabású majmok, és emberek súlyos meiotikus defektust mutattak. A mutáció érintette a DSB-k keletkezését és a crossover termékek lokalizációját; élesztőben a DSB-k 84%-a “eltűnt” Set1 hiányában (Sollier et al. 2004; Borde et al. 2009), Prdm9 KO egérben pedig a rekombináns termékek kevés számban és rossz helyen jelentek meg a genomban, például esszenciális gének belsejében, kromoszóma szerkezeti elemekben (centromerekben / telomerekben), amely infertilitáshoz vezetett (Auton et al. 2012; Parvanov et al. 2010; Camerini-Otero 2014; Myers et al. 2010; Parpanov et al. 2010; Brick et al. 2012). Fontos megemlíteni, hogy egyes gerincesekben (pl. kutyákban, madarakban) és növényekben nem találták meg a Prdm9-et vagy annak funkcionális ortológját, azonban a meiotikus DSB-k továbbra is nyitott, nukleoszóma-mentes kromatin régiókban (többnyire promóterekben) lokalizálódtak (Singhal et al. 2015; Choi et al. 2013; Auton et al. 2013), hasonlóan a Prdm9-függő hotspot-okhoz. A rekombinációs forrópontok tehát fajok között és populáción belül viszonylag stabilnak mondhatók, amelynek feltételezhető oka, hogy az evolúció során hotspot-ként szelektálódott kromoszómális elemek egyéb esszenciális folyamatokban is részt vesznek (pl. transzkripció, kromoszóma/sejtmag szerkezet), ezért e genomi régiók folyamatos szelekciós nyomás alatt vannak (Lichten 2015). Az evolúciós stabilitás ellenére az egyedek között jelentős variabilitás tapasztalható, amelynek oka a meiotikus kromoszómák térszerkezetében keresendő (Zuo et al. 2021; Jin et al. 2021; Fillér et al. 2018).

TÉRBELI GENOMSZERKEZET ÉS HOMOLÓG REKOMBINÁCIÓ

A kromatin funkciója szorosan kapcsolódik a genom háromdimenziós szerkezetéhez, amely természetesen érinti a homológ rekombináció folyamatát is. Interfázisos sejtekben a DNS pakolódása fraktálszerű (önhasonló) egységek egymásba-ágyazásával, skálázott molekuláris mechanizmusokkal jön létre a kisebb kromatin hurkoktól a többezer megabázisos territóriumokig (Bolzer et al. 2005; Lieberman-aiden et al. 2009). A hierarchia első szintjét a nukleoszómák képviselik, amelyek ~30 nm-es szolenoid rostokba majd 20-200 kb-os hurok doménekbe szerveződnek, amely lehetővé teszi a genom hatékony funkcionális elrendezését. A közelmúltban végzett molekuláris vizsgálatok a

genomszerveződés további szintjeit tárják fel, az ún. topológiailag asszociált doméneket (TAD, ~1 Mb) és az A/B kromatin kompartmenteket (~1-10 Mb) (Rao et al. 2014). Előbbiek a kromoszóma territóriumok alapvető strukturális és funkcionális építőkövei, melyek az enhanszerek hatását a TAD-on belüli génekre korlátozva lehetővé teszik a szövetspecifikus génexpressziós mintázatok kialakulását (Symmons et al. 2014). Utóbbi struktúrák pedig a klasszikus aktív eukromatinnak ("A") és inaktív heterokromatinnak ("B") feleltethetők meg (Lieberman-aiden et al. 2009). Ezt a felosztást különféle genomszintű vizsgálatokkal tovább finomították és egyedi tulajdonságokkal rendelkező funkcionális állapotokba ("states") sorolták a kromatin doméneket (Di Pierro et al. 2017; Ernst and Kellis 2012; Dekker and Mirny 2016), azonban ezek kialakulási mechanizmusa kevéssé ismert. A TAD-okról tudjuk, hogy határaik nagymértékben átfednek a különféle sejttípusok között (Feng et al. 2014; Dixon et al. 2015) és szerkezetük stabilan öröklődik az ivarsejtképzés során (Battulin et al. 2015). A TAD-okon belül elhelyezkedő kromatin hurkok ezzel szemben rendkívül dinamikusak, különböznek az egyes sejttípusok között, és erősen korrelálnak a sejtek transzkripciós aktivitásával (Rao et al. 2014). Emlősökben néhány strukturális fehérjét már jellemeztek, melyek fontos szerepet játszanak a kromatin hurkok kialakulásában és mutációjuk gyakran kapcsolódik fejlődési rendellenességekhez és más betegségekhez. Ilyen fehérje a CCCTC-kötő faktor (CTCF) (Wit et al. 2015; Vietri Rudan et al. 2015; Guo et al. 2015; Nagy et al. 2016; Nichols and Corces 2015; Tang et al. 2015), a YY1 (Weintraub et al. 2017), és a kohezin komplex tagjai (Haarhuis et al. 2017; Fudenberg et al. 2016; Mizuguchi et al. 2014; Sofueva et al. 2013; Costantino et al. 2020; Rao et al. 2017; Lazar-Stefanita et al. 2017; Busslinger et al. 2017), amelyek kötődését a helyi hiszton környezet, másodlagos DNS struktúrák és a DNS szekvencia szabályozza. Meiotikus sejtekben ezidáig a CTCF-paralóg BORIS szerepét igazolták (Rivero-Hinojosa et al. 2021), illetve azt is tudjuk, hogy a mitotikus kohézinek meiózis-specifikus kohézinekre cserélődnek (Rec8, Rad21L), amelyek a kromatin hurkok szervezéséhez szükségesek és ellenállnak a szeparáz hasításnak, amely a testvérkromatid kohézió fennmaradásához szükséges a meiózis I anafázisában (Watanabe 2012). A meiotikus kohézinek ún. tengelyfehérjékkel együtt (pl. Red1, Hop1, Zip1, Zip3) hozzák létre a homológ kromoszómák központi tengelyét ("axis"), amely az ezután kialakuló szinaptonémális komplex laterális elemeként funkcionálva proximitásba hozza az anyai és apai homológokat a rekombináció iniciációja során. A tengelyfehérjék citológiai és genomikai térképezése igazolta, hogy azok kolokalizálnak az esszenciális DSB-fehérjékkel (**8. ábra**; élesztőben: Spo11, Mer2, Mre11, Xrn2, Rad50, Ski8, Mei4, Rec102, Rec104) (Blat et al. 2002; Arora et al. 2004; Maleki et al. 2007; Panizza et al. 2011; Murakami et al. 2020; Karányi et al. 2018), vagyis a rekombináció biokémiai lépései (DNS törés, száláthelyezés, hibajavítás, rekombináns termékek keletkezése) a kromoszóma tengelyhez kötötten játszódnak le. A

DSB hotspot-ok viszont meglepő módon nem ezekre a szakaszokra esnek, hanem a tengellyel "ellentétes" kromatin hurok régiókra (Blat et al. 2002). Ezt a szerkezeti paradoxont a Nancy Kleckner által javasolt "hurok-tengely" modell bevezetése oldotta fel ("tethered loop-axis", (Blat et al. 2002)), amely szerint a kromatin hurkok (ahol a DSB forrópontok vannak) és a kromoszómák tengelye (ahol a töréseket okozó enzimek és rekombinoszóma fehérjék vannak) fizikai kölcsönhatásba kell lépjenek egymással (**10. ábra**).



10. ábra A hurok-tengely modell Blat és munkatársai nyomán. A kromatin tengelyt a Rec8 kohézin kötőhelyek definiálják, ahol a meiotikus DSB komplex összeszerelődik. A kromatin hurkokat a Rec8-kötőhelyek közé eső genomi szakaszok alapján azonosították. A DSB hotspot-ok (rad50S DSB-k) kizárólag a hurkokra esnek, amely a modell szerint a hurok szakaszok tengelyhez történő lefűződésével lehetséges. A fizikai kontaktus neve: "lefűződött hurok-tengely" komplex. Forrás: (Blat et al. 2002)

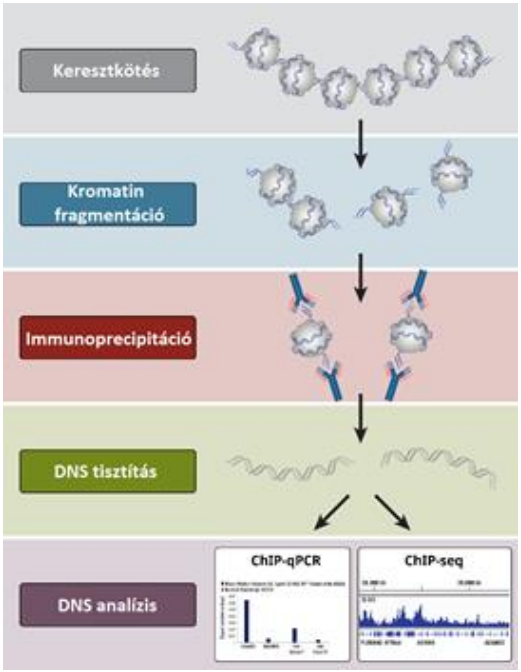
A feltételezett hurok-tengely kölcsönhatás molekuláris mechanizmusa egészen a közelmúltig ismeretlen maradt; a résztvevő molekuláris komponenseket a kutatócsoportunk karakterizálta francia és ausztrál kollaborációs partnerekkel együtt (Acquaviva et al. 2013; Karányi et al. 2018, 2020) amely az "*Eredmények*" fejezetben kerül bemutatásra.

RELEVÁNS KROMATINSZERKEZET VIZSGÁLÓ MÓDSZEREK

E fejezetben elvi szinten áttekintjük azokat kromatinszerkezet vizsgáló módszereket, amelyeket a kísérleteinkben alkalmaztuk illetve a kutatásaink szempontjából relevánsak.

1) A kromatinkötő fehérjék (pl. struktúrfehérjék, transzkripciós faktorok, hiszton módosító enzimek, hiszton modifikációk) genomi kötőhelyeinek a meghatározása és azok funkcionális annotálása elsősorban kromatin immunprecipitációval (ChIP) történik, amely manapság rutin eljárásnak számít. A ChIP egy immunológiai technika, amely során élő sejteket formaldehiddel fixálunk, a térhálósított kromatint ultrahanggal feldaraboljuk, majd a specifikus fehérje-DNS komplexeket szelektíven feldúsítjuk az adott fehérje ellen termeltetett specifikus antitestekkel (**12. ábra**). Az immunprecipitált fehérje-DNS komplexeket ezután megtisztítjuk és a DNS fragmentumokat mélyszekvenálással (NGS), vagy qPCR-ral azonosítjuk és kvantifikáljuk (ChIP-seq, illetve ChIP-qPCR). Az NGS read-eket a referenicagenomra térképezve bioinformatikai módszerekkel meghatározzuk a

statisztikai dúsulást mutató kromoszómális pozíciókat, amelyet kötőhelyeknek (peak-ek) hívunk. Az annotált kötőhelyek összessége a cisztrom, amely alapján a vizsgált fehérje célpontjai és biológiai funkciói prediktálhatók, illetve funkcionális vizsgálatokkal validálhatók.



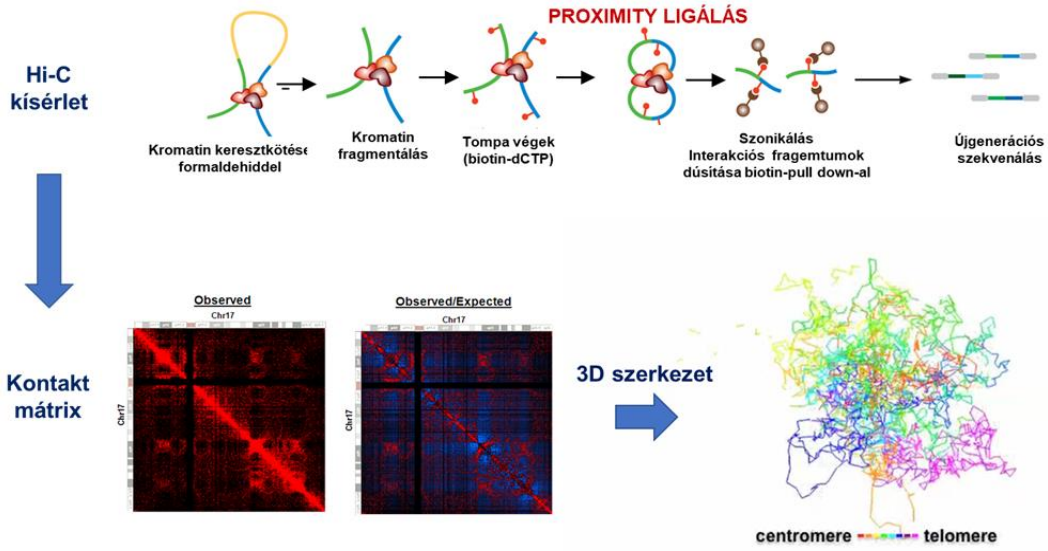
12. ábra A kromatin immunoprecipitáció (ChIP) lépései. Első lépés a fehérjék keresztkötése a DNS-sel. Ezt követi a kromatin fragmentációja (általában szonikálással), majd az immunoprecipitáció, amely során egy specifikus antitesttel szelektíven feldúsítjuk a vizsgálni kívánt fehérje-DNS komplexeket. Tisztítási lépésekkel eltávolítjuk az aspecifikus termékeket, majd újgenerációs szekvenálással (NGS) vagy qPCR technikával azonosítjuk a genomi kötőhelyeket.

2) Az R-hurkok elektronmikroszkóppal történő felfedezése óta több molekuláris technika áll rendelkezésünkre e struktúrák azonosítására. A terület egyik legfontosabb mérföldköve egy RNS-DNS hibrid-specifikus monoklonális antitest (S9.6) kifejlesztése volt (Hu et al. 2006), amely szekvencia kontextustól függetlenül nagy affinitással kötődik az R-hurkok RNS-DNS hibrid részéhez. Az S9.6 antitest lehetővé teszi az R-hurok *in vitro* és *in vivo* tanulmányozását különféle molekuláris technikákkal, mint például az immunfluoreszcens képalkotás, microarray elemzés, vagy a NGS szekvenálás. Az egyik legelterjedtebb módszer az DNS-RNS immunprecipitáció szekvenálás (DRIP-seq, (Ginno et al. 2012), amely során az extrahált nukleinsavat szonikálással vagy restrikciós enzimekkel fragmentáljuk, majd az RNS-DNS hibrideket S9.6 antitesttel immunprecipitáljuk. Végül a tisztított fragmentumokat NGS vagy qPCR segítségével kvantifikáljuk és térképezzük. Általában *in vitro* RNázH (*E. coli*) kezelést használunk negatív kontrollként, amely az S9.6 IP előtt specifikusan degradálja az izolált nukleinsavban található R-hurkok RNS-szálát. Néhány évvel az eredeti DRIP protokoll után számos kiegészítő módszert fejlesztettek ki, amelyeket az immunprecipitáció célpontja (DNS vagy protein), a szekvenált molekula típusa (DNS vagy RNS), és az NGS könyvtárkészítés módja szerint lehet csoportosítani. Az S1-DRIP-seq S1 nukleáz kezelést alkalmaz a szimplaszálú DNS szakaszok eltávolítására az RNS-DNS hibrid immunprecipitáció előtt, ami jobb jel/zaj arányt eredményez (Wahba et al. 2016). A módszer hátránya, hogy az S1 nukleáz savas puffere (pH 4.6) DNS károsodást okoz (elsősorban nick-eket és depurinációt), amelyeket az S1 specifikusan felismer és hasít. Az ún. DRIP-RNS-seq és DRIPc-seq módszerek a DRIP protokoll lépéseit követik az immunprecipitációs lépéséig (Nadel et al. 2015; Sanz and Chédin 2019), majd az RNS-DNS hibrideket denaturálják és a DNS szálakat DNázI emésztéssel eltávolítják. A fennmaradó RNS molekulákból

szálspecifikus RNS-seq könyvtárat készítenek, majd megszekvenálják azt. Utóbbi megközelítésnek egyértelmű előnye, hogy a hibridek szál-orientációjáról is információt nyújtanak, hátrányuk a rendkívül komplikált protokoll és az inkomplett DNázI emésztés által fennmaradó DNS kontamináció. Hasonló módszer az ún. ssDRIP-seq (Xu et al. 2017), amelyben a feldúsított RNS-DNS hibrideket szonikáják és hődenaturálják, hogy egyszálú DNS-t kapjanak, amelyhez szimplaszálú adaptert ligálnak RNS-ligáz enzimmel a könyvtárkészítés és a szekvenálás előtt. Az ún. DNS-RNS *in vitro* dúsítás (DRIVE-seq) és az R-ChIP módszerek katalitikusan inaktív, mutáns RNázH1 enzimet alkalmaznak az S9.6 antitest helyett, amelyek specifikusan megkötik de nem hasítják az RNS-DNS hibrideket (Ginno et al. 2012; Chen et al. 2017). A DRIVE-seq egy affinitásos *in vitro* „pulldown" vizsgálat, míg az R-ChIP egy *in vivo* kromatin immunprecipitációs módszer, amely a mutáns RNázH1 expresszióján alapul. Az R-ChIP hátránya, hogy a vizsgálni kívánt sejteket először stabilan transzfektálni kell az RNázH1-GFP plazmiddal, amely bizonyos esetekben nehezen kivitelezhető illetve bizonyos genomi szakaszokhoz (pl. tRNS gének, rDNA gének) nem kötődik megfelelően a konstrukt (ismeretlen okok miatt). A legújabb módszer az ún. bisDRIP-seq, mely natív (nem-denaturáló) biszulfit-modifikációt és S9.6 immunoprecipitációt kombinálva lehetővé teszi az R-hurkok szálspecifikus feltérképezését a genomban (Dumelie and Jaffrey 2017). A módszer koncepciója, hogy a natív körülmények között végzett biszulfit kezelés szelektíven átalakítja a nem-metilált citozint uracilra az R-hurok szimplaszálú DNS-szakaszában, míg a duplaszálú RNS-DNS hibrid rész illetve a duplaszálú DNS védett marad a C-U konverzióktól. A bisDRIP-seq legfőbb hátránya a citozinok és a DNS metiláció egyenetlen eloszlása a genomban, illetve a biszulfit kezelés okozta DNS károsodás és mintaveszteség csökkenti az S9.6 immunprecipitáció hatékonyságát és specificitását. További technológia az ssDNS-t felismerő replikációs fehérje A (RPA) használata az immunprecipitációs lépésben az S9.6 antitest helyett (RPA ChIP), amely azonban az R-hurkoktól függetlenül keletkező szimplaszálú DNS szakaszokat is feldúsítja. A fenti módszereket rendszeresen használják R-hurkok detektálására, ám a specificitás és szenzitivitás fajtól és kísérleti rendszertől függően jelentős eltéréseket mutat. Az inkonzisztencia oka sokáig ismeretlen maradt, ennek felderítésére egy szisztematikus vizsgálat keretén belül a kutatócsoportunk vállalkozott (Roszik et al. 2017; Halász et al. 2017).

3) A genom térszerkezetének tanulmányozásához két fő megközelítés létezik: mikroszkópos és molekuláris vizsgálatok. A hagyományos fénymikroszkópos technikák kisfelbontású (max. 250 nm) fizikai információt nyújtanak a kromoszómák eloszlásáról és alakjáról az egyedi sejtekben, a felbontóképesség azonban különféle szuperrezolúciós megoldásokkal a nanométeres tartományig növelhető (pl. STORM, STED, PALM, TIRF). A molekuláris megközelítések az ún. kromoszóma

konformáció léképezés technológián és szekvenáláson alapulnak (Hi-C), amely a mikroszkópos módszerekkel ellentétben lehetővé teszi egyedi kromoszóma szakaszok és funkcionális elemek kölcsönhatásának tömeges vizsgálatát egy sejtpopulációban. A Hi-C felbontóképessége sokat javult az elmúlt évek során, az első 1 Mb-os Hi-C mátrixhoz képest (Lieberman-aiden et al. 2009) mára elérte a szubkilobázisos / nukleoszómális feloldást (~146 bp; (Hsieh et al. 2016). A Hi-C és Hi-C-variáns módszerek (pl. 4C, 5C, 6C, ChIA-PET, ChIP-Loop, HiChIP, micro-C) térbeli interakciós valószínűségeket jelenítenek meg bármely két genomi szakasz között egy sejtpopulációban (Jerkovic´ and Cavalli 2021), amelyek azonos vagy eltérő kromoszómán találhatóak (**13. ábra**). A kapott intra- és interkromoszómális interakciók gyakorisága egy bizonyos dinamikus tartományon belül fordítottan arányos a két szakasz térbeli távolságával, tehát minél nagyobb a térképezett Hi-C read-ek száma, annál kisebb a két szakasz térbeli távolsága. Érdemes megjegyezni, hogy mára egyedi-sejtes megközelítések is rendelkezésre állnak a Hi-C elemzéshez (single-cell Hi-C, (Nagano et al. 2015), azonban ezek csupán a térbeli interakció meglétéről adnak információt (van / nincs), de annak mértékét nem kvantifikálják (a detektált interakciók kis számossága miatt). Az újabban kifejlesztett tiramidszignál amplifikáció szekvenálás (TSA-seq) képes a kromoszómális szakaszok valós citológiai távolságának számítására a teljes-genomban (Chen et al. 2018), ötvözve a Hi-C módszer és a fénymikroszkópia előnyös tulajdonságait.



13. ábra A genomkonformáció leképezés (Hi-C) elve. A térben közel elhelyezkedő DNS szálakat formaldehiddel keresztkötik, restrikciós enzimekkel fragmentálják és biotin-jelölt ragadós végeket képeznek. Intramolekuláris ligálással összekapcsolják a "kiméra" szálakat (proximity ligálás), amelyek a genom különböző pontjaiból származnak, akár eltérő kromoszómákról. A ligált kiméra fragmentumokat anti-biotin gyöngyökkel dúsítják majd paired-end szekvenálással térképezik. A 3D kromatin interakciók jellegzetes mintázatot adnak a hőtérképen. Minden genomi fragmentum önmagával adja a legnagyobb interakciós frekvenciát, ezzel egy erős jelet kapunk az átlós tengely mentén. A DNS hurkok önmagukba visszahajlása pontszerű mintázatot adnak, míg a TAD-ok az átlós tengely mentén elhelyezkedő négyzet alakú struktúrát

alkotnak. A kromoszómák közötti interakciók az átlós tengelytől távolabb elhelyezkedő mezőt alkotnak. Az interakciós gyakoriságok matematikai módszerekkel (pl. multidimenzionális skálázás) fizikai távolsággá konvertálhatók, amely alapján megrajzolható a kromoszómák (átlagos) 3D szerkezete.

A fenti módszereket egyre több fajban és kísérleti körülmények között alkalmazzák az interfázisos és mitotikus kromoszómák térszerkezetének felderítésére (Hoencamp et al. 2021), azonban a meiotikus kromoszómák felépítéséről és dinamikájáról nagyon keveset tudunk. A meiózis stádiumainak tanulmányozásához homogén sejtpopulációra illetve szövetmintára, sejtszinkronizálásra, és specifikus mutánsokra van szükség, amely magasabbrendű fajokban csak rendkívül speciális körülmények között kivitelezhető. Élesztő modellorganizmusban ezidáig két meiotikus Hi-C tanulmány született (Muller et al. 2018; Schalbetter et al. 2019), amely megerősítette a korábbi citológiai vizsgálatok eredményét (pl. Rabl konfiguráció, bouquet stádium, Rec8 kohézin szerepe), azonban a meiotikus homológ rekombináció folyamata és annak kromatinszerkezeti aspektusai mindezidáig feltáratlan maradt.

CÉLKITŰZÉSEK

1) Kutatómunkám során célom volt a meiotikus homológ rekombináció iniciációjának molekuláris vizsgálata, a DSB képződés epigenetikai szabályozásának mélyebb megértése, a főbb molekuláris komponensek és fehérje-fehérje interakciós partnerek azonosítása. Kiemelt figyelmet fordítottam a Set1C/COMPASS hiszton metiláz komplex szerepére, mivel a hiszton H3K4 trimetiláció és a rekombinációs forrópontok statisztikai korrelációja már korábban ismert volt (Borde et al. 2009), azonban az ok-okozati kapcsolat bizonyítása nem sikerült. További célom volt a "hurok-tengely komplex" hipotézis alátámasztása vagy elvetése a kapott adatok alapján. Célom volt megvizsgálni más hiszton modifikációk szerepét is a fenti folyamatokban, amelyek nem az N-terminális hiszton részekre koncentrálódnak, hanem a nukleoszómák központi globuláris doménjére esnek (pl. H3K56 acetiláció).

A fenti kísérleteket pékélesztő (*Saccharomyces cerevisiae*) modellorganizmusban végeztem és az alábbi publikációk születtek a témában:

1. Karányi, Z., Hornyák, L., Székvölgyi, L. Histone H3 Lysine 56 Acetylation Is Required for Formation of Normal Levels of Meiotic DNA Breaks in S. cerevisiae. *Front. Cell Dev. Biol.* **7**, 364 (2020).

2. Karányi, Z. *et al.* Nuclear dynamics of the Set1C subunit Spp1 prepares meiotic recombination sites for break formation. *Journal of Cell Biology* **217**, 3398–3415 (2018).

3. Székvölgyi, L., Ohta, K. & Nicolas, A. Initiation of Meiotic Homologous Recombination: Flexibility, Impact of Histone modifications, and chromatin remodelling. *Cold Spring Harb Perspect Biol.* **7**, 447–462 (2015).

4. Acquaviva, L* & Székvölgyi, L* *et al.* The COMPASS subunit Spp1 links histone methylation to initiation of meiotic recombination. *Science* **339**, 215–8 (2013). * joint first authors

5. Székvölgyi, L. & Nicolas, A. From meiosis to postmeiotic events: homologous recombination is obligatory but flexible. *FEBS J.* **277**, 571–89 (2010).

2) Az R-hurkok azonosítása és kromoszómális térképezése kulcsfontosságú a genom stabilitását szabályozó folyamatok megértéséhez és a repair faktorok és DNS léziók kromatinszerkezeti aspektusainak feltárásához. Célom volt egy olyan univerzális genomikai térképezési módszer kidolgozása, amely az "alap" DRIP módszer inkonzisztenciáit kikerülve alkalmas az R-hurkok nagypontosságú azonosítására bármely modellorganizmusban. További célom volt, hogy összefüggést találjak a kromatin hurok letapadási pontok, a DNS törések és az R-hurkok sejtmagi lokalizációja között. A térbeli genomszerkezet vizsgálatok kapcsán kiemelt célom volt a Hi-C módszer

implementálása a laboratóriumomban. A fenti eredményekre és metodikákra alapozva célul tűztem ki új R-hurok regulátor funkcióval rendelkező fehérjék azonosítását és genomikai karakterizálását. A lúdfű (*Arabidopsis thaliana*) kicsiny genomja (~130 Mb) és genetikai manipulálhatósága miatt tökéletes alanya e komplex vizsgálatoknak, elsőként ebben a modellorganizmusban tanulmányoztam a *Nodulin homeobox* (NDX) transzkripciós faktor szerepét az R-hurok képződés, DNS metiláció, és a magasabbrendű 3D kromatinszerkezet összefüggésében. Humánspecifikus R-hurok szabályozó fehérjékkel pedig farmakogenomikai elemzéseket végeztem, hogy összefüggést találjak az R-hurok regulátorok expressziója és különféle rákos megbetegedésekben mutatott túlélés-asszociációk és kemoterápiás szerekkel szembeni érzékenység/rezisztencia kialakulása között. Utóbbi vizsgálatok, az alapkutatási jelentőségükön túl, utat nyitottak az R-hurkok és R-hurok szabályozó fehérjék klinikai kiaknázásához diagnosztikai markerként vagy terápiás célpontként. Ebben a témakörben az alábbi publikációk születtek:

1. Karányi Zs. Mosolygó-L Á, Feró O, Horváth A, Boros-Oláh B, Nagy É, Hetey Sz, Holb I, Szaker H, Miskei M, Csorba T, Székvölgyi L: NODULIN HOMEOBOX is required for heterochromatin homeostasis in Arabidopsis. *Nature Communications.* **13** 5058, 2022. (D1, IF: 17.694)

2. Boros-oláh, B. *et al.* Drugging the R-loop interactome: RNA-DNA hybrid binding proteins as targets for cancer therapy. *DNA Repair.* **84**, 102642 (2019).

3. Hegedüs, É. *et al.* Endogenous single-strand DNA breaks at RNA polymerase II promoters in Saccharomyces cerevisiae. *Nucleic Acids Res.* **46**, 10649–10668 (2018).

4. Roszik, J., Fenyőfalvi, G., Halász, L., Karányi, Z., Székvölgyi, L. In Silico Restriction Enzyme Digests to Minimize Mapping Bias in Genomic Sequencing. *Mol. Ther. Methods Clin. Dev.* **6**, 66–67 (2017).

5. Halász, L. *et al.* RNA-DNA hybrid (R-loop) immunoprecipitation mapping: an analytical workflow to evaluate inherent biases. *Genome Res.* **27**, 1063–1073 (2017).

6. Székvölgyi, L. *et al.* Ribonucleoprotein-masked nicks at 50-kbp intervals in the eukaryotic genomic DNA. *Proc. Natl. Acad. Sci. U. S. A.* **104**, 14964–9 (2007).

ANYAGOK ÉS MÓDSZEREK

Élesztő mutánsok

A kísérletekben használt *Saccharomyces cerevisiae* törzsek az SK1 genomi háttérből származtak, azonosítójuk és genotípusuk az 1. táblázatban található. Az összes géndeléciót az EUROFAN II géndeléciós könyvtárból (http://www.euroscarf.de/index.php?name=Description) származó PCR-amplifikált KanMX4 vagy HphMX vagy NatMX deléciós kazettákkal való transzformációval és keresztezéssel vittük be a szülői sejtekbe. Minden géndeléciót Southern blot-tal és PCR-rel validáltunk. A plazmid shuffle mutánsokat izogén *hht1Δ*::HphMX *hht2Δ*::KanMX pARS-CEN(*HHT2*, *HHF2*-URA3) és *hht1Δ*::HphMX *hht2Δ*::KanMX *dmc1Δ*::LEU2 pARS-CEN(*HHT2*, *HHF2*-URA3) törzsek transzformálásával hoztunk létre, majd a haploid szülők keresztezésével és tetrádelemzéssel homozigóta diploidokat készítettünk. A mutáns hiszton klasztereket a pARS-CEN(*hht2*-"mut", *HHF2*-TRP1) plazmiddal vittük be a haploid szülői sejtekbe, amellyel kicseréltük ("shuffle") a vad típusú hisztonokat kódoló URA3 plasmidot. A hiszton pontmutációkat a QuikChange II PCR mutagenezis kit-tel (Agilent) vittük be a fenti TRP1 plazmidokba. A vad típusú URA3 plazmidot éjszakán át, gazdag táptalajon (YPD) végzett növesztéssel szegregáltattuk, majd egyedi telepeket növesztettünk YPD plate-en, amelyeket ezt követően ura-/trp- drop-out lemezekre replikáltunk. Végül a trp+/ura- telepeket azonosítottuk. A hiszton pontmutációk jelenlétét Sanger szekvenálással igazoltuk. A Set1C/COMPASS alegységek az UAS_{GAL} szekvenciához történő célzott irányításához használt konstrukciókat PCR-rel állítottuk elő, a géneket NcoI/Pst1 vagy BamHI/PstI fragmentumok formájában vittük be a pASIN plazmidba a Gal4 DNS-kötő domént (GBD) kódoló szekvenciától 3'-irányban. A GBD-hez való in-frame fúziókat Sanger szekvenálással validáltuk, a fúziós fehérjék expresszióját pedig western blot-tal ellenőriztük anti-Gal4 monoklonális antitestek (Euromedex) alkalmazásával. A GBD-Set1C fúziós fehérjéket a konstitutív pADH1 promóter szabályozása alatt fejeztük ki az endogén *trp1-1* lókuszba integrálva. Minden GBD-törzsben delécióval inaktiváltuk a megfelelő vad típusú Set1C allélt, hogy elkerüljük a fehérje kompetíciót a GBD- és endogén allélok között az UAS_{GAL} szekvenciához való kötődésért. A co-IP kísérletekhez az Spp1 fehérjét 3xHA epitóppal fúzionáltattuk, amelyet az *SPP1* gén C-terminális részébe vittünk be a pFA6a-3HA-KanMX6 plazmidról. Az SPP1-3xHA fehérje expressziója a 2 mikronos pBD827 plazmidról történt. Az SPP1 PHD doménjének delécióját két PCR amplikon fuzionálásával végeztük, amelyek -569 pozíciótól (ATG-hez képest) +3-ig és +235-től +1062-ig terjedtek. Ezt a PCR terméket a Yiplac211 *URA3* integratív plazmidba klónoztuk, amelyet NruI-el linearizáltunk. A kromoszómába integrált plazmidot hordozó sejteket 5-fluoroorotsav (FOA) lemezekre oltottuk, hogy kiszelektálhassuk az spp1Δ4-234

mutáns allélt hordozó, spp1ΔPHD-nek elnevezett rekombinánsokat. Az Spp1 cink-ujj (CXXC) motívumának (C263GYC266) delécióját úgy végeztük el, hogy a GBD-SPP1-et hordozó integratív vektorból (pASIN-SPP1) 12 bp-t kitöröltünk a Quickchange site-directed mutagenesis PCR segítségével. A GBD-SPP1Δ789-801 mutáns a GBD-SPP1ΔCXXC nevet kapta. A pASIN-SPP1ΔCXXC plazmidot XbaI-gyel linearizáltuk, majd az endogén *trp1-1* lókuszba integráltuk.

1. Táblázat. A felhasznált élesztő mutánsok listája.

#	Azonosító	Genotípus	Referencia
1	LAK20	sae2::KanMX	Acquaviva L & Székvölgyi L (2013)
2	LAK21	sae2::KanMX set1::KanMX	Acquaviva L & Székvölgyi L (2013)
3	LAK22	sae2::KanMX bre2::KanMX	Acquaviva L & Székvölgyi L (2013)
4	LAK23	sae2::KanMX sdc1::TRP1	Acquaviva L & Székvölgyi L (2013)
5	LAK24	sae2::KanMX swd1::KanMX	Acquaviva L & Székvölgyi L (2013)
6	LAK25	sae2::KanMX swd3::KanMX	Acquaviva L & Székvölgyi L (2013)
7	LAK26	sae2::KanMX shg1::TRP1	Acquaviva L & Székvölgyi L (2013)
8	LAK27	sae2::KanMX spp1::HphMX	Acquaviva L & Székvölgyi L (2013)
9	LAK20G-A	sae2::KanMX gal4::NatMX trp1::GAL4BD-TRP1	Acquaviva L & Székvölgyi L (2013)
10	LAK21G-B	sae2::KanMX gal4::NatMX set1::KanMX trp1::pADH1-GAL4BD-SET1-TRP1	Acquaviva L & Székvölgyi L (2013)
11	LAK21G-C	sae2::KanMX gal4::NatMX set1::KanMX trp1::pADH1-GAL4BD-set1G951S-TRP1	Acquaviva L & Székvölgyi L (2013)
12	LAK21/2G-B	sae2::KanMX gal4::NatMX set1::KanMX trp1::pADH1-GAL4BD-SET1-TRP1 bre2::KanMX	Acquaviva L & Székvölgyi L (2013)
13	LAK21/1G-B	sae2::KanMX gal4::NatMX set1::KanMX trp1::pADH1-GAL4BD-SET1-TRP1 spp1::HphMX	Acquaviva L & Székvölgyi L (2013)
14	LAK21/8G-B	sae2::KanMX gal4::NatMX set1::KanMX trp1::pADH1-GAL4BD-SET1-TRP1 bre1::HphMX	Acquaviva L & Székvölgyi L (2013)
15	LAK21/11G-B	sae2::KanMX gal4::NatMX set1::KanMX trp1::pADH1-GAL4BD-SET1-TRP1 hht1::URA3	Acquaviva L & Székvölgyi L (2013)
16	LAK21/12G-B	sae2::KanMX gal4::NatMX set1::KanMX trp1::pADH1-GAL4BD-SET1-TRP1 hht1::URA3 hht2-K4R	Acquaviva L & Székvölgyi L (2013)
17	LAK21/10G-B	sae2::KanMX gal4::NatMX set1::KanMX trp1::pADH1-GAL4BD-SET1-TRP1 spo11::KanMX	Acquaviva L & Székvölgyi L (2013)
18	LAK27G-F	sae2::KanMX gal4::NatMX spp1::HphMX trp1::pADH1-GAL4BD-SPP1-TRP1	Acquaviva L & Székvölgyi L (2013)
19	LAK07G-F	gal4::NatMX spp1::HphMX trp1::pADH1-GAL4BD-SPP1-TRP1	Acquaviva L & Székvölgyi L (2013)

20	LAK27/2G-F	sae2::KanMX gal4::NatMX spp1::HphMX trp1::pADH1-GAL4BD-SPP1-TRP1 bre2::KanMX	Acquaviva L & Székvölgyi L (2013)
21	LAK27/1G-F	sae2::KanMX gal4::NatMX spp1::HphMX trp1::pADH1-GAL4BD-SPP1-TRP1 set1::KanMX	Acquaviva L & Székvölgyi L (2013)
22	LAK27/8G-F	sae2::KanMX gal4::NatMX spp1::HphMX trp1::pADH1-GAL4BD-SPP1-TRP1 bre1::HphMX	Acquaviva L & Székvölgyi L (2013)
23	LAK27/11G-F	sae2::KanMX gal4::NatMX spp1::HphMX trp1::pADH1-GAL4BD-SPP1-TRP1 hht1::URA3	Acquaviva L & Székvölgyi L (2013)
24	LAK27/12G-F	sae2::KanMX gal4::NatMX spp1::HphMX trp1::pADH1-GAL4BD-SPP1-TRP1 hht1::URA3 hht2-K4R	Acquaviva L & Székvölgyi L (2013)
25	Mer2-myc, Spp1-HA	Mata/alpha, lys2/lys2, ho::LYS2/ho::LYS2, trp1-1/trp1-1, ura3/ura3, leu2/LEU2,HIS4/his4, SPP1-3xHA::KanMX4/SPP1-3xHA::KanMX4, MER2-9xmyc::TRP1/MER2-9xmyc::TRP1	Karányi Zs et al. (2018)
26	Mer2-HA, Bre2-myc	Mata/alpha, lys2/lys2, ho::LYS2/ho::LYS2, trp1-1/trp1-1, ura3/ura3, leu2/LEU2, HIS4/his4, arg4/arg4, BRE2-13xmyc::NaTMX4/BRE2-13xmyx::NatMX4, MER2-3xHA-int::NatMX4/MER2-3xHA-int::NatMX4	Karányi Zs et al. (2018)
27	Spp1PHDΔ-myc, Mer2-HA	Mata/alpha, ho::LYS2/ ho::LYS2, ura3/ura3, lys2/lys2, arg4/arg4, trp1/trp1, his4/his4, leu2/leu2, MER2-HA3-int/MER2-HA3-int, spp1ΔPHD-Myc9::KanMX/spp1ΔPHD-Myc9::KanMX	Karányi Zs et al. (2018)
28	Spp1CXXCΔ-myc, Mer2-HA	Mata/alpha, ho::LYS2/ ho::LYS2, ura3/ura3, lys2/lys2, arg4/arg4, trp1/trp1, his4/his4, leu2/leu2, MER2-HA3-int/MER2-HA3-int, spp1Δcxxc-13Myc-NatMX/spp1Δcxxc-13Myc-NatMX	Karányi Zs et al. (2018)
29	Spp1-myc	Mata/alpha, SPP1-13xmyc-KanMX4/SPP1-13xmyc-KanMX4, leu2/LEU2, HIS4/his4, trp1/trp1, ura3/ura3	Karányi Zs et al. (2018)
30	Bre2-myc	Mata/alpha, BRE2-13xmyc-NatMX4/BRE2-13xmyc-NatMX4, leu2/LEU2, HIS4/his4, trp1/trp1, ura3/ura3	Karányi Zs et al. (2018)
31	Spp1PHDΔ-myc	Mata/alpha, spp1PHDΔ-13xmyc-KanMX4/spp1PHDΔ-13xmyc-KanMX4, leu2/LEU2, HIS4/his4, trp1/trp1, ura3/ura3	Karányi Zs et al. (2018)
32	Spp1CXXCΔ-myc	Mata/alpha, spp1CXXCΔ-13xmyc-NatMX4/spp1CXXCΔ-13xmyc-NatMX4, leu2/LEU2, HIS4/his4, trp1/trp1, ura3/ura3	Karányi Zs et al. (2018)
33	Spp1-HA, H3	Mata/alpha, lys2/lys2, ho::LYS2/ho::LYS2,trp1-1/trp1-1, ura3/ura3,leu2/LEU2,HIS4/his4, hht1Δ::HphMX/hht1Δ::HphMX, hht2Δ::KanMX4/hht1Δ::HphMX, SPP1-3xHA::KanMX4/SPP1-3xHA::KanMX4, pCEN-ARS(HHT2,HHF2,TRP1)	Karányi Zs et al. (2018)
34	Spp1-HA, H3R2A	Mata/alpha, lys2/lys2, ho::LYS2/ho::LYS2, trp1-1/trp1-1, ura3/ura3, leu2/LEU2, HIS4/his4, hht1Δ::HphMX/hht1Δ::HphMX, hht2Δ::KanMX4/hht1Δ::HphMX, SPP1-3xHA::KanMX4/SPP1-3xHA::KanMX4, pCEN-ARS(HHT2R2A,HHF2,TRP1)	Karányi Zs et al. (2018)
35	Spp1-HA, H3K4R	Mata/alpha, lys2/lys2, ho::LYS2/ho::LYS2, trp1-1/trp1-1, ura3/ura3, leu2/LEU2, HIS4/his4, hht1Δ::HphMX/hht1Δ::HphMX, hht2Δ::KanMX4/hht1Δ::HphMX, SPP1-3xHA::KanMX4/SPP1-3xHA::KanMX4, pCEN-ARS(HHT2K4R,HHF2,TRP1)	Karányi Zs et al. (2018)
36	GFP-Spp1	Mata/alpha, SPP1-13xmyc-KanMX4/NatMX4-pCUP1-1-yGFP-SPP1, leu2/LEU2, HIS4/his4, trp1/trp1, ura3/ura3	Karányi Zs et al. (2018)
37	GFP-Set1	Mata/alpha, 9xmyc-SET1::LEU2/NatMX4-pCUP1-1-yGFP-SET1, leu2/leu2, HIS4/his4, trp1/trp1, ura3/ura3	Karányi Zs et al. (2018)
38	GFP	Mata/alpha, leu2/leu2, his4/his4, trp1/trp1, ura3/ura3, pRS426 (URA3, GFP-NLS)	Karányi Zs et al. (2018)

39	NHY1210	Mat a /α, ho::hisG / ho::hisG, leu2::hisG / leu2::hisG, ura3(ΔSmaI-PstI) / ura3(ΔSmaI-PstI), HIS4::LEU2-(BamHI;+ori) / his4-X::LEU2-(NgoMIV; +ori)--URA3	Karányi Zs et al. (2020)
40	asf1Δ	Mat a /α, ho::hisG / ho::hisG, leu2::hisG / leu2::hisG, ura3(ΔSmaI-PstI) / ura3(ΔSmaI-PstI), HIS4::LEU2-(BamHI;+ori) / his4-X::LEU2-(NgoMIV; +ori)--URA3, asf1Δ::KanMX / asf1Δ::KanMX	Karányi Zs et al. (2020)
41	hst3/4Δ	Mat a /α, ho::hisG / ho::hisG, leu2::hisG / leu2::hisG, ura3(ΔSmaI-PstI) / ura3(ΔSmaI-PstI), HIS4::LEU2-(BamHI;+ori) / his4-X::LEU2-(NgoMIV; +ori)--URA3, hst3Δ::KanMX / hst3Δ::KanMX, hst4Δ::HphMX / hst4Δ::HphMX	Karányi Zs et al. (2020)
42	rtt109Δ	Mat a /α, ho::hisG / ho::hisG, leu2::hisG / leu2::hisG, ura3(ΔSmaI-PstI) / ura3(ΔSmaI-PstI), HIS4::LEU2-(BamHI;+ori) / his4-X::LEU2-(NgoMIV; +ori)--URA3, trp1-1 / trp1-1, rtt109Δ::TRP1 / rtt109Δ::TRP1	Karányi Zs et al. (2020)
43	H3 (ctrl) DMC1	Mat a /α, ho::hisG / ho::hisG, hht1::HphMX / hht1::HphMX, hht2::KanMX / hht2::KanMX, trp1-1 / trp1-1, his4 / his4, ura3 / ura3, pARS-CEN(HHT2, HHF2, TRP1)	Karányi Zs et al. (2020)
44	H3 (ctrl) dmc1Δ	Mat a /α, ho::hisG / ho::hisG, dmc1::LEU2 / dmc1::LEU2, hht1::HphMX / hht1::HphMX, hht2::KanMX / hht2::KanMX, trp1-1 / trp1-1, his4 / his4, ura3 / ura3, pARS-CEN(HHT2, HHF2, TRP1)	Karányi Zs et al. (2020)
45	H3K56A DMC1	Mat a /α, ho::hisG / ho::hisG, hht1::HphMX / hht1::HphMX, hht2::KanMX / hht2::KanMX, trp1-1 / trp1-1, his4 / his4, ura3 / ura3, pARS-CEN(hht2-K56A, HHF2, TRP1)	Karányi Zs et al. (2020)
46	H3K56A dmc1Δ	Mat a /α, ho::hisG / ho::hisG, dmc1::LEU2 / dmc1::LEU2, hht1::HphMX / hht1::HphMX, hht2::KanMX / hht2::KanMX, trp1-1 / trp1-1, his4 / his4, ura3 / ura3, pARS-CEN(hht2-K56A, HHF2, TRP1)	Karányi Zs et al. (2020)

Arabidopsis minták

Kísérleteinkben a következő *Arabidopsis thaliana* (lúdfű) ökotípusokat használtuk: Col-0, ndx1-4, flag-NDX (genotípus: flag-NDX / ndx1-1(FRI) / flc-2) és NDX-GFP (genotípus: NDX- GFP / ndx1-1(FRI) / flc-2). A magokat sterilizáltuk és agar-MS lemezekre vetettük, majd 4 °C-on, sötétben tartottuk 2 napig. A lemezeket ezután 21 °C-os, hosszú nappalos (LD, 16 óra világos, 8 óra sötét) környezetbe helyeztük 10 napra. Minden méréshez 10 napos csíranövényeket használtunk.

Humán sejtkultúra

Az R-hurkokkal kapcsolatos kísérletekben használt Jurkat sejteket (humán lymphoblastoid leukémia) 10% (v/v) magzati borjúszérummal (FBS), 2 mM glutaminnal, penicillinnel és sztreptomicinnel kiegészített RPMI-1640 tápfolyadékban (Sigma, R5886) tenyésztettük 5% széndioxid koncentráció mellett.

Meiózis és sporuláció élesztőben

A szinkronizált sporuláció elindításához a diploid törzseket gazdag tápközegben (YPD) növesztettük 24 órán át, majd pre-sporulációs tápoldatba (SPS) oltottuk át őket át és éjszakán át növesztettük

~4x10^7 sejt/ml sűrűségig. Az SPS tenyészeteket centrifugálással összegyűjtöttük, majd 1% kálium-acetáttal mostuk és előmelegített sporulációs tápoldatban szuszpendáltuk 2x10^7 sejt/ml sűrűségig (SPM; nem-fermentálható szénforrásként 1% kálium-acetátot tartalmaz, aminosavakkal és nukleotidokkal kiegészítve az auxotróf markereknek megfelelően). A meiotikus DNS-replikáció és szinkronitás követése áramlási citometriás méréssel történt, amelyhez a meiotikus sejteket 70%-os etanolban fixáltuk, majd 1X PBS-ben végzett rehidratálás után a mintákat legalább 3 órán át RNázA-val (0,5 mg/ml) inkubáltuk 37 °C-on. A sejteket 50 µg/ml propidium-jodidban 1X PBS-ben reszuszpendáltuk 20 percig szobahőmérsékleten, majd újabb PBS-mosás után 5 µg/ml propidium-jodidban vettük fel. A mintákat pár másodpercig ultrahanggal kezeltük a sejtcsomók eltávolításához, majd a DNS-tartalmat FACSCalibur áramlási citométerrel (Becton Dickinson) határoztuk meg. A fertilitást (spóra életképesség) tetrád elemzéssel határoztuk meg az életképes spórák arányából.

Southern blot és DSB analízis

A meiotikus DSB-k kimutatása és kvantifikálása *sae2*Δ vagy *dmc1*Δ mutánsokban történt, amely megállítja a DNS törések processzállását. YPD tápközegben végzett vegetatív növekedés után a diploid sejteket pre-sporulációs (SPS) majd sporulációs táptalajba (SPM) vittük, és különböző időpontokban összegyűjtöttük. A genomi DNS-t fenol-kloroformmal extraháltuk, majd megfelelő restrikciós enzimekkel emésztettük és Southern blot-tal analizáltuk. A pulzáltatott erőterű gélelektroforézishez (PFGE) a genomiális DNS-t agarózdugókban készítettük elő. A DNS-t Hybond-N+ membránra (Amersham) blottoltuk, UV-sugárzással keresztkötöttük és hibridizáltuk a megfelelő radioaktív próbákkal. A PCR-által generált próbákat ^{32}P-dCTP-vel jelöltük Ready Prime kit (GE Healthcare) segítségével, és a radioaktív jeleket Storm Phosphorimager (Molecular Dynamics) vagy FLA5100 (Fuji) szkenner segítségével detektáltuk. A DSB frekvenciákat az ImageQuant szoftverrel (Molecular Dynamics) határoztuk meg.

Mer2/Spp1 koimmunprecipitáció

30 ml meiotikus élesztő sejtet gyűjtöttünk a jelzett időpontokban. A sejteket egyszer mostuk Tris-HCl pH8-ban és folyékony nitrogénben lefagyasztva -80 °C-on tároltuk későbbi felhasználáshoz. A pelletet 500 µl TMG50 lízis pufferben szuszpendáltuk (10 mM Tris-HCl pH 8, 1 mM MnCl2, 10% glicerin, 50 mM NaCl, 0,1 mM DTT) azonos térfogatú üveggyönggyel, és a sejteket Precellys® 24 bead-beater-rel (Bertin Technologies) tártuk fel. A lizátumokat 4 percig centrifugáltuk (13 000 ford./perc) 4 °C-on. Minden mintához 2,5 µl egér anti-myc monoklonális antitestet (9E10) és 2,5 µl Tween 20-at adtunk, 2 órán át 4 °C-on, forgatva. Ezután 10 µl protein G mágneses gyöngyöt (Invitrogen) adtunk a mintákhoz és további 2 órán át 4°C-on rotálva inkubáltuk. A gyöngyöket

háromszor mostuk TMG50+0,5% Tween20 pufferrel, egyszer pedig TMG50-el. A fehérjéket 20 µl vízben eluáltuk. A western blot-hoz 2x Laemmli puffer-ben vettük fel a mintákat, a fehérjéket 5 percig 95 °C-on denaturáltuk, majd 10% SDS-poliakrilamid gélen megfuttattuk. A gélt elektroblottoltuk, a membránokat pedig először anti-myc antitesttel inkubáltuk, hogy ellenőrizzük az IP hatékonyságát, majd anti-HA antitesttel (12CA5, Roche). Az IP-minták foszfatáz kezelését a következőképpen végeztük: az IP-ket 30 µl 1 mM $MnCl_2$-ot tartalmazó foszfatáz pufferben szuszpendáltuk és 800 U λ foszfatázzal emésztettük (New England Biolabs) 1 órán át 30 °C-on. A felülúszókat anti-myc-hez kapcsolt mágneses gyöngyökkel elválasztottuk, az IP-ket és a felülúszókat western blot-tal elemeztük. Az IP-minták DNázI kezelését hasonló protokollal végeztük, kivéve, hogy 200U DNáz I-et használtunk (Roche).

Western blot

Az Spp1-3xHA, GFP-Spp1, Spp1-myc, Rap1, H3K4me3, és béta-aktin fehérjék kimutatása élesztő teljes sejt-kivonatokban a következőképpen történt: a TCA-val kicsapott fehérje kivonatot SDS-PAGE-val megfuttatuk, nitrocellulóz membránra transzferáltuk majd a membránokat 5% tej/PBS-ben blokkoltuk. Anti-HA (12CA5), anti-myc (9E10), anti-GFP (Abcam), anti-Rap1 (V. Géli lab), anti-H3K4me3 (Upstate), és anti-béta-aktin (Abcam) antitesteket használva a membránokat 3 órán át 5%-os tej/PBS-ben inkubáltunk. Az infravörös fluorokrómmal jelzett másodlagos antitesteket (IRDye 800 anti-nyúl kecske IgG-t (Rockland) és az Alexa Fluor 680 anti-egér kecske IgG-t (Molecular Probes) 1:7500 hígításban használtuk. A másodlagos antitestek fluoreszcencia emisszióját Odyssey Imager (LI-COR Bioscience) szkennerrel detektáltuk.

Élesztő 2-hibrid screen

Az *Saccharomyces cerevisiae* teljes-genom könyvtár ("ORFome") szűrését yeast 2-hibrid screening rendszerrel végeztük a Hybrigenics SA (Párizs) cég segítségével. Csaliként pB27 vektorba klónozott *SPP1* ORF-et használtunk (lexA, C-terminális fúzió).

Rekombináns fehérje expresszió és GST-pull down

A megfelelő Spp1 plazmidokkal transzformált *E. coli* BL21 sejteket 2xYT-ben tenyésztettük 25 °C-on 0,1 OD600 értékig. 0,5 mM IPTG-vel való indukció után az inkubálást 3 órán át folytattuk. A GST-t és a GST-Mer2 fúziós fehérjét a bakteriális lizátumokból Glutathione Sepharose 4B-n tisztítottuk az ajánlásoknak megfelelően (GE Healthcare). Minden lépést 4°C-on vagy jégen végeztünk. A fehérjéket eluáló pufferrel (10 mM redukált glutathione, 50 mM Tris-Cl pH 8,0, 0,01% NP-40 és 10% glicerin) eluáltuk, és -80 °C-on tároltuk. A teljes hosszúságú Spp1 fehérje és trunkált Spp1-fragmentumok *in*

vitro expressziójához a plazmid DNS-eket TNT-kapcsolt *in vitro* transzlációs rendszerrel (Promega) írtuk át mRNS-é és fehérjévé, összesen 50 ul térfogatban, a gyártó utasításai szerint. A tisztított GST-t vagy a GST-Mer2 fúziós fehérjét (0,5-1,0 µg) az *in vitro* transzlált [^{35}S]-metionin-jelölt fehérjékkel inkubáltunk. A fehérjefoszforiláció *in vitro* kötődési hatékonyságra gyakorolt hatásának tesztelésére a reakciókat 10 egység borjúbél-foszfatázzal egészítettük ki (CIP; New England Biolabs). A kötőpuffer 1 ml-es aliquotját 15 µl glutation-sepharose-val (GE Healthcare) inkubáltuk 1 órán át, szobahőmérsékleten, majd a gyantát ülepítettük és háromszor mostuk 1 ml 0,01% NP-40-et tartalmazó PBS-sel. A fehérjéket 40 µl NuPAGE minta pufferrel eluáltuk és 4-12%-os Bis-Tris NuPAGE gélen (Invitrogen) megfuttattuk. A géleket 30 percig fixáló oldatban (10% ecetsav, 40% metanol) fixáltuk, majd 30 percig 1 M nátrium-szalicilát oldattal kezeltük és vákuumban 3M whatman papíron szárítottuk. A szárított géleket Fujifilm filmen 24-72 órán keresztül exponáltuk és BAS-1800II képelemző készüléken szkenneltük.

Kromatin immunprecipitáció és kompetíciós kromatin immunprecipitáció (ChIP és c-ChIP) szekvenálás

A meiotikus élesztősejteket (4×10^7 sejt/ml) 50 ml térfogatban gyűjtöttük össze az ábrákon jelzett időpontokban és 1% formaldehiddel 20 percig szobahőmérsékleten fixáltuk. A reakciót 125 mM glicinnel, szobahőmérsékleten történő inkubálással állítottuk le 5 perc alatt, majd a sejteket háromszor jéghideg 1x TBS, pH 7,5 (20 mM Tris-HCl, pH 7,5, 150 mM NaCl) oldatban mostuk. *Arabidopsis* csíranövények (GFP-NDX) esetén magát a növényi mintát fixáltuk a fenti módon, majd folyékony nitrogén-tartalmú dörzsmozsárban mechanikusan homogenizáltuk a sejteket. Ezt követően a sejteket 500 µl lízis pufferben (50 mM Hepes, KOH pH 7,5, 140 mM NaCl, 1 mM EDTA, 1 % Triton X-100, 0,1 % Na-deoxikolát, 1 tabletta teljes protein inhibitor koktél (Roche)), savval mosott üveggyöngyökkel lizáltuk 10 percig FastPrep bead beater készülék segítségével. A kromatin mintákat ultrahanggal (Bioruptor, Diagenode) fragmentáltuk átlagosan 300 bp méretűre. A teljes sejtextraktumból kivettünk 50 µl immunprecipitáció előtti mintát (input), és teljes sebességgel 10 másodpercig centrifugáltuk a sejttörmelék elválasztásához (felülúszó: teljes sejtextraktum). A maradék mintát 20 percig centrifugáltuk (12000 fordulat/perc, 4 °C) a sejttörmelék elválasztásához. Az immunoprecipitációt (IP) úgy hajtottuk végre, hogy 450 µl extraktumot adtunk a mágneses protein G dynabeads (Dynal) pellethez, amelyből 50 µl 2×10^7 gyöngynek felelt meg. A gyöngyöket előinkubáltuk a megfelelő antitestekkel, amely “hagyományos” ChIP esetén 9E11 monoklonális egér anti-myc antitest volt (ab56, Abcam), c-ChIP esetén pedig 9E11 anti-myc antitest (ab56, Abcam) a konstitutív allél ellen és anti-GFP poliklonális nyúl antitest (ab290, Abcam) a kompetíciós (indukált)

allél ellen, éjszakán át 4 °C-on. Egyéb yeast ChIP kísérletekben anti-GAL4BD (Euromedex) és anti-H3K4me3 (Abcam) antitesteket is használtunk. Az *Arabidopsis* ChIP mérésekhez anti-flag (#2044, New England Biolabs) és anti-GFP (#ab290 Abcam) antitesteket használtunk. Az IP mintákat kétszer mostuk lízis pufferrel, majd újabb kétszer 360 mM NaCl-dal kiegészített lízispufferrel, kétszer mosó pufferrel (10 mM Tris HCl pH 8,0, 250 mM LiCl, 0,5 % NP-40, 0,5 % Na-deoxikolát, 1 mM EDTA), és végül egyszer 1x TE-vel, pH 7,5, Dynal mágneses szeparátor segítségével. A keresztkötéseket TE/1% SDS oldatban oldottuk fel, éjszakán át, 65 °C-on. Ezután a fehérjéket proteináz K-val emésztettük 3 órán át 65 °C-on. A nukleinsavat PCR clean-up oszlopon tisztítottuk, majd RNS-emésztést (10 ug RNáz) végeztünk 1 órán át 37 °C-on. A ChIP DNS-t végül 50 µl nukleáz-mentes dH2O-ban oldottuk fel. Az NGS könyvtárakat az Illumina TruSeq ChIP mintaelőkészítési protokollja szerint állítottuk elő. A könyvtárakat 150 bázispáros paired-end read-ekre szekvenáltuk Illumina HiSeq 2500 szekvenátorral (EMBL Genomics Core Facility, Heidelberg, Németország). A nyers fastq file-ok feldolgozása és a peak meghatározás a DRIP-seq módszerhez hasonlóan történt. A publikált ChIP-seq adatok elérhetőek a JBrowse genomböngészőben: http://geneart.med.unideb.hu/pub/2018-spp1, login: spp1; password: cerevisiae7. A nyers NGS adatok a GEO-n (Gene Expression Omnibus) találhatók: GSE107967.

Kromatin immunprecipitáció (ChIP) microarray (ChIP-on-Chip)

Néhány élesztő ChIP mérést még az "NGS korszak" előtt végeztünk DNS-chip technológia segítségével. Ezeket a mintákat Agilent 4x44k Yeast teljes genom microarray-khez hibridizáltuk (Agilent), amely 44,000 próbával a teljes élesztő genomot lefedte (az rDNS klaszter kivételével). Az immunprecipitációt 3 µl nyúl poliklonális anti-Rfa1 antitesttel (Steven Brill), 3 µl nyúl poliklonális anti-H3K56ac antitesttel (Alain Verreault, Montreal), és 3 µl nyúl poliklonális anti-H3Cter antitesttel végeztük (Abcam1791). Az input mintákat Cy3 festékkel, az IP mintákat pedig Cy5 festékkel jelöltük és egyenlő arányban hibridizáltuk a csipek felszínére. A mosásokat, a szkennelést, az adatok normalizálását és az elemzést az Agilent cég gyári protokollja szerint végeztük. A hiszton H3K56ac adatokat (log2 arányok) a meiózis során a korábban publikált hiszton H3 kötődésre normalizáltuk (GSM274190 és GSM274191). Minden microarray eredmény hozzáférhető JBrowse-ban a következő url-en: http://geneart.med.unideb.hu/pub/h3k56ac, login: h3k56ac; jelszó: Mozaic4. A nyers adatok a GEO-n (GSE37487) érhetők el.

Kromatin turnover ráta becslése c-ChIP adatokból

Az Spp1 és Set1 genomi kötőhelyek turnover rátáját a GFP-Spp1 és a Spp1-myc kompetíciós ChIP-seq adatok alapján határoztuk meg. Kiszámítottuk a GFP-Spp1 / MYC-Spp1 "okkupancia" arányokat

minden egyes időpontban és minden genomi kötőhelyre (peak-re), majd exponenciális modellt illesztettünk az adatokra: $\frac{GFP}{myc}$= (1-e$^{-\lambda t}$), ahol a *GFP/myc* az "okkupancia" arány, λ a turnover ráta, t pedig a kompetíciós allél indukciójától számított idő. A modell illesztése után kiszámítottuk a becslések standard hibáját és t-próbát végeztünk a modell illeszkedésének jóságának becslésére. Összesen 977 c-ChIP kötőhelyet és turnover rátát írtunk le a modellel.

Meiotikus transzkriptom microarray

20 ml sporuláló (SPM) élesztő tenyészetekből Trizol reagenssel totál RNS-t izoláltunk a sporuláció kezdőpontjában (0h) és az azt követő időpontokban (2h, 4h, 6h). A cDNS szintézis során aminoallyl-dUTP-t (Fermentas) építettünk be a mintákba. A 0h mintát referenciának tekintettük és Cy3 festékkel jelöltük (Amersham), míg a 2-6h mintákat Cy5 festékhez kapcsoltuk. A Cy3-mal jelölt referencia cDNS-t azonos mennyiségben hibridizáltuk a Cy5-jelzett mintákkal 8x15K transzkriptom microarray-khez (Agilent). A mosásokat, a szkennelést, az adatok normalizálását és az elemzést a gyártó protokollja szerint végeztük. Az eredmény file a 0h időponthoz képest mért relatív mRNS expressziós változásokat mutatja.

RNS-DNS hibrid immunprecipitáció (DRIP)

Humán sejtek keresztkötéséhez 1%-os paraformaldehid oldatot használtunk 10 percig, aminek a hatását 2.5 M glicinnel szüntettünk meg 5 perc alatt szobahőmérsékleten. *Arabidopsis* csíranövények (Col-0 és *ndx1-4* minták) esetén magát a növényi mintát fixáltuk a fenti módon, majd folyékony nitrogén-tartalmú dörzsmozsárban mechanikusan homogenizáltuk a mintát. Az előkészített preparátumokat 1 ml lízis pufferben (500 µl 2x lízis és 500 µl TE) lizáltuk. Humán minták esetén a sejtlízist két különböző hőmérsékleten is elvégeztük: 65 °C 7 óra vagy 37 °C, éjszakán át. A nukleinsav izolálást NucleoSpin Tissue Kit segítségével végeztük és 100 µl elúciós pufferben eluáltuk a mintákat. A tisztított nukleinsav preparátumot 300 µl Tris-HCl (pH 8.5) oldatban fragmentáltuk szonikálással kétszer 5 percig, hogy átlagosan 500 bázispár méretű DNS fragmenteket kapjunk. A fragmentum analízist 1%-os agaróz gélelektroforézissel végeztük el. Amennyiben szükséges volt, további szonikálást alkalmaztunk. A szonikált DNS mintákat NucleoSpin Gél és PCR tisztító kittel tisztítottuk, majd 100 µl elúciós pufferben eluáltuk. 12 µg DNS-t 100 µl térfogatra hígítottunk 5 mM Tris-HCl (pH 8.5) oldattal. A minták 2%-át input DNS-ként eltettük. A minták felét 80 µl végtérfogatban 8 µl *E. coli* RNase H-val kezeltük (NEB) 37 °C-on, éjszakán át. Az immunszelekcióhoz Dynabeads Protein A mágneses gyöngyöket PBS/EDTA tartalmú 0.5%-os BSA-val blokkoltunk. Az S9.6 antitest immobilizációhoz 50 µl blokkolt gyöngyöt 10 µg S9.6 antitesttel inkubáltunk folyamatos forgatás

közben IP pufferben 4 °C-on 4 órán át. 6 µg fragmentált genomi DNS-t adtunk a keverékhez és gyengéden forgattuk egy éjszakán át 4 °C-on. A gyöngyöket ülepítettük és mostuk egymás után 1 ml lízis pufferrel (alacsony só cc. mellett), 1 ml lízis pufferrel (magas só cc. mellett), 1 ml mosó pufferrel és 1 ml TE-vel 4 °C-on, kétszer. Az elúció 100 µl elúciós pufferrel történt 15 percen keresztül 65 °C-on. NucleoSpin Gél tisztítás és PCR Clean-up kit után, a nukleinsavakat 55 µl elúciós pufferben eluáltuk. A kinyert DNS-t kvantitatív valós-idejű PCR készülékkel (LightCycler 480, SYBR Green I Master) mértük le és QuantStudio 12K Flex valós-idejű PCR rendszerrel elemeztük ki. Az adatokat a relatív dCT módszerrel analizáltuk. Az RNS-DNS hibrid feldúsulást az IP/Input arányból számoltuk ki.

RNS-DNS hibrid szekvenálás (DRIP-seq)

Az NGS könyvtárakat az Illumina TruSeq ChIP protokoll alapján készítettük el. Röviden, az immunprecipitáció során feldúsult DRIP molekulák végét reparáltuk és az inzertekhez index adaptereket ligáltunk. A tisztított ligációs termékeket PCR segítségével amplifikáltuk, majd szekvenáltuk. A leolvasott szekvenciákat az emberi (hg19) vagy *Arabidopsis thaliana* (tair10) referencia genomra illesztettük BWA-MEM algoritmussal alapbeállítások mellett. Kiszűrtük azokat a leolvasott szekvenciákat, amelyek alacsony térképeződési értékkel rendelkeznek, illetve PCR duplikátumok voltak vagy feketelistás régiókon lokalizálódtak. A replikákat összevontuk és MACS2-t használtunk a szignifikáns genomi régiók azonosításához input-normalizálás mellett. A feldolgozott és összevont illesztésekből bamCoverage alkalmazásával szignál fájlokat generáltunk. Az RPKM értékeket 20 bázispáros ablakokban határoztuk meg minden egyes mintára egy 60 bázispáros csúszóablak alkalmazásával. Az elkészített szignál fájlokat R-ben ábrázoltuk ggplot2 és ggbio csomagok segítségével. A DRIP-seq adatok elérhetők JBRowse-ban (http://geneart.med.unideb.hu/pub/2021-ndx, user: ndx, pw: athaliana21), illetve az SRA (Sequence Read Archive) adatbázisban (https://www.ncbi.nlm.nih. gov/sra, #SRP095885).

S9.6 slot blot

A genomi DNS-t 10 napos Col-0 és ndx1-4 csíranövényekből tisztítottuk fenol/kloroform extrakcióval. Humán minták esetén 1 millió Jurkat sejtből izoláltunk hasonló módszerrel. 50, 100 és 200 ng gDNS prep-et csöppentettünk nitrocellulóz membránokra (Hybond-N+, GE Amersham) - RNaseH kezeléssel vagy anélkül. A transzfert a Bio-Dot SF Microfiltration Apparatus segítségével végeztük el gyári beállítások szerint. A membránokat UV-vel (2000) kétszer keresztkötöttük, majd tej/TTBS pufferben blokkoltuk (5% tej 1X TTBS-ben; 17 mM Tris, 130 mM NaCl, PH7,5, 1% Triton-

X100) 1 órán át, szobahőmérsékleten. Az S9.6 antitestet tej/TTBS pufferben 1:1500 arányban hígítottuk és a membránokat éjszakán át 4 °C-on inkubáltuk. A membránokat tej/TTBS-ben mostuk háromszor, minden alkalommal 10 percig. A másodlagos antitestet (kecske anti-egér-HRP) 1 órán át szobahőmérsékleten alkalmaztuk, majd tej/TTBS-ben 5 percig és 1X TTBS-ben háromszor 10 percig mostuk. Az S9.6 jelet ECL reagens segítségével detektáltuk. Loading kontrolnak a membránok metilénkék festését használtuk, amely alapján ImageJ-vel kvantifikáltuk a felvitt nukleinsav mennyiségét.

A DRIP és ChIP peak-ek genomi annotációja

A ChIP-seq és DRIP-seq adatok elemzése során az NGS szekvenálási adatok feldolgozásából kapott .bed fájlokkal dolgoztunk. A DRIP és ChIP peak-ek genomi eloszlásának a meghatározásához a GenomicRanges R csomagot alkalmaztuk, amellyel a DRIP/ChIP régiók és az annotációs kategóriák átfedő területeit meghatároztuk. Az átfedő területeket számítógép-generált véletlen átfedésekhez hasonlítottuk. A Bedtools szoftver segítségével a fájlokban tárolt kötőhely pozíciókat véletlenszerűen megváltoztattuk a kromoszómákon belül, ezzel az eredeti fájlokkal megegyező peak számot és méretet, de eltérő pozíciókat kaptunk. A szignifikáns dúsulást mutató helyeket a Jbrowse genomböngésző programmal vizualizáltuk.

KisRNS szekvenálás (sRNS-seq)

Arabidopsis Col-0 és *ndx1-4* palántákból totál RNS-t tisztítottuk a northern blot-nál leírt módszerrel. Az RNS-minták minőségét Agilent bioanalizátorral ellenőriztük (RIN > 9). Az NGS könyvtárakat az Illumina NEBNext® Multiplex Small RNA Library Prep protokollja szerint készítettük és Illumina NextSeq500 szekvenátorral szekvenáltuk 1x50 bp leolvasással. Az eredményeket az sRNAnalyzer pipeline segítségével elemeztük (Wu et al. 2017) az alábbiak szerint: az Illumina adaptereket Cutadapt segítségével levágtuk, majd a leolvasások méretét a 19-25 nt tartományba korlátoztuk. Az sRNAnalyzer segítségével a kiválasztott méretű read-eket a miRbase adatbázishoz (https://www.mirbase.org/) és egy nemrégiben közzétett átfogó kis RNS-lókusz-adatbázishoz (Hardcastle et al. 2018) annotáltuk. A Col-0 és *ndx1-4* minták közötti differenciális sRNS expressziót a Deseq2 algoritmussal határoztuk meg ($p<0,05$, abs(log2(fc)) >log2(1,5)) és az eredményeket R-ben ábrázoltuk. A targetkereséshez a psRNATarget algoritmust alkalmaztuk az alapértelmezett beállításokkal (Dai et al. 2018). Az sRNS-seq adatokat sRNS rt-qPCR-rel validáltuk az alábbiak szerint: 5 µg totál RNS-t DNázI-el (NEB, #M0303) kezeltünk a gyártó utasításai szerint, majd a mintákat etanolban kicsaptuk és steril vízben felszuszpendáltuk. Egy mikrogramm DNázI-kezelt totál RNS-t és sRNS- vagy U6 referencia RNS-specifikus PCR primereket használtunk a cDNS reakcióhoz (NEB,

#E6560) a gyártó utasításai szerint. A qPCR-eket qPCR Master Mix-ben (NEB, #M3003) végeztük LightCycler® 96 Real-Time PCR gép (Roche) segítségével. Az adatokat LightCycler® 96 szoftverrel (1.1.0.1320-as verzió) elemeztük.

In situ Hi-C

A Hi-C-t 10 napos *Arabidopsis thaliana* csíranövényeken végeztük Col-0 és *ndx1-4* mutáns genomi háttérben. A növényi mintát 1%-os formaldehiddel fixáltuk, majd folyékony nitrogén-tartalmú dörzsmozsárban mechanikusan homogenizáltuk. A sejtmag izoláláshoz Miracloth szűrőt használtunk az Arima Genomics protokollja alapján. A Hi-C kísérlet kivitelezéséhez Arima-HiC Kit-et használtunk (https://arimagenomics.com/kit, Arima Genomics Ltd.). A kromatin feldarabolását DpnII és HinfI restrikciós enzimekkel végeztük, a végreparálást és biotinilálást a kit reagenseivel végeztük a gyártó ajánlása szerint. A keresztkötött/biotinilált fragmentumokat a permeabilizált sejtmagokon belül (*in situ*) ligáltuk, majd szonikálás után NGS könyvtárat készítettünk az Illumina TrueSeq protokoll alapján. A paired-end NGS szekvenálás során mintánként közel 200 millió read-et kaptunk, amely a jelenlegi legjobb felbontás az *Arabidopsis thaliana* 130 megabázis méretű genomjáról. A leolvasásokat ún. "kiméra read"-ként kaptuk vissza, mivel a read-ek eltérő genomi szakaszokról származtak. Az interakciós fragmentumokat a TAIR10 referencia genomra térképeztünk a Juicer pipeline-al (Durand et al. 2016). A Juicer először a nyers .fastq adatokat transzformálta Hi-C interakciós adatokká, majd ezeket normalizálta és interakciós mátrixokká alakította és annotálta. A Juicer ezenkívül annotálta a DNS hurkokat, loop anchor motívumokat, és kontakt doméneket, melyeket a HiCCUPS algoritmus (Rao et al. 2014) segítségével azonosított. A .hic fájlokból származó térbeli interakciókat .bed fájlokba konvertáltuk 1 kilobázis (kB), 5kb, 10kb és 25kb nagyságú genomi felbontásokkal.

In silico restrikciós enzim emésztés

Adott restrikciós enzim-kombinációk ("RE koktélok") által generált elméleti fragmenthossz eloszlás kiszámításához az emberi, Arabidopsis, és élesztő genomot *in silico* feldaraboltuk a DECIPHER R csomag segítségével. A hasítóhely pozíciók alapján kiszámoltuk a restrikciós fragmentumok hosszát és eloszlását. A keletkezett fragmenthossz-eloszlások statisztikai összehasonlításához 300 véletlenszerűen kiválasztott értéket hasonlítottunk össze Wilcoxon Rank Sum teszttel, 100 alkalommal. A p-értékeket Benjamini & Hochberg módszerrel korrigáltuk.

Konfokális lézerpásztázó mikroszkópia (CLSM)

A humán sejteket minden mérés előtt 1%-os formaldehiddel fixáltuk. A CLSM képeket Olympus FluoView 1000 konfokális mikroszkóppal vettük fel, amelyhez 60x olajimmerziós objektívet (NA 1,35) használtunk. A gerjesztési és emissziós szűrők a következő volt: EGFP, 488 nm gerjesztés, 500-540 nm detektálás. Tíz, 0,7-1,1 µm vastagságú optikai szeletet gyűjtöttünk minden egyes mag esetében, Kalman szűrőmódot alkalmazva a zaj és alternatív gerjesztés csökkentése érdekében és a cross-talk kizárására. A kolokalizáció mértékét az ImageJ program JACoP plugin-jével számoltuk ki.

Fluoreszcencia-visszatérés fotoelhalványodás után (FRAP)

A FRAP méréseket GFP-Spp1 és GFP-Set1 élesztő sejteken végeztük Olympus FluoView 1000 IX-81 invertált konfokális mikroszkóppal, UPlanAPO 60x (NA 1.2) vízimmerziós objektívvel. Az GFP-t 488 nm-en gerjesztettük argon-ion lézerrel és a fluoreszcenciát egy 500-550 nm-es sávszűrőn detektáltuk. A FRAP mérésekben kioltási régióknak négyzet alakú területeket választottunk a magon belül. A sporuláló élesztősejtekben elvégzett FRAP kísérletekhez (0-6 óra SPM-ben) óránként vettünk mintákat és a méréseket 1% kálium-acetáttal bevont mikroszkópos tárgylemezeken végeztük el. A Set1-GFP vagy Spp1-GFP fehérjéket expresszáló sejteket véletlenszerűen választottuk ki 100 µM CuSO4 indukciót követően, majd öt előképet (256 × 256 képpontos terület, 15x zoom, ~ 9 µW lézer teljesítmény) készítettünk és 500 ms-os kioltási idő mellett, 100 %-os lézer teljesítménnyel (900 µW) végeztük el a sejtmag fotokioltását. A képeket másodpercenként gyűjtöttük 1 percen át. Az *Arabidopsis thaliana*-ban végzett FRAP mérések esetén az NDX-GFP jelet 10 napos csíranövények gyökércsúcsában mértük a fenti paraméterekkel azonos beállításokkal, 25 másodpercig.

Fluoreszcencia-korrelációs spektroszkópia (FCS)

A GFP-Set1 és GFP-Spp1 FCS kísérletekhez Olympus FluoView1000 konfokális mikroszkópot használtunk. Meiotikus élesztő sejteket mintavételeztünk minden órában folyékony sporulációs tenyészetekből, majd a sejteket 1% kálium-acetát agarral (SPM) bevont mikroszkópos tárgylemezekre cseppentettük. Mérés előtt a GFP-Set1 és GFP-Spp1 expressziót 100 µM $CuSO_4$ oldattal indukáltuk. Az autokorrelációs görbéket ALV-5000E korrelációs kártyával számoltuk ki az egyes sejtmagokban, három véletlenszerűen kiválasztott pontban, 10 × 8 másodperces mérést követően. Gazdag táptalajon (YPD) növesztett mitotikus tenyészetek esetében a sejteket A260 = 1,0 optikai denzitás értékig növesztettük, majd YPD agarral bevont tárgylemezen, 100 µM $CuSO_4$ indukciót követően mértük. Minden mérés szobahőmérsékleten (22 °C) történt. Az autokorrelációs függvényeket és a származtatott FCS paramétereket a QuickFit 3.0 szoftverrel számítottuk ki

(Krieger, Jan; http://www.dkfz.de/Macromol/quickfit/). Az *Arabidopsis thaliana*-ban végzett FCS mérések esetén az NDX-GFP jelet 10 napos csíranövények gyökércsúcsában mértük a fenti paraméterekkel azonos beállításokkal.

Farmakogenomikai elemzések

Összesen 36 R-hurok regulátor fehérjét vizsgáltunk (**2. táblázat**) 33 primer tumor típusban (**3. táblázat**), amelyek génexpressziós és túlélés adatait a TCGA adatbázisból töltöttük le (The Cancer Genome Atlas Research Network et al. 2013). A túléléselemzés Kaplan-Meier görbéit a "survival" szoftverrel generáltuk R-ben. Az R-hurok regulátorok génexpresszió-függő gyógyszer-szenzitivitás asszociációit 276 FDA által engedélyezett szerre számítottuk ki a GDSC adatbázisból letölthető tumorsejtvonalak adatai alapján (Yang et al. 2013). Az asszociációkat az IC_{50} értékek (félhalálos dózis) és az adott R-hurok regulátor gén medián mRNS-seq értékek Spearman rangkorrelációja alapján ábrázoltunk. A statisztikai számításoknál a p-értékeket a Benjamini-Hochberg módszerrel korrigáltuk.

2. Táblázat A farmakogenomikai elemzésekben vizsgált R-hurok regulátor gének listája

Gén	Ensemble ID	Funkció
BRCA1	ENSG00000012048	DNS hibajavítás
BRCA2	ENSG00000139618	DNS hibajavítás
FANCD2	ENSG00000144554	DNS hibajavítás
FANCM	ENSG00000187790	DNS hibajavítás
GADD45A	ENSG00000116717	DNS hibajavítás
TOP1	ENSG00000198900	DNA topológia
TOP3B	ENSG00000100038	DNA topológia
AQR	ENSG00000021776	helikáz
ATXN1	ENSG00000124788	helikáz
ATXN2	ENSG00000204842	helikáz
BLM	ENSG00000197299	helikáz
DDX19, DDX19A	ENSG00000168872	helikáz
DHX9	ENSG00000135829	helikáz
PIF1	ENSG00000140451	helikáz
RTEL1	ENSG00000258366	helikáz
SETX	ENSG00000107290	helikáz
CARM1	ENSG00000142453	hiszton arginin metiláz
PRMT1	ENSG00000126457	hiszton arginin metiláz
BUB3	ENSG00000154473	mitózis regulator
BUGZ, ZNF207	ENSG00000010244	mitózis regulátor
RNASEH1	ENSG00000023191	ribonukleáz
RNASEH2A	ENSG00000104889	ribonukleáz
RNASEH2B	ENSG00000136104	ribonukleáz
RNASEH2C	ENSG00000172922	ribonukleáz
EWSR1	ENSG00000182944	RNS-kötő fehérje
SRSF1	ENSG00000136450	RNS splicing
SRSF2	ENSG00000161547	RNS splicing
THOC1	ENSG00000079134	RNS splicing
THOC2	ENSG00000125676	RNS splicing
THOC3	ENSG00000051596	RNS splicing
THOC4	ENSG00000183684	RNS splicing
THOC5	ENSG00000100296	RNS splicing
THOC6	ENSG00000131652	RNS splicing
THOC7	ENSG00000163634	RNS splicing
TREX1	ENSG00000213689	RNS splicing
U2AF1	ENSG00000160201	RNS splicing

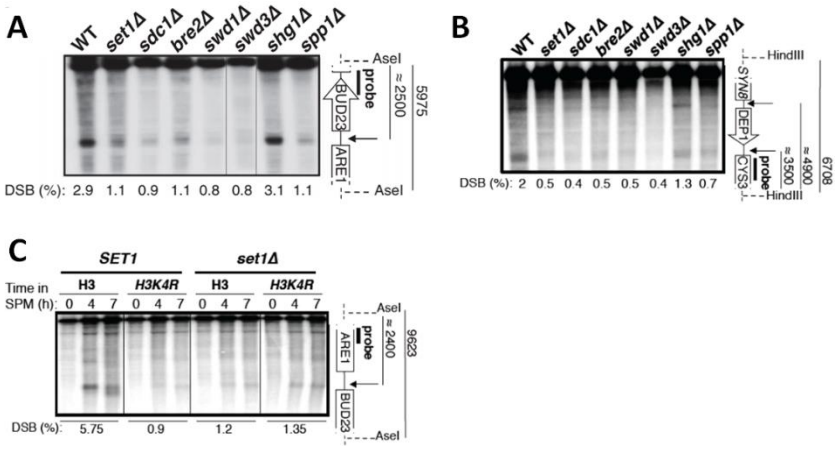
3. Táblázat A farmakogenomikai elemzésekben vizsgált TCGA tumor típusok listája

Azonosító	Tumor típus
LAML	Acute Myeloid Leukemia
ACC	Adrenocortical carcinoma
BLCA	Bladder Urothelial Carcinoma
LGG	Brain Lower Grade Glioma
BRCA	Breast invasive carcinoma
CESC	Cervical squamous cell carcinoma and endocervical adenocarcinoma
CHOL	Cholangiocarcinoma
LCML	Chronic Myelogenous Leukemia
COAD	Colon adenocarcinoma
ESCA	Esophageal carcinoma
FPPP	FFPE Pilot Phase II
GBM	Glioblastoma multiforme
HNSC	Head and Neck squamous cell carcinoma
KICH	Kidney Chromophobe
KIRC	Kidney renal clear cell carcinoma
KIRP	Kidney renal papillary cell carcinoma
LIHC	Liver hepatocellular carcinoma
LUAD	Lung adenocarcinoma
LUSC	Lung squamous cell carcinoma
DLBC	Lymphoid Neoplasm Diffuse Large B-cell Lymphoma
MESO	Mesothelioma
MISC	Miscellaneous
OV	Ovarian serous cystadenocarcinoma
PAAD	Pancreatic adenocarcinoma
PCPG	Pheochromocytoma and Paraganglioma
PRAD	Prostate adenocarcinoma
READ	Rectum adenocarcinoma
SARC	Sarcoma
SKCM	Skin Cutaneous Melanoma
STAD	Stomach adenocarcinoma
TGCT	Testicular Germ Cell Tumors
THYM	Thymoma
THCA	Thyroid carcinoma
UCS	Uterine Carcinosarcoma
UCEC	Uterine Corpus Endometrial Carcinoma
UVM	Uveal Melanoma

EREDMÉNYEK ÉS MEGBESZÉLÉSÜK

A Set1 komplex szerepe a meiotikus DNS törések kialakulásában

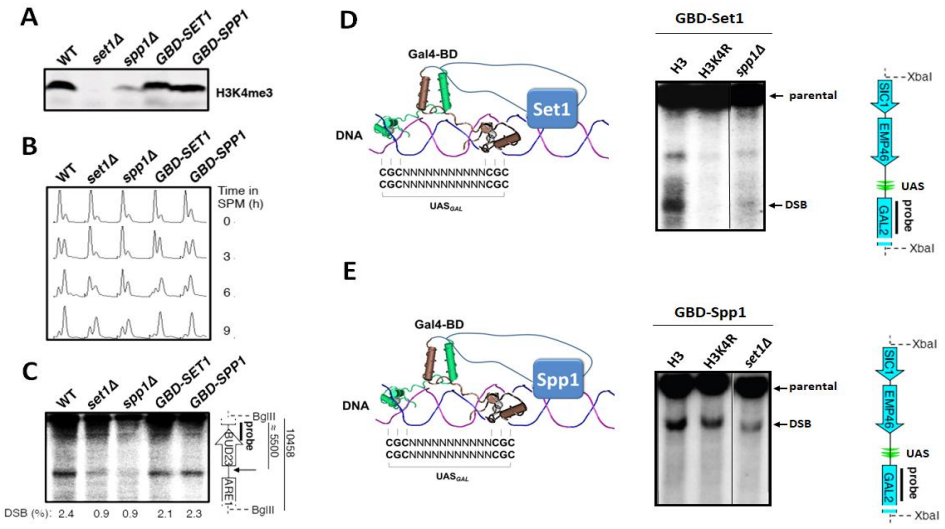
Ivarosan szaporodó fajokban a genetikai allélek kombinációi a szülői kromoszómák közötti homológ rekombináció révén keletkeznek a meiózis során. A kétezres évek közepétől tudjuk, hogy a folyamat jelentős epigenetikai kontroll alatt áll és bizonyos hiszton modifikációk és hisztonmódosító enzimek fontos szerepet játszanak a rekombináció elindításában (Székvölgyi and Nicolas 2010). A H3 lizin 4 trimetiláció és a Set1 komplex (Set1C/COMPASS) "megjelöli" a rekombinációs forrópontokat élesztőtől az emlősökig (Sollier et al. 2004; Borde et al. 2009; Parvanov et al. 2010), azonban azt nem ismertük, hogy ez a rendkívül abundáns hisztonmódosítás hogyan lép kapcsolatba a DSB hotspot régiókkal és a DNS duplaszáltöréseket (DSB-ket) okozó Spo11 komplexxel. Ennek tisztázására olyan Set1C-mutánsokkal végeztünk funkcionális vizsgálatokat, amelyekben egyesével inaktiváltuk a komplex összes alegységét és megvizsgáltuk, hogy bizonyos jól karakterizált rekombinációs forrópontokban megváltozik-e a DNS törések gyakorisága és eloszlása. Southern blot kísérleteink szerint a *BUD23-ARE1* és *DEP1-CYS3* hotspot régiókban minden Set1C-mutánsban (Set1, Sdc1, Bre2, Swd1, Swd3, Spp1) szignifikánsan csökkent a meiotikus DSB-k mennyisége (az Shg1 alegység kivételével) (**14/A-B ábra**), és a DSB-szint csökkenése a módosítható lizin 4 oldalláncától és így a H3K4 metilációtól függött (**14/C ábra, bal panel**).



14. ábra Meiotikus DSB-k vizsgálata Set1C mutánsokban. A Southern blot méréseket az ábrán jelzett deléciós mutánsokban vagy hiszton pontmutánsokban vagy kettős mutánsokban végeztük PCR-ral előállított végjelölt radioaktív próbákkal. Az A-B paneleken az 5 órás meiotikus időpont (5h SPM) látható, míg a C panelen a 0h, 4h, 7h meiotikus időpontokat mutatjuk. A Southern blot próbát a gél bal oldalán jelzett hotspot régió restrikciós fragmentumának legvégére terveztünk, amely két domináns band-et ad: egy hasítatlan parentális fragmentumot illetve egy elhasított DSB fragmentumot. A DSB band-et nyíllal jeleztük, kvantifikálását a gél alatti számok mutatják.

A H3K4R/*set1Δ* dupla-mutánsok hasonló DSB-fenotípust mutattak, mint az egyszeres H3K4R és *set1Δ* mutánsok (**14/C ábra, jobb panel**), amely a két allél recesszív episztatikus kölcsönhatására és közös biokémiai útvonalon történő szabályozásukra utal. Ezen megfigyelések a Set1C és a H3K4 metiláció funkcionális szerepét jelzik a DSB képződés szabályozásában.

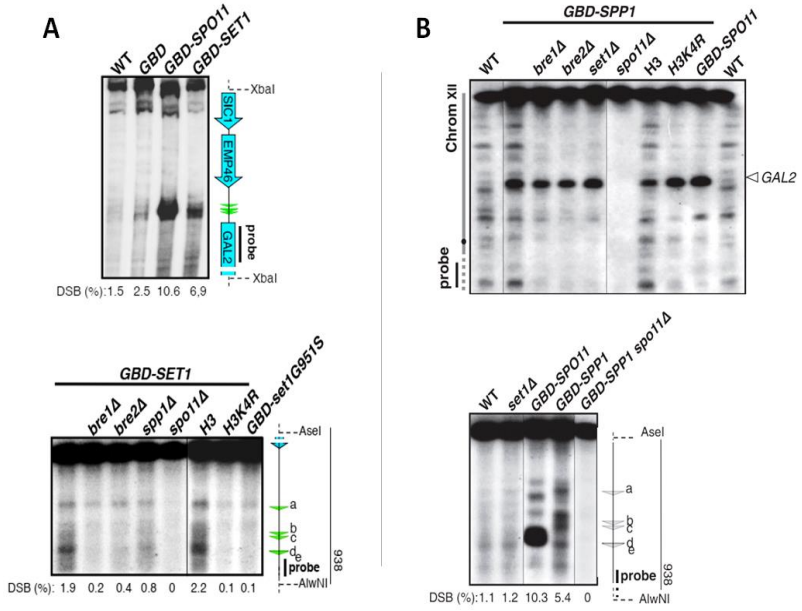
A H3K4 trimetiláció és a DNS törések mechanikai kapcsoltságának bizonyításához egy olyan kísérleti rendszert terveztünk, amelyben a kromoszómák előre meghatározott pontjain (célzottan) elhelyeztünk hiszton H3K4me3 módosításokat, majd ugyanitt detektáltuk a DNS töréseket. Ez a funkcionális megközelítés a CRISPR korszak előtt kis túlzással "forradalminak" számított, amellyel azt kívántuk bizonyítani, hogy amikor elhelyezzük a kromoszómán a hiszton-jelet, akkor valóban a várt pozícióban törik el a DNS és nem máshol, ok-okozati kapcsolatot igazolva a két változó között a korábbi statisztikai összefüggés helyett. Ehhez egy olyan géntechnológiai eszközt fejlesztettünk, amely egyrészt felelős a vizsgált H3K4me3 kialakulásáért, másrészt szelektíven odairányítható a DNS molekula előre megadott szakaszaihoz egy Gal4 DNS kötő domént (GBD) tartalmazó „célzó" egység segítségével. A vizsgált genomi régiók jól karakterizált "hideg" (coldspot) szakaszok voltak, amelyek normál körülmények között nem vesznek részt rekombinációs eseményekben, viszont UAS_{GAL4} motívumokat tartalmaztak a konstrukt megkötéséhez. A GBD-Set1 és GBD-Spp1 törzsek minden tekintetben úgy viselkedtek, mint a vad típusú sejtek (H3K4me3 mennyisége, meiotikus progresszió (FACS profil), DSB képződés a *BUD23-ARE1* hotspot-ban; **15/A-C ábra**), kivéve, hogy a *GAL2* coldspot régióban, ahol vad típusú sejtekben egyáltalán nincs rekombináció (**16/A-B ábra**), a fúziós fehérjék DNS töréseket váltottak ki (**15/D-E ábra, első oszlop;**), így bizonyítva az ok-okozati összefüggést a H3K4me3 jel és a rekombináció iniciációja és között.



15. ábra A GBD-Set1 és GBD-Spp1 rekombinációs „hideg" kromatin területre irányítása. *A-C.* Mindkét fehérje normál (wt) H3K4me3-t, FACS profilt és DSB fenotípust mutat a *BUD23-ARE1* forrópontban. *D-E.* A *GAL2* coldspot-ban mindkét konstrukt DSB-t indukál, amely a GBS-Set1 esetén függ a H3K4 oldalláncától és az Spp1-től, míg a GBD-Spp1 esetén nem függ sem a hiszton oldalláncától sem a Set1 fehérje jelenlététől.

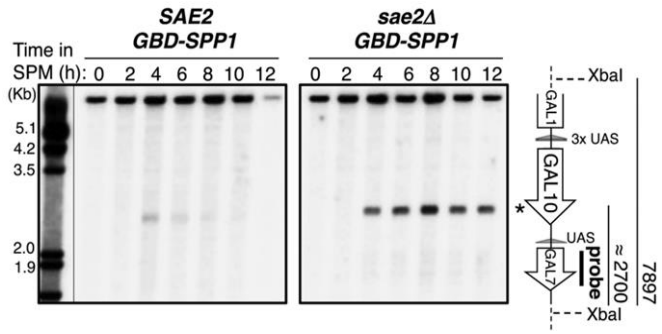
A GBD-targeting kísérletek váratlan és elsőre megmagyarázhatatlan eredménye volt, hogy amíg a Set1 katalitikus egység által kiváltott meiotikus DSB-k képződése a módosítható H3K4 oldalláncától és az Spp1 jelenlététől függött, addig az Spp1 hatása független volt a H3K4 oldallánc és a Set1 jelenlététől (**15D-E ábra, 2-3. oszlop**). A GBD-Spp1 fehérje a H3K4R és *set1Δ* mutánsokban tömeges meiotikus DNS

töréseket váltott ki a *GAL2* coldspot régióban, annak hatása sokszorosan felülmúlta a Set1 hatását. A jelenség genetikai hátterét tovább boncolva kiderült, hogy GBD-Set1 hatása a fenti faktorokon kívül függ még a Set1C/COMPASS katalitikus aktivitásától (GBD-set1G951S mutáns), a H2B monoubiquitinációért felelős Bre1/2 enzimektől (amely a H3K4me3 előfeltétele), és a Spo11 jelenlététől (*spo11Δ*; **16/A ábra**). A GBD-Spp1-el elvégzett hasonló genetikai elemzés viszont azt mutatta, hogy az Spp1 hatása csupán a Spo11 nukleáztól függ (**16/B ábra**).



16. ábra A GBD-Set1 és GBD-Spp1 indukált meiotikus DNS törések genetikai determinánsai. *A. Felső panel:* A vad típusú törzsben és a GBD "only" törzsben nem keletkeznek meiotikus DSB-k a *GAL2* coldspot régióban, ellenben a GBD-Set1 konstrukt tömeges DNS töréseket indukál ugyanitt. Alsó panel: A GBD-Set1 fehérje DSB indukciós hatása függ az oszlopokban utatott fehérjéktől. *B.* A GBD-Spp1 fehérje DSB-indukáló hatása kizárólag a Spo11 jelenlététől függ. A Southern blot-ok alatti számok a DSB sávok intenzitását mutatják.

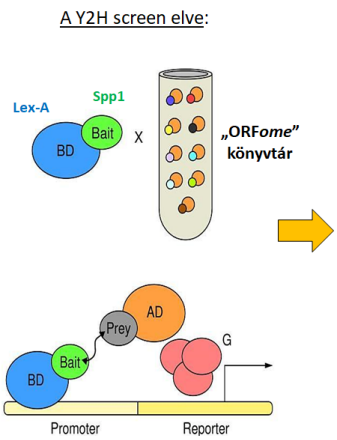
A GBD-Spp1 által indukált meiotikus DSB-ket megvizsgáltuk *SAE2*(+) genomi háttérben is, amely nem állítja le a DSB-k processzálását és a hibahavító útvonalakat (**17. ábra**). A kísérlet szerint a *GAL10* coldspot-ban indukált DSB-k normál kinetikával javítódnak, vagyis a mesterségesen indukált DSB-k reparációjukat tekintve megegyeznek a "természetesen" keletkező meiotikus DSB-kkel.



17. ábra GBD-Spp1 indukálta meiotikus DNS törések hibajavítást-megengedő és nem-engedő sejtekben. *SAE2(+)* genomi háttérben megtörténhet a DSB végek processzálása és javítása, míg *sae2Δ* háttérben nem keletkeznek 3'OH végű visszavágott preszinaptikus filamentumok, amely a rekombinációs javítás előfeltétele. Előbbi sejtekben a DSB-k nem akkumulálódnak, hanem 4-6 óránál rekombinációval javítódnak. Utóbbi sejtekben a DSB-k folyamatosan akkumulálódnak, amely a nagypontosságú és érzékeny DSB térképezés fontos technikai feltétele.

A fenti eredmények legegyszerűbb magyarázata, hogy az Spp1 valamilyen módon fizikai kontaktusba lép a meiotikus DSB-komplex egyik (vagy több) tagjával, esetleg magával a Spo11-el, amelyet odatoboroz a rekombinációs coldspot régiókhoz, ahol emiatt DNS törések keletkeznek. Hipotézisünket élesztő 2-hibrid (Y2H) screen

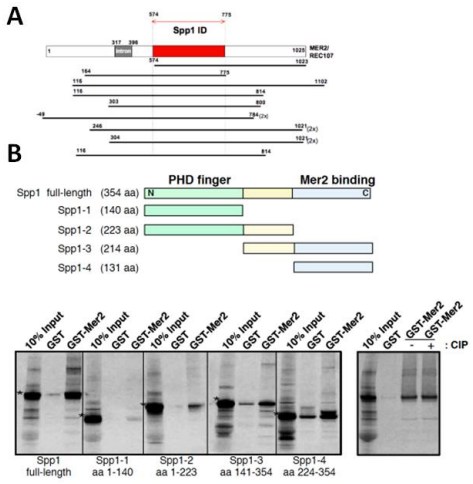
kísérlettel igazoltuk *in vitro*, amelyben a teljes (~6000) élesztő fehérjekódoló génekre vonatkozóan ("ORF*ome*") elvégeztünk egy szűrést oly módon, hogy csaliként az Spp1 LexA-fúziós fehérjét használtuk (**18. ábra**).



Clone	ID	Gene	Gene name	Start	Stop	Frame	Orientation
YPL138C	1aA-29	YHR119W	(SET1)	2156	(3239)	OOF2	Sense
YPL138C	1aA-30	YHR119W	(SET1)	2088	2588	IF	Sense
YPL138C	1aA-2	YHR119W	(SET1)	1659	(3239)	IF	Sense
YPL138C	1aA-64	YHR119W	(SET1)	1386	(3239)	IF	Sense
YPL138C	1aA-93	YHR119W	(SET1)	2098	2380	OOF1	Sense
YPL138C	1aA-142	YHR119W	(SET1)	2098	2380	OOF1	Sense
YPL138C	1aA-226	YHR119W	(SET1)	2056	2631	OOF1	Sense
YPL138C	1aA-176	YHR119W	(SET1)	1386	2448	IF	Sense
YPL138C	1aA-136	YHR119W	(SET1)	2156	2457	OOF2	Sense
YPL138C	1aA-144	YHR119W	(SET1)	2156	2457	OOF2	Sense
YPL138C	1aA-183	YHR119W	(SET1)	2088	2588	IF	Sense
YPL138C	1aA-160	YHR119W	(SET1)	1671	2406	IF	Sense
YPL138C	1aA-36	YHR119W	(SET1)	2156	(3239)	OOF2	Sense
YPL138C	1aA-1	YHR119W	(SET1)	2156	(3239)	OOF2	Sense
YPL138C	1aA-8	YHR119W	(SET1)	2156	(3239)	OOF2	Sense
YPL138C	1aA-37	YHR119W	(SET1)	2156	(3239)	OOF2	Sense
YPL138C	1aA-41	YHR119W	(SET1)	2156	(3239)	OOF2	Sense
YPL138C	1aA-85	YHR119W	(SET1)	2060	(3239)	OOF1	Sense
YPL138C	1aA-48	YHR119W	(SET1)	2056	2563	OOF1	Sense
YPL138C	1aA-60	YHR119W	(SET1)	1973	(3239)	OOF2	Sense
YPL138C	1aA-40	YHR119W	(SET1)	2088	2588	IF	Sense
YPL138C	1aA-200	YHR119W	(SET1)	2056	(3239)	OOF1	Sense
YPL138C	1aA-104	YHR119W	(SET1)	1671	2401	IF	Sense
YPL138C	1aA-170	YHR119W	(SET1)	1671	2406	IF	Sense
YPL138C	1aA-127	YJR021C	(MER2/REC107)	574	1023	OOF1	Sense
YPL138C	1aA-154	YJR021C	(MER2/REC107)	164	775	OOF2	Sense
YPL138C	1aA-114	YJR021C	(MER2/REC107)	116	1102	OOF2	Sense
YPL138C	1aA-119	YJR021C	(MER2/REC107)	116	814	OOF2	Sense
YPL138C	1aA-42	YJR021C	(MER2/REC107)	303	800	IF	Sense
YPL138C	1aA-18	YJR021C	(MER2/REC107)	-49	784	IF	Sense
YPL138C	1aA-13	YJR021C	(MER2/REC107)	-49	784	IF	Sense
YPL138C	1aA-31	YJR021C	(MER2/REC107)	246	(1021)	IF	Sense
YPL138C	1aA-212	YJR021C	(MER2/REC107)	304	(1021)	IF	Sense
YPL138C	1aA-202	YJR021C	(MER2/REC107)	304	(1021)	IF	Sense

18. ábra Élesztő 2-hibrid screen az Spp1 fizikai interakciós partnereinek az azonosítására. Az Spp1-et LexA-fúziós fehérje formájában expresszáltuk "csaliként", hogy ~6000 potenciális fehérje partner közül ("ORF*ome*") megtaláljuk a valódi interakciós partnereket. Méréseinkben két szignifikáns találatot azonosítottunk: a Set1-et (amely nem volt váratlan, hiszen az Spp1 a Set1-komplex része) és a Mer2 DSB fehérjét (pirossal kiemelve).

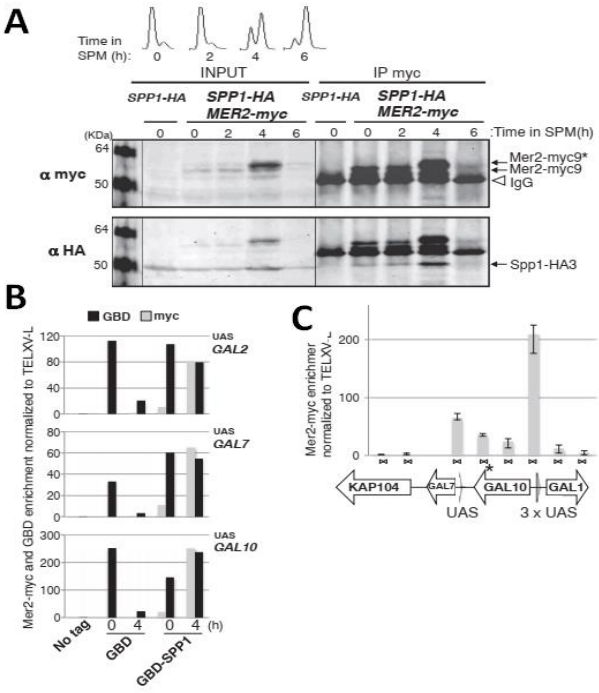
A screen eredményeként a Mer2 rekombinoszóma fehérjét azonosítottuk az Spp1 fehérje-fehérje interakciós partnereként, amely a Spo11 közvetlen partnere és aktivátora (Arora et al. 2004; Zhang et al. 2020; Rousova et al. 2021) és a kromoszóma tengelyre lokalizálódik (Panizza et al. 2011). A kölcsönhatást glutation S-transzferáz / Mer2 pull-down kísérletekkel validáltuk különböző hosszúságú (trunkált) *in vitro* transzlált Spp1 polipeptidek felhasználásával, amelyben a Mer2-vel való kölcsönhatáshoz szükséges domént az Spp1 131 aminosavból álló C-terminális régiójára térképeztük (**19. ábra**).



19. ábra Spp1-Mer2 kölcsönhatás *in vitro* vizsgálata GST pull down kísérlettel. *A.* A Y2H screen-ben azonosított Mer2 fragmentumok sematikus ábrázolása a teljes hosszúságú Mer2/Rec107 mentén. A feltételezett Spp1-interakciós Zn-finger (CXXC) domént (Murton et al. 2010) pirossal mutatjuk. *B.* A rekombináns GST-Mer2 fehérjét *E. coli*-ban expresszáltuk és tisztítottuk, amelyhez *in vitro* transzlált teljes hosszúságú Spp1 fehérjét vagy trunkált Spp1-et adtunk. Az *in vitro* transzlált [^{35}S]-metioninnal jelölt Spp1 fehérjéket az alsó panelen csillaggal jelöltük. Az input a reakciókban résztvevő összes radioaktív anyag 10%-át mutatja. Az Spp1 oldaláról egy 131 aminosavból álló C-terminális régiót azonosítottunk, amely szükséges az Spp1-Mer2 interakcióhoz.

A két fehérje fizikai kölcsönhatását *in vivo* kísérletekben is megerősítettük. Co-IP mérésekben Spp1-3xHA, Mer2-9xmyc "dupla tag" törzseket vittünk szinkron-meiózisba majd anti-myc és anti-HA antitestekkel western blot-ot végeztünk, amely a Mer2-n keresztül feldúsította az Spp1-et, és fordítva (**20/A ábra**). Ezután ChIP-qPCR-mérésekben kimutattuk a GBD-Spp1 kötődését a *GAL2*, *GAL7*, és *GAL10* coldspot régiók *GAL4*$_{UAS}$ motívumaihoz, amely időben egybe

esett a DSB-képződés tipikus időpontjával SK1 élesztő sejtekben (ez a meiózis negyedik órája; **20/B ábra**). Ugyanebben az időpontban a Mer2-myc specifikus kromatin kötődését is detektáltuk (**20/C ábra**).

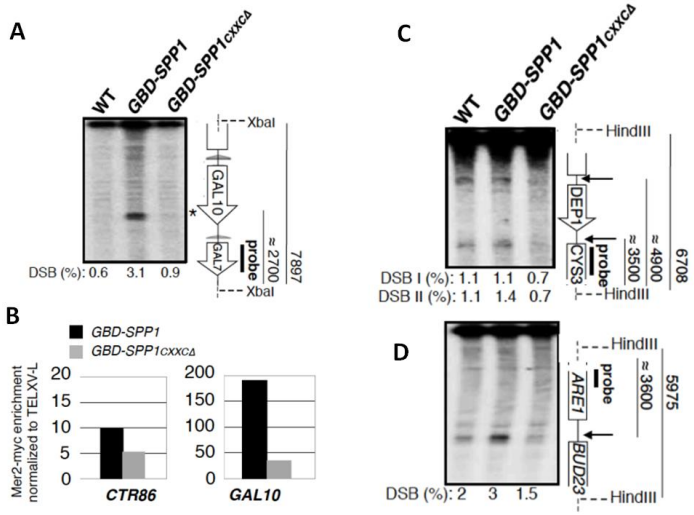


20. ábra Spp1-Mer2 kölcsönhatás kimutatása *in vivo*. *A.* Co-IP kísérlet Spp1-3xHA és Mer2-9xmyc fehérjéket expresszáló meiotikus élesztő sejtekben. A meiózis stádiumait FACS-al követtük nyomon (0-6 h SPM). TEljes-sejt extraktumból western blot-ot csináltunk anti-myc, majd anti -HA antitestekkel, amely kimutatta a Mer2-n keresztül az Spp1 partnert. *B-C.* A GBD-Spp1 és Mer2-myc fehérje specifikus dúsulást mutatott a *GAL2*, *GAL7*, és *GAL10* coldspot régiók $GAL4_{UAS}$ motívumainál a meiózis negyedik órájában, amely a DSB képződés tipikus időpontja vad típusú sejtekben. Mintajelölés: "No tag": myc- és GBD-nélküli kontrol törzs. "GBD": csak a Gal4 domént expresszáló kontrol törzs.

Összegezve, a fenti *in vitro* és *in vivo* kísérletekben bizonyítottuk, hogy a Set1C hisztonmódosító enzimkomplex egyik alegysége, az Spp1, fizikai kölcsönhatásba lép a meiotikus DSB-komplex egyik tagjával, a Mer2 fehérjével, s e két fehérje $GAL4_{UAS}$ célszekvenciákhoz történő kötődése időben egybeesik a DSB-képződés időpontjával a meiózis során.

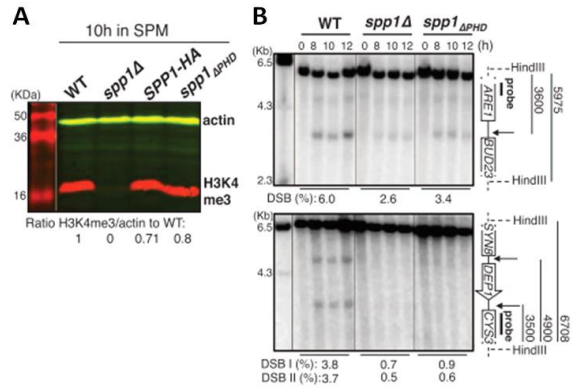
Az Spp1 fehérje Mer2-interakciós doménjének közvetlen szerepét további funkcionális kísérletekben igazoltuk. Ehhez az Spp1-ből kitöröltük a fehérje korábban leírt cinkujjszerű (CXXC) doménjét (Murton et al. 2010), amely egybeesett az általunk felfedezett Mer2-interakciós doménnel, majd megvizsgáltuk, hogy a deléció befolyásolja-e a DSB képződést és a Mer2 kötődését a *GAL10* coldspot régióban. A GBD-$Spp1^{CXXC\Delta}$ mutánsban mind a DSB-képződés mind a Mer2-kötés szignifikánsan csökkent a GBD-Spp1-hez képest (**21/A-B ábra**), amely azt jelzi, hogy az Spp1-Mer2 interakciós motívum szükséges a Mer2 kromatin kötéséhez és a DSB keletkezéshez. Az Spp1 CXXC motívum

hiánya hasonló DSB-csökkenést okozott két természetes hotspot esetén (**21/C-D ábra**), amely tovább erősíti az Spp1-Mer2 interakció funkcionális szerepét a rekombináció iniciációjában.



21. ábra A CXXC motívum deléciója meggátolja a GBD-Spp1 indukált DSB képződést és a Mer2 kromatinkötődését. *A.* Gal10 coldspot régióban nem keletkezik GBD-Spp1 indukált DSB, ha az Spp1-ből kitöröljük a CXXC motívumot, amely a feltételezett Mer2-interakciós domén. A Southern blot mérés 4h SPM időpontban történt. *B.* Reprezentatív meiotikus ChIP-qPCR kísérlet (4h SPM) egy kontrol szakaszon (*CTR86*) és a *GAL10* coldspot régióban. Az Spp1 CXXC motívum deléciója meggátolja a Mer2-myc fehérje kromatinkötődését. *C-D.* Az Spp1 CXXC motívum szükséges a DSB képződéshez a *DEP1-CYS3* és *BUD23-ARE1* "természetes" hotspot régiókban. CXXC motívum hiányában szignifikánsan kevesebb DSB keletkezik.

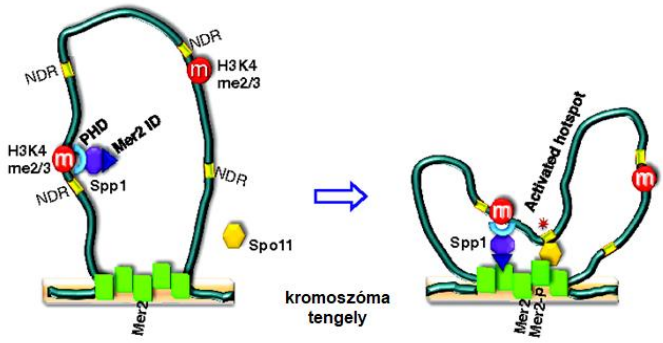
A hiszton H3K4 trimetilációnak van-e közvetlen mechanikai szerepe az Spp1-Mer2 interakció létrejöttében és a meiotikus DSB-k keletkezésében? Ez egy kulcsfontosságú kérdés, amely megválaszolásához az Spp1 fehérjéből kitöröltük a H3K4me3 jel kiolvasásáért felelős PHD domént (*spp1PHDΔ* mutáns) és megvizsgáltuk a DSB-keletkezés gyakoriságát a *BUD23-ARE1* és *DEP1-CYS3* hotspot régiókban (**22. ábra**).



22. ábra Az Spp1 PHD finger doménje szükséges a DSB-k keletkezéséhez. *A.* A H3K4me3 mennyiségének mérése western blot-tal (10h SPM meiotikus időpontban). A *spp1PHDΔ* mutáns nem befolyásolja a H3K4me3 szintjét. *B.* DSB térképezés Southern blot módszerrel a *BUD23-ARE1* és *DEP1-CYS3* hotspot régiókban. A *spp1PHDΔ* mutánsban hasonló DSB-szint csökkenést látunk, mint a *spp1Δ* deléciós mutánsban. A domináns DSB band-ek intenzitását a gélek alatti számok mutatják.

Az *spp1PHDΔ* mutáció, a várakozásnak megfelelően, nem befolyásolta a H3K4me3 modifikáció szintjét (**22/A ábra**), viszont szignifikánsan csökkentette a meiotikus DSB-k számát a vizsgált szakaszokon (**22/B ábra**). Mindebből az következik, hogy az Spp1 H3K4me3-olvasó PHD doménje specifikus szerepet játszik a meiotikus DSB-k kialakulásában. A fenti eredményeket összegezve, a meiotikus homológ rekombináció iniciációjához szükség van H3K4me3 modifikációra, amelyet a Set1C/COMPASS katalizál, szükség van Mer2 DSB-fehérjére, amely a Spo11-kompexet rögzíti a kromoszóma tengelyhez (Panizza et al. 2011), és szükség van Spp1-re, amely a

PHD-finger és Mer2-ID doménjén keresztül fizikai kapcsolatot teremt az epigenetikai jel és a DNS törések között (**23. ábra**).



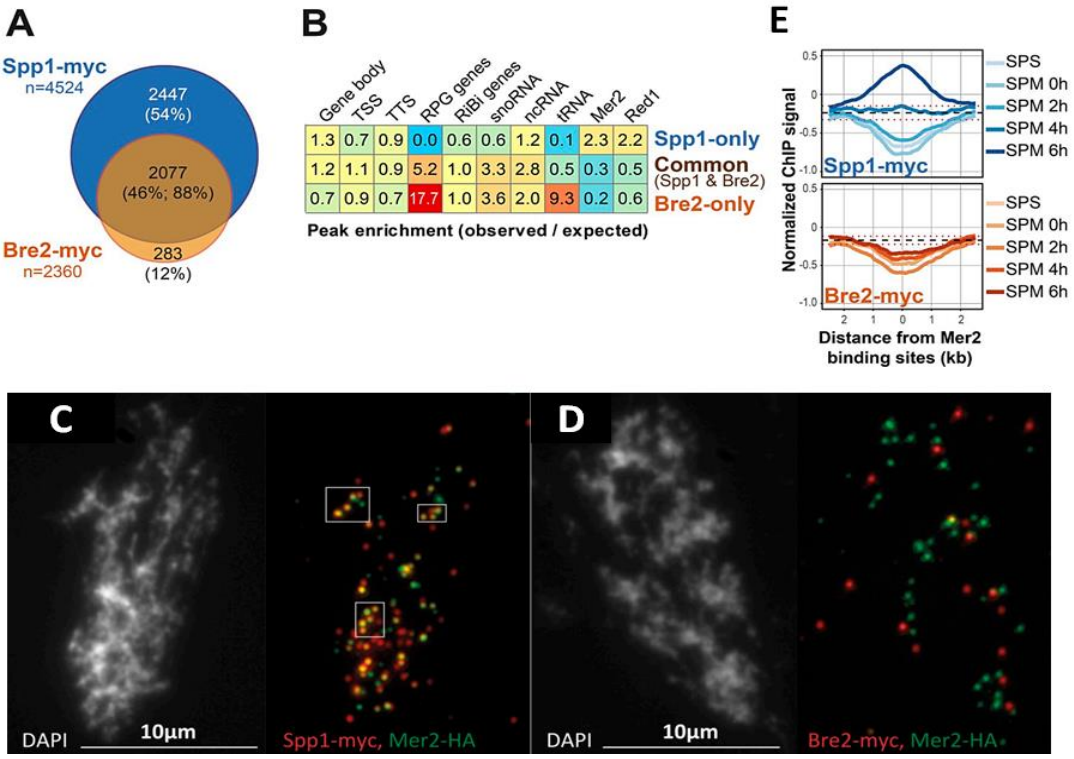
23. ábra A meiotikus homológ rekombináció iniciációjának új hurok-tengely modellje. Az Spp1 fehérje a PHD finger doménjén keresztül a nukleoszóma-mentes szakaszokat (NDR) szegélyező H3K4me3 módosításhoz kötődik, majd a Mer2-kötő doménjén keresztül a kromoszóma tengelyen található rekombinószóma komplexszel lép fizikai kölcsönhatásba a Mer2 fehérjén keresztül. Ezáltal a kromatin hurkokon található rekombinációs forrópont „odairányítódik" a kromoszóma tengelyen lokalizálódó Spo11 enzimhez, amely DNS kettősszál törést indukál. Forrás: (Acquaviva & Szekvolgyi et al. 2013).

Kísérleteink alapján egy új rekombinációs modellt vezettünk be, amely egyrészt kiterjesztette a régebbi hurok-tengely modellt (Blat et al. 2002), másrészt az Spp1-Mer2 kölcsönhatás alapján molekuláris szinten megmagyarázta a meiotikus DNS-törések kapcsolatát a Set1C-katalizált H3K4me3 módosítással.

Az Spp1 kromatinkötődési dinamikájának vizsgálata a meiózis során

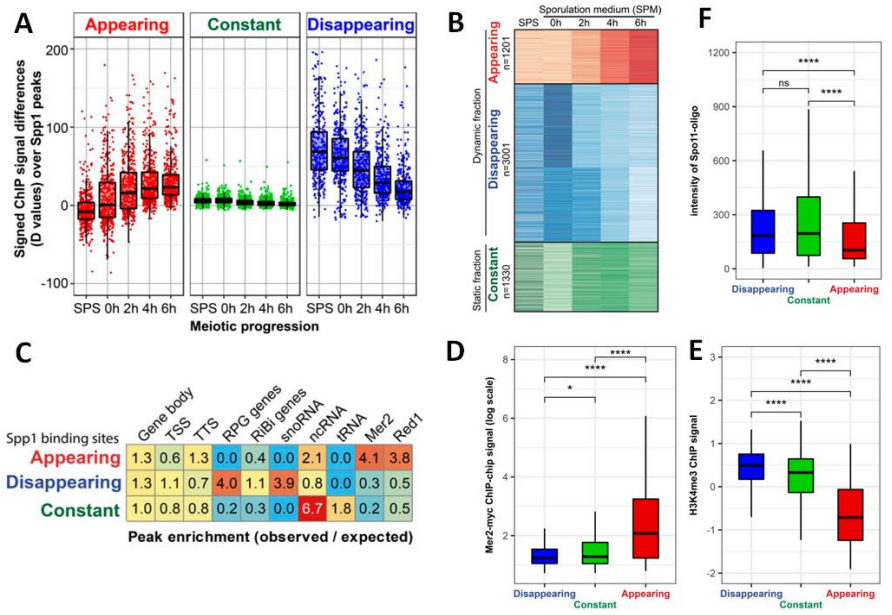
A fenti hurok-tengely modell alapján szerettük volna részleteiben megérteni az Spp1 speciális DSB-segítő funkcióját a meiózis során. Például azt, hogy az Spp1 a Set1C/COMPASS hisztonmetiláz komplex részeként fejti-e ki rekombináció-potencírozó hatását, vagy esetleg a Set1C-ről leválva, attól függetlenül működik. A kérdés megválaszolásához 9xmyc epitóppal jelölt Spp1 és Bre2 fehérjéket expresszáltunk meiotikusan differenciálódó élesztő sejtekben (Spp1-9xmyc, Bre2-9xmyc) majd ChIP-szekvenálással meghatároztuk a genomi eloszlásukat (cisztrom) a meiózis profázisa során (0-6h SPM). A Bre2 fehérje a Set1 komplex egyik stabil alegysége, ezért genomi eloszlása markerként használható a Set1C pozíciójának jelölésére. Az azonosított kromatinkötő helyek átfedésének az elemzése azt mutatta, hogy az Spp1 csúcsok mintegy 46%-a egybeesik a Bre2/Set1C kötőhelyekkel a meiózis alatt, míg 54%-uk attól független genomi eloszlást mutat (**24/A ábra**). A különbség és metszet halmazok annotációja szerint a Set1C-független Spp1 frakció ("Spp1-only" kötőhelyek; n=2447) jelentős dúsulást mutattak a Mer2 és Red1 kromoszómatengely kötőhelyek felett, míg ugyanitt a "Bre2-only" (n=283) és a közös Spp1/Bre2 csúcsok ("common", n=2077) nem dúsultak (**24/B ábra**). Az Spp1-Mer2 kolokalizációt immunfluorescens mikroszkópiával is megerősítettük meiotikus kromatin preparátumokon (a Mer2 fehérjét ebben az esetben 3xHA tag-gel fúzionáltattuk; **24/C ábra**). E mérésekben a Bre2 jelentős antikolokalizációt mutatott a Mer2-vel (**24/D ábra**). A

mikroszkópos adatok tehát alátámasztották a ChIP-seq eredményeket. További érdekes megfigyelés, hogy az Spp1-hez hasonlóan a Bre2 kötőhelyeknek is létezik egy Set1C-től független részhalmaza ("Bre2-only" csúcsok; n=283 (12%)), amely a riboszómális fehérjéket kódoló génekhez (RPG) és a tRNS génekhez kötődött preferenciálisan, azonban ezen minor Bre2-frakció biológiai jelentősége nem ismert. Fontos megjegyezni, hogy az Spp1 kromatinkötése progresszív növekedést mutatott a Mer2 kötőhelyeken a meiózis profázisa alatt (**24/C ábra**), míg ugyanitt a Bre2/Set1C peak-ek egyáltalán nem mutattak dúsulást. Eszerint, a meiotikus differenciáció során az Spp1 molekulák mintegy fele leválik a Set1 komplexről és "átkerül" a Mer2 kötőhelyekre, s mindez a Set1C/COMPASS alapfunkciójától függetlenül történik.



24. ábra Az Spp1 és Bre2 genomi eloszlása a meiózis profázisa során. *A.* A ChIP-seq kísérletben azonosított Spp1-myc és Bre2-myc kötőhelyek átfedése. A Venn diagram elemzéshez az egyes meiotikus időpontokban azonosított Spp1 és Bre2 peak-ek únióját képeztük az időpontok szerint (SPS, 0h, 2h, 4h, 6h SPM), majd a két úniós peak-set átfedését ábrázoltuk. *B.* A peak-ek különbség halmazainak és metszet halmazának annotációja különféle genomi annotációs kategóriákon. Az "Spp1-only" peak-ek (vagyis az Spp1 molekulák Set1C-től független frakciója) jelentős dúsulást mutatnak a Mer2 és Red1 kromoszómatengely kötőhelyek felett, ugyanitt a "Bre2-only" és közös (common Spp1/Bre2) peak-ek nem dúsulnak. Érdekes megfigyelés, hogy a Bre2-nek is van egy Set1C-től független kisebb részhalmaza (összesen 283 peak), amely a riboszómális fehérjéket kódoló géneken (RPG) és a tRNS géneken dúsul. A cellák a megfigyelt peak-ek arányát mutatja a randomizált peak-ekhez képest. *C-D.* Immunfluoreszcens mikroszkópia meiotikus kromoszóma preparátumokon (4h SPM) az Spp1-Mer2 és Bre2-Mer2 kolokalizáció vizsgálatára. Az Spp1-myc és Bre2-myc fehérjéket piros szín jelzi, míg a Mer2-HA zöldben látszik. A DNS (DAPI festés) szürkében látható. Az Spp1 és Mer2 jelentős átfedést mutat (sárga szín), a Bre2-Mer2 viszont nem. *E.* Spp1 és Bre2 metaplot elemzés, amely különböző meiotikus időpontokban mutatja az átlagos ChIP-seq jelet a Mer2 kromoszómális kötőhelyek felett. Az Spp1 progresszív dúsulást mutat a Mer2 kötőhelyeken, míg a Bre2 esetén depléció látható. A szaggatott vonal mutatja a ChIP-seq szignál genomi átlagát.

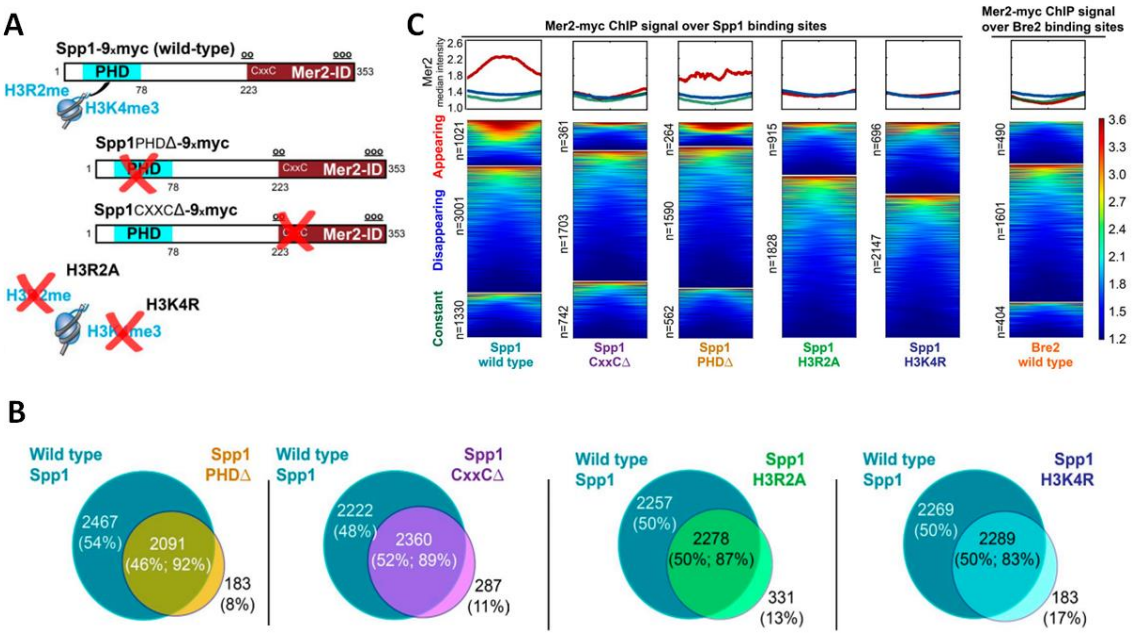
Az Spp1 kromatin dinamikájának mélyebb megértéséhez időfüggő klaszteranalízist végeztünk az azonosított ChIP-seq kötőhelyeken, így az Spp1 csúcsokat osztályokba tudtuk sorolni a hasonlóságuk alapján. Két fő kinetikus csoportot különítettünk el az Spp1 kötőhelyekre jellemző ChIP-seq értékek időbeli változása alapján: 1. dinamikus Spp1 kötőhelyek, amelyek a meiózis előrehaladtával fokozatosan megjelentek (ún. megjelenő alcsoport) vagy eltűntek (ún. eltűnő alcsoport), 2. statikus helyek, amelyek időben állandó asszociációt mutattak az Spp1-el (**25/A-B ábra**). A dinamikus és statikus kötőhelyek funkcionális annotációja alapján i) a megjelenő Spp1 csúcsok erősen dúsultak a kromoszóma tengely mentén (a Mer2 és Red2 kötőhelyeken); ii) az eltűnő Spp1 kötőhelyek ugyanitt depléciót mutattak, ellenben a riboszómális fehérje (RPG) és snoRNS géneken felülreprezentáltak voltak; iii) a statikus Spp1 peak-ek erősen kötődtek a ncRNS génekhez (**25/C ábra**).



25. ábra Klaszteranalízissel azonosított Spp1 kötőhelyek kinetikus osztályai. *A-B.* A megjelenő (piros) és eltűnő (kék) Spp1 kötőhelyek dinamikusan növekvő illetve csökkenő ChIP-seq értéket mutatnak a meiózis előrehaladása során, míg a konstans Spp1 kötőhelyek (zöld) nem mutatnak jelentős időbeli változást. *C.* A három kinetikus frakció funkcionális annotációja. A cellák a random peak-ekhez képest mért statisztikai dúsulás mértékét mutatják. *D-F.* Az eltűnő, konstans, és megjelenő Spp1 kötőhelyek felett mért Mer2 ChIP, H3K4me3 ChIP, és Spo11-oligo szignálok eloszlása. Az Spp1 kötőhelyek dinamikus osztályai eltérő feldúsulást mutatnak a Mer2 kötődésében (D panel, Panizza és munkatársai, 2011), a H3K4me3 szintjében (E panel, Borde et al., 2009) és a Spo11-oligok mennyiségében (F panel,, Pan et al 2011). A box-whisker diagrammok a Mer2/H3K4me3/Spo11-oligo ChIP jelek mediánját mutatják (interkvartilis tartományokkal). Statisztikailag szignifikáns különbség: *, P < 0,05; ****, P < 0,0001; ns, nem-szignifikáns; Mann-Whitney U teszt.

Egy másik (komplementer) megközelítésben megvizsgáltuk a Mer2, H3K4me3, és Spo11-oligo eloszlást a három kinetikus Spp1 frakcióban (publikus ChIP-seq adatok alapján; (Borde et al. 2009; Panizza et al. 2011; Pan et al. 2011), amely szerint a Mer2 nagyobb affinitással kötődik a megjelenő Spp1 peak-ekhez, mint a konstans és megjelenő peak-ekhez, amelyek viszont a H3K4me3 modifikációval és a Spo11-oligókkal (tehát a DNS duplaszál törésekkel) mutattak szoros asszociációt (**25/D-F ábra**). Mindebből arra a következtetésre jutottunk, hogy az Spp1 kromatinkötődési

dinamikája és az ez alapján azonosított kinetikus frakciók erősen korrelálnak az Spp1 Set1C-független (meiózis specifikus) és Set1C-függő (általános) biológiai funkcióival. Hogy jobban megértsük az Spp1 kromatinkötésének molekuláris mechanizmusát, funkcionális ChIP-seq elemzést végeztünk különféle Spp1 funkcióvesztéses mutánsokban, amelyekben az Spp1 nem tud kötődni a H3K4me3 módosításhoz (*Spp1$_{PHD\Delta}$* mutáns), vagy nem kapcsolódik a Mer2 fehérjéhez (*Spp1$_{CXXC\Delta}$* mutáns), vagy nincs jelen a H3K4me3 modifikáció (H3K4R és H3R2A hiszton mutánsok), amelyet az Spp1 a Mer2-höz horhonyozhatna a PHD doménjén keresztül (**26/A ábra**).



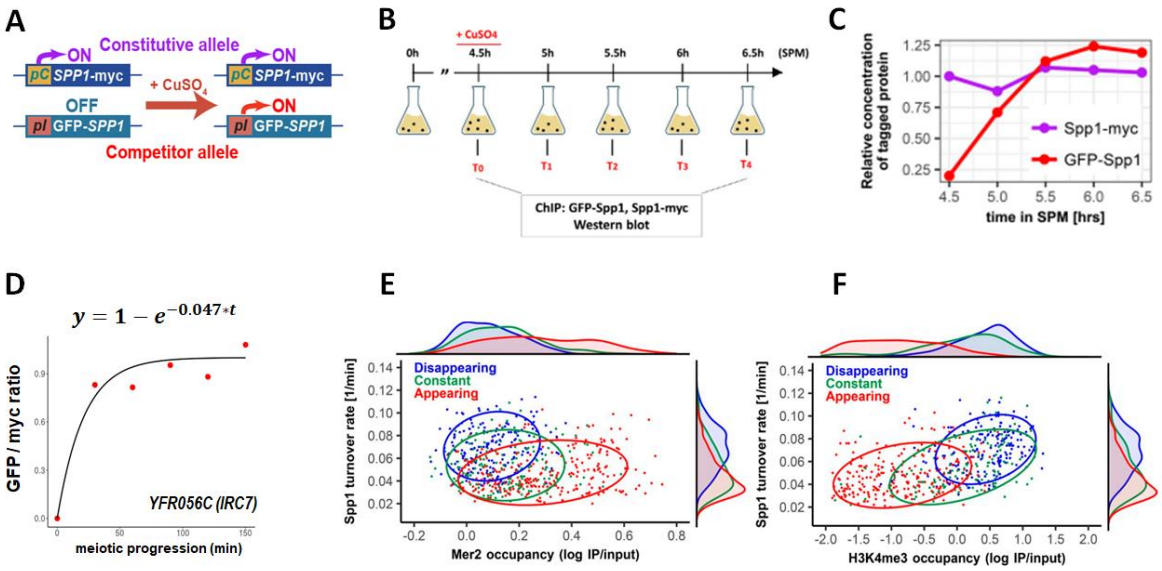
26. ábra Meiotikus Spp1 kötőhelyek funkcionális analízise. *A.* ChIP-seq elemzésben használt mutánsok, amelyekben az Spp1 nem kötődik a H3K4me3.hoz (Spp1PHDΔ), vagy nem kapcsolódik a Mer2-höz (Spp1CXXCΔ), vagy nincs jelen a H3K4me3 modifikáció (H3K4R és H3R2A), amelyet a Mer2-höz kapcsolhatna az Spp1. *B.* Vad típusú és mutáns sejtekben azonosított Spp1 kötőhelyek átfedését bemutató Venn diagramm. A mutációk a vad típusú sejtekben azonosított Spp1 kötőhelyek kb. 50%-át eliminálják. A mutánsokban megmardt Spp1 csúcsok körülbelül 90%-a átfed a vad típusú Spp1 helyekkel. A mutánsokban detektált Spp1 csúcsok körülbelül 10%-a ektópikus, vagyis *de novo* alakul ki a mutáció hatására, amely oka ismeretlen. *C.* A Mer2 kötődése a mutánsokban azonosított Spp1 kötőhelyek dinamikus osztályai felett. A bal oldali metaplot a Mer2-myc ChIP szignált mutatja az Spp1 csúcsokra centrálva, míg a jobb oldali metaplot ugyanazt mutatja a Bre2 kötőhelyekre. Mindhárom mutáció szignifikánsan csökkenti a Mer2 kötődését a "megjelenő" Spp1 csoportban (pirossal jelezve).

Mindhárom esetben a feltételezett hurok-tengely kölcsönhatás megszűnését és a meiotikus DSB-k szintjének jelentős csökkenését várhatjuk (lásd: **15. ábra**, **21-22. ábrák**). A mutánsokban ChIP-seq elemzést végeztünk a meiózis különböző stádiumaiban (0h, 2h, 4h, 6h SPM) és meghatároztuk az Spp1 kötőhelyek cisztromját. A vad típusú sejtekben azonosított kötőhelyekhez viszonyítva a mutánsokban detektált Spp1 kötőhelyek kb. 50%-a eltűnt (**26/B ábra**), vagyis az Spp1 fehérje PHD doménje, CXXC doménje, illetve a H3K4me3 módosítás szükséges az Spp1 stabil kromatinkötéséhez.

A mutánsokban megmaradó Spp1 csúcsok ~90%-a átfedett a vad típusú Spp1 helyekkel, tehát a mutációk többsége nem változtatta meg az Spp1 kötőhelyek genomi eloszlását. A peak-ek ~10%-a viszont ektópikus lokalizációt mutatott, vagyis ezek az Spp1 kötőhelyek *de novo* alakultak ki a mutációk hatására és a vad típusú sejtekben egyáltalán nem voltak jelen. A **26/C ábrán** látható metaplot elemzés szerint a Mer2 fehérje kötődése (publikus ChIP-seq adat alapján, (Panizza et al. 2011) szignifikánsan csökkent az *Spp1$_{CXXC\Delta}$*, H3R2A és H3K4R mutánsokban azonosított Spp1 kötőhelyeken. Az *Spp1$_{PHD\Delta}$* mutánsban az Spp1 kötőhelyek 75%-a (264/1021) eltűnt, azonban a fennmaradó 25%-nyi kötőhelyből 130 Spp1 csúcs (összesen 12%) továbbra is jelentős Mer2-dúsulást mutatott (összehasonlításképpen, az *Spp1$_{CXXC\Delta}$*, H3R2A és H3K4R mutánsokban az Spp1 kötőhelyek 100%-ában megszűnt a Mer2 dúsulás). A *Spp1$_{PHD\Delta}$* mutáció hatása azonban így rendkívül markáns volt, amely összhangban van a DSB-adatokkal (**22. ábra**). Mindebből arra következtettünk, hogy az Spp1 H3K4me3-olvasó PHD doménje, a Mer2-interakciós CXXC doménje, illetve a H3K4me3 módosítás jelenléte szükséges a Mer2-vel történő kromatin-kölcsönhatáshoz. Összehasonlításképpen kielemeztük a Bre2 ChIP-seq kötőhelyek Mer2-vel való asszociációját, amely a várakozásnak megfelelően egyáltalán nem mutatott Mer2-dúsulást (**26/C ábra, jobb panel**).

A továbbiakban az Spp1 kromatin-dinamikáját kvantitatívan jellemeztük különféle genomikai és mikroszkópos módszerek segítségével (c-ChIP, FRAP, FCS). A kompetíciós ChIP (c-ChIP) eljárásban (Lickwar et al. 2012, 2013) egy konstitutív és egy indukálható Spp1 izoformát "versenyeztettünk" egymással a genomi kötőhelyekhez történő kötődésért, amely alapján ún. kicserélődési rátát számoltunk (turnover rate), amely az időegység alatt történő Spp1 kicserélődések számát mutatja az egyes Spp1 kötőhelyekre vonatkozóan (Spp1 kicserélődés / perc / kötőhely; mértékegysége: [1/perc]). Mérésünkben 9xmyc tag-gel jelöltük a konstitutív Spp1 allélt amelyet endogén promóteréről expresszáltunk (Spp1-9xmyc), míg a kompetitor Spp1 allélt GFP-vel fúzionáltattuk egy réz-indukálható promoter (p*CUP1*) szabályozása alatt (GFP-Spp1) (**27/A ábra**). A megfelelő időbeli-felbontóképesség elérése érdekében a réz-indukciót követően 30 percenként vettünk mintát a meiotikus tenyészetekből (**27/B ábra**), majd western blot elemzéssel ellenőriztük a kompetitor allél expressziójának exponenciális növekedését (**27/C ábra**). (Az indukciót úgy optimalizáltuk az előkísérletek során, hogy elkerüljük a GFP-Spp1 overexpressziót, amely ektópikus kötőhelyek megjelenéséhez vezetett volna.) Ezt követően kromatin immunprecipitációt (ChIP) végeztünk anti-myc és anti-GFP antitestekkel, majd genomi kötőhelyeket azonosítottunk a "hagyományos" ChIP-seq elemzéshez használt bioinformatikai módszerekkel. Kicserélődési rátát azokra a peak-ekre definiáltunk, amelyek illeszthetőek voltak az elméleti exponenciális modell alapján ($\frac{GFP}{myc}$= (1-e$^{-\lambda t}$),

ahol λ a turnover ráta, t pedig a kompetitor allél indukciójától számított idő. Egy reprezentatív Spp1 kötőhely (*IRC7*) illesztése a **27/D ábrán** látható. A modell illesztése után kiszámítottuk a becslések standard hibáját (SE), amely figyelembe vételével összesen 977 c-ChIP kötőhelyet és Spp1 turnover rátát definiáltunk.

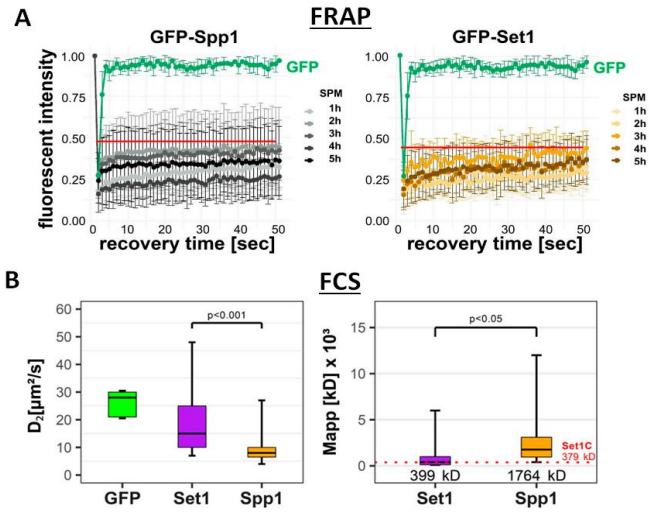


27. ábra Spp1 kötőhelyek turnover rátájának elemzése kompetíciós ChIP-el. *A-B.* A c-ChIP kísérlet sémája. Konstitutív allél: Spp1-myc; indukálható allél; GFP-Spp1 (p*CUP1* réz-indukálható promóterről expresszálva). A sporuláló élesztő sejtekhez réz-szulfátot adva indukáljuk a kompetitor allél expresszióját, majd 30 percenként mintát veszünk a tenyészetből és western blot-ot illetve ChIP-seq mérést csinálunk anti-myc és anti-GFP antitestekkel. *C.* Reprezentatív western blot kísérlet kvantifikációja, amely a kompetitor allél expresszió exponenciális növekedését mutatja a meiotikus időkinetika során. *D.* Reprezentatív Spp1 c-ChIP kötőhely (IRC7) illesztése az exponenciális modell alapján. Az egyenletből a turnover ráta kifejezhető. *E-F.* A c-ChIP kísérletben illesztett 977 Spp1 kötőhely turnover rátája (y tengely) a Mer2 ChIP értékek ("occupancy") és H3K4me3 ChIP értékek függvényében. A klaszterelemzés során azonosított kinetikus Spp1 frakciókat (megjelenő, eltűnő, állandó) piros, kék, és zöld halmazokként jelöltük.

Elemzésünkből kiderült, hogy az Spp1 c-ChIP kötőhelyek markánsan szétválaszthatók turnover rátájuk alapján, amelyek nagyban átfednek az Spp1 kötőhelyek kinetikus viselkedésük szerinti klasztereivel (megjelenő, eltűnő, konstant Spp1 peak-ek). Megállapítottuk, hogy minél erősebb egy adott kromatin régió Mer2-asszociációja, annál kisebb az Spp1 kicserélődési rátája (**27/E ábra**), ami azt jelenti, hogy a Mer2-gazdag területek stabilan kötik az Spp1-et, illetve az Spp1 Mer2-axiális helyekkel való kapcsolata lecsökkenti annak mobilitását. A hagyományos ChIP-seq adatokkal összhangban, a "megjelenő" Spp1 frakció mutatta a legnagyobb Mer2-asszociációt és legkisebb turnover rátát (piros halmaz), míg az "eltűnő" Spp1 peak-ek pont fordítva viselkedtek (kék halmaz). A H3K4me3 módosítás kapcsán a Mer2-vel kapott összefüggéseknek pontosan az ellentettje derült ki, vagyis minél magasabb a H3K4 trimetiláció mértéke, annál nagyobb Spp1 turnover rátája (**27/F ábra)**. Itt az "eltűnő" Spp1 frakció (kék halmaz) mutatta a legnagyobb H3K4me3 szintet és legkisebb

turnover rátát, a "konstans" Spp1 peak-ek ehhez hasonlóan viselkedtek (zöld halmaz) de az átlagos turnover ráta kisebb volt, míg az "megjelenő" Spp1 kötőhelyek (piros halmaz) ellenkezően viselkedtek. Eszerint az Spp1 rendkívül dinamikus a H3K4me3-gazdag régiókban, az Spp1 peak-ek és a H3K4 trimetiláció szintje pedig egyenesen arányos az Spp1 kicserélődési rátájával. Összefoglalva, az Spp1 kötőhelyek differenciális turnover értékei kétféle Spp1 pool jelenlétét bizonyítják a meiózis során, amelyek markánsan eltérő asszociációt mutatnak a H3K4me3 modifikációval (Set1C-függő frakció) és a meiotikus DSB-komplexxel (Set1C-független frakció).

A c-ChIP eredményeket kvantitatív mikroszkópos módszerekkel validáltuk élő (nem-fixált) meiotikus sejtekben. Ehhez az Spp1-et és Set1-et GFP fluoreszcens riporterrel jelöltük az N-terminálisukon (GFP-Spp1 és GFP-Set1) majd FRAP és FCS technikákkal megmértük a mobilitásukat (**28. ábra**).

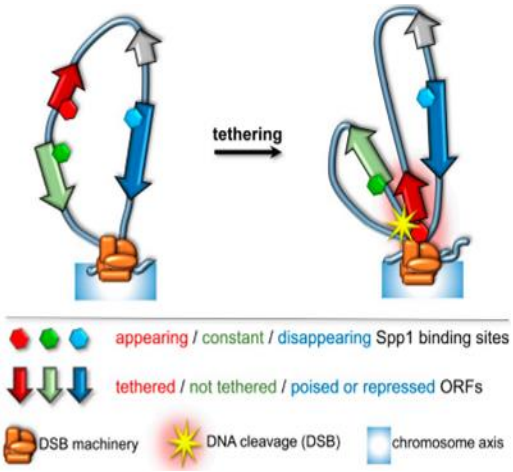


28. ábra Az Spp1 és Set1 kinetikai vizsgálata FRAP és FCS technikával. *A.* GFP-Spp1 (szürke), GFP-Set1 (sárga), és GFP (zöld) FRAP görbék különböző meiotikus időpontokban (SPM 1-6h). A fluoreszcens jel 50 másodpercen belül eléri a plató fázist. A piros vízszintes vonal feletti rész jelzi az immobilis hányadot. *B.* Az FCS mérésből származtatott autokorrelációs függvények kiértékelése. A bal oldani panel a GFP (zöld), GFP-Set1 (lila), és GFP-Spp1 (sárga) molekulákra számolt diffúziós állandók eloszlását mutatja. A jobb oldali panel a Stokes-Einstein egyenletből számított látszólagos molekulatömegek (Mapp) eloszlását mutatja a Set1 és Spp1 molekulákra vonatkozóan. A Set1C tömegét pirossal jeleztük. Statisztikai szignifikanciát Mann-Whitney próbával számoltunk ($p<0.05$).

A FRAP kísérletben a teljes sejtmagot kioltottuk, majd fluoreszcens visszatérést vizsgáltunk a meiózis első öt órájában SPM sporulációs táptalajon. A GFP-Spp1 és a GFP-Set1 molekulák hasonló visszatérési kinetikát követtek, a fluoreszcencia intenzitások nagyjából 50 másodpercen belül elérték a szaturációs plató fázist. A FRAP jel ~50%-a sosem tért vissza a kezdeti fluoreszcencia kioltást követően, amely azt jelzi, hogy az Spp1 és a Set1 molekulák kb. fele rendkívül szorosan kötődik a kromatinhoz (vagy egy másik nagy-molekulatömegű komplexhez) a meiózis során. Ez a stabil hányad reprezentálja az ún. immobilis frakciót. A két fehérje diffúziós paramétereinek kvantifikálásához FCS-el végeztük, amely szerint az Spp1 diffúziós együtthatója (D) szignifikánsan kisebb volt a Set1-hez képest. Ez egyértelműen az Spp1 kisebb nukleális mobilitására utal. Ezután megbecsültük a GFP-Spp1 és a GFP-Set1 látszólagos molekulatömegét a Stokes-Einstein egyenlet alkalmazásával gömb alakú objektumokra ($D = kT/(6\pi\kappa\eta R)$) és feltételezve a molekulatömeg arányosságát R^3-al (ahol D a diffúziós együttható, k a

Boltzmann-állandó, T az abszolút hőmérséklet, η a viszkozitás, R pedig a molekulasugár). Számításaink szerint a GFP-Set1 látszólagos molekulatömege (M_{app}= 399 kDa) kb. háromszorosa volt a fehérje valódi molekulatömegének (M_{Set1}=124 kDa), amely pontosan megegyezik a Set1C/COMPASS holokomplex tömegével ($M_{COMPASS}$= 379 kDa). A Set1 fehérje tehát jellemzően a Set1 komplex részeként működik a sejtekben, amely egy független mikroszkópos módszerrel alátámasztja a korábbi biokémiai/proteomikai mérések eredményeit (Trésaugues et al. 2006). Ezzel szemben a GFP-Spp1 kb. 43-szor nagyobb molekulatömegű volt (M_{app}= 1764 kDa) a fehérje valós molekulatömegéhez képest (M_{Spp1}= 41 kDa), amely arra utal, hogy az Spp1 egy hatalmas megakomplexhez kapcsolódik a meiózis során, amely sokkal nagyobb, mint maga a Set1 komplex. Az Spp1 látszólagos és valódi molekulatömege közötti hatalmas különbség nehezen igazolható pusztán diffúzibilis faktorokkal történő nukleoplazmatikus kölcsönhatásokat feltételezve; a kromatinnal történő tranziens asszociációk (Mer2-höz kötődés) viszont teljesen megmagyarázzák az Spp1 diffúziós viselkedését, amely összecseng a ChIP és c-ChIP kísérletek eredményeivel.

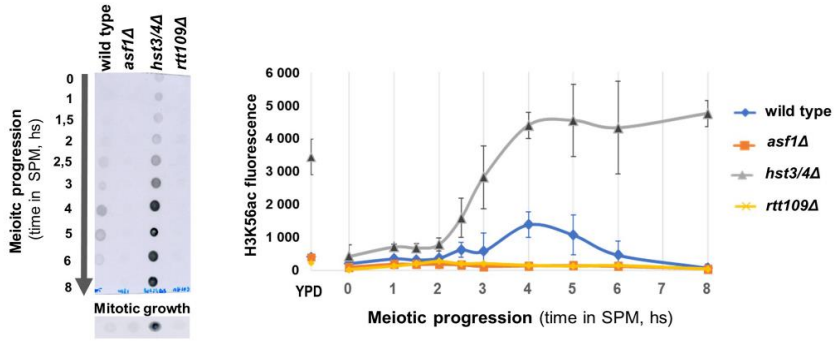
Összegezve a fentieket, a Set1 és Spp1 kromatin dinamikája és diffúziós tulajdonságai jelentős eltéréseket mutat a meiózis során, amely oka az Spp1 nagy-molekulatömegű makromolekuláris komplexekhez, feltehetően a kromoszóma axiális részein található DSB-komplexhez, történő kötődése. Az Spp1 a meiózis során Set1C-független működésre vált, amely dinamikai tulajdonságainak megváltozásában és kromoszómális relokalizációjában tettenérhető. Az Spp1 dinamikájával kiegészített új hurok-tengely modell a **29. ábrán** látható, amely bemutatja a kísérletesen azonosított dinamikus Spp1-frakciók kromoszóma tengelyhez történő kapcsolódását és a rekombinációs iniciációs helyek aktiválását.



29. ábra Az Spp1 dinamikájával kiegészített hurok-tengely modell. Az Spp1 kötőhelyek egy része (piros hatszögek) a kromoszóma tengelyhez kötődik (piros ORF-ek), amelyek megfelelnek a ChIP-seq kísérletekben azonosított "megjelenő" Spp1 frakciónak. A megjelenő Spp1 peak-ek kölcsönhatásba lépnek a Mer2-vel és így a DSB hotspot-okat a tengelyhez kötik, ahol a Spo11 elhasítja őket (sárga csillag). Az "állandó" Spp1 kötőhelyek (zöld hatszögek) változatlanok maradnak a meiózis alatt, nem lépnek kölcsönhatásba a Mer2-vel, és a Set1C-vel kolokalizálnak a gének 5' UTR régiójában (zöld ORF-ek). Az "eltűnő" Spp1 peak-ek (kék hatszögek) főként az RPG és snoRNS génekhez kapcsolódnak (kék ORF-ek). Forrás: (Karányi et al. 2018)

A hiszton H3K56 acetiláció szerepe a meiotikus homológ rekombináció iniciációjában

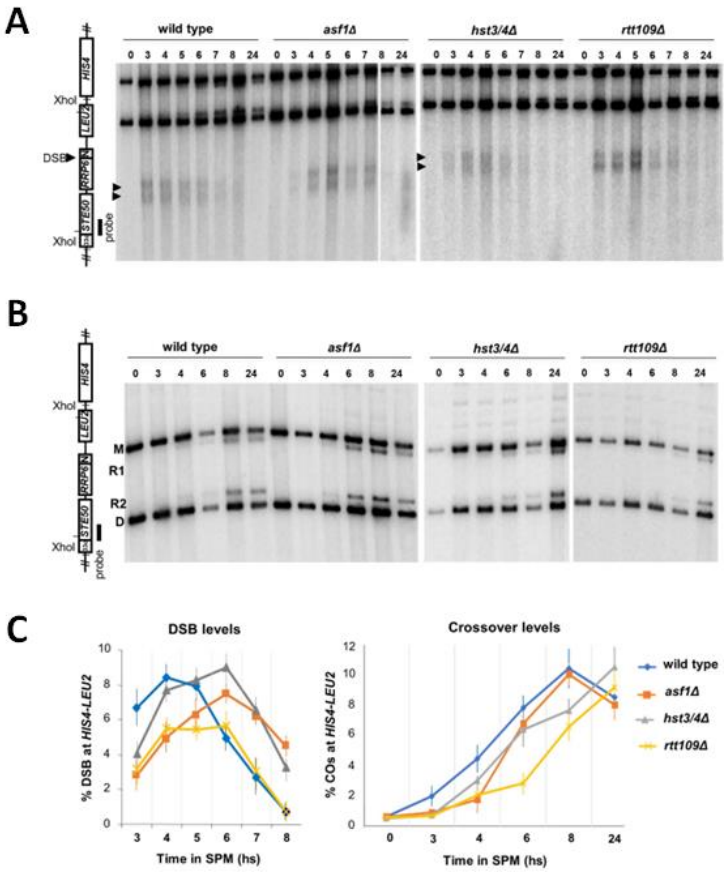
A hiszton H3K4 trimetiláció a genom számos pontján beindíthatja a homológ rekombinációt, azonban más hisztonmódosítások is szükségesek a folyamat flexibilitásához és a rekombináns haplotípusok sokféleségéhez (Székvölgyi and Nicolas 2010). További hisztonmodifikációk és alternatív epigenetikai útvonalak azonosítása tehát segíthet megérteni a meiotikus crossover mintázatok kialakulását és azok evolúciós plaszticitását. A legtöbb kémiai módosítás a hiszton fehérjék N-terminális részére koncentrálódik, amely nem változtatja meg a nukleoszómák szerkezetét (White et al. 2001; Biswas et al. 2011). A hisztonok központi ("core") doménjén található poszttranszlációs modifikációk azonban jelentős nukleoszóma-szerkezet változásokat indukálhatnak (Luger et al.; Biswas et al. 2011). Kísérleteinkben ezért a H3 hiszton globuláris doménjén elhelyezkedő lizin 56 acetiláció szerepére összpontosítottunk (H3K56ac), amely a DNS be- és kilépési pontjainál destabilizálhatja a hiszton/DNS kölcsönhatásokat (Buning and Van Noort 2010; Simon et al. 2011). A H3K56ac egy gyorsan kialakuló és eltűnő ("tranziens") kromatin szignál, amely mitotikus sejtekben erősen korrelál a DNS-replikációval és a transzkripció során megfigyelhető hiszton leválással ("eviction") (Rufiange et al. 2007; Watanabe et al. 2010, 2013). Élesztőben kimutatták, hogy a H3K56ac modifikáció lehetővé teszi a nukleoszómák gyors össze- és szétszerelését a replikáció és transzkripció során, az Asf1 hiszton chaperon pedig (amely az Rtt109 H3K56-acetiltranszferázt aktiválja (Abshiru et al. 2013) szükségesnek bizonyult a meiotikus reproduktív kapacitás kifejlődéséhez élesztőben és egérben (Recht et al. 2006; Govin et al. 2010; Messiaen et al. 2016). A H3K56 acetiláció kapcsolata a meiotikus rekombinációval azonban feltáratlan maradt. Kísérleteinkben ezért funkcionális vizsgálatokat végeztünk egy sor élesztő mutánsban, amelyekben lecsökkentettük vagy megnöveltük a H3K56ac szintjét a meiotikus differenciáció során (**30. ábra**).



30. ábra Eltérő H3K56 acetilációt mutató élesztő mutánsok. Anti-H3K56ac antitesttel végzett reprezentatív dot blot kísérlet, amelyben vad típusú sejtekben és izogénikus *asf1Δ*, *rtt109Δ*, és *hst3/4Δ* mutánsokban detektáltuk (bal panel) és kvantifikáltuk (jobb panel) a H3K56ac szignált a különböző meiotikus időpontokban. A mérés 3 független kísérlet átlagát mutatja. Hibasáv: SEM.

Dot blot méréseink szerint a vad típusú sejtek H3K56ac-szintje a DSB képződés időpontjában érte el globális maximumát (4-5h SPM), majd ezután gyorsan alapszintre csökkent (**30. ábra**). A H3K56

acetiláció szinte teljesen eltűnt az *asf1Δ* és *rtt109Δ* mutánsokban, ami összhangban van az Asf1 (hiszton chaperon) és Rtt109 (K56-specifikus acetiltransferáz) ismert biokémiai szerepével a H3K56 acetiláció kromatinra írásával kapcsolatban (Tsubota et al. 2007). A K56ac szignál jelentősen megnőtt a *hst3/4Δ* kettős-mutánsban, amely megakadályozza a H3K56ac deacetilációját (Celic et al. 2006).

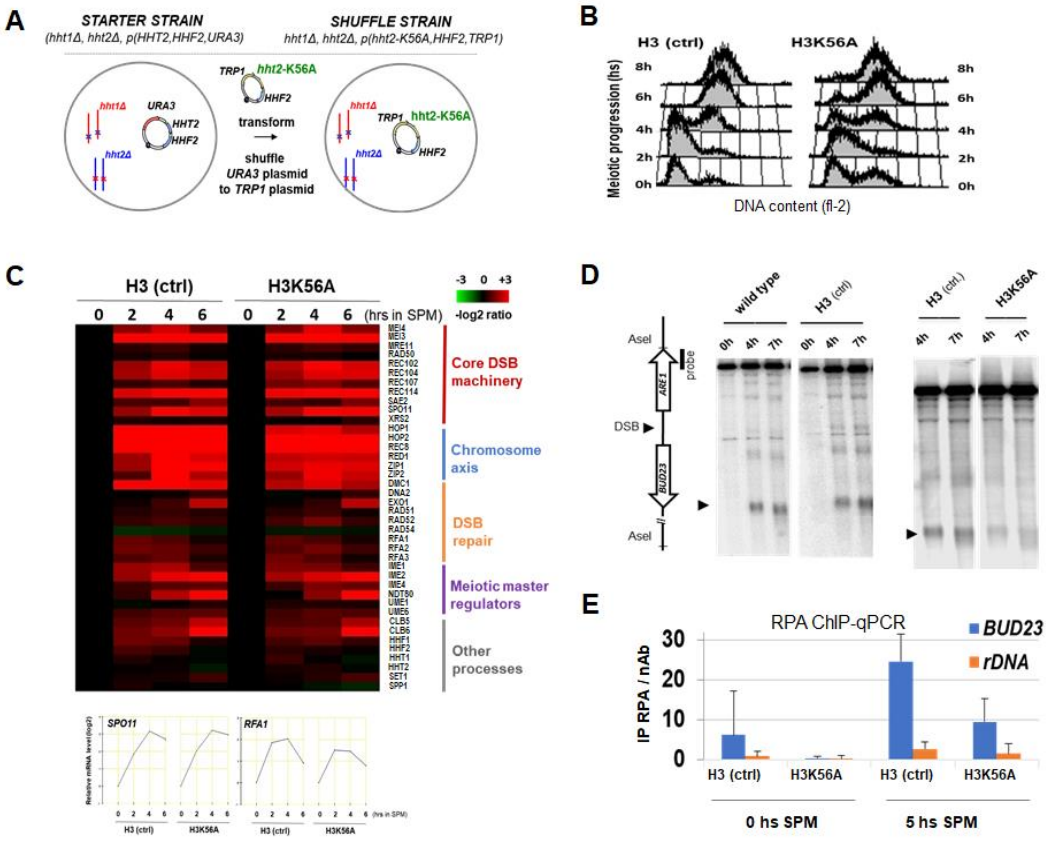


31. ábra Meiotikus DSB-k vizsgálata H3K56ac mutánsokban. A Southern blot méréseket az ábrán jelzett deléciós mutánsokban végeztük PCR-ral előállított végjelölt radioaktív próbákkal. Az 'A' paneleken látható DSB band-ek (nyilak) kvantifikációja a 'C' panel bal oldalán található, míg a 'B' panelen mutatott R1/R2 crossover rekombináns molekulák kvantifikációját a 'C' panel jobb oldali ábrája mutatja. A DSB-k és R1/R2 rekombinánsok mennyiségét a szülői sávokra (M: mom, D: dad) normáltuk. A Southern blot próbát a gél bal oldalán jelzett *HIS4-LEU2* régió restrikciós fragmentumának legvégére terveztünk.

(Mitotikus körülmények között a *hst3/4Δ* mutáns kivételével nem tudtunk H3K56 acetilációt kimutatni ezzel a módszerrel, amely valószínűleg a H3K56 acetiláció rendkívül tranziens természete miatt volt). Ezután a *HIS4-LEU2* "mesterséges" rekombinációs hotspot régióban Southern blot módszerrel térképeztük és kvantifikáltuk a meiotikus DSB-ket és a crossover rekombinánsok mennyiségét. (Az *E.coli* eredetű *HIS4-LEU2* konstrukciót kiterjedten használják DSB és crossover térképezésre és kvantifikációra (Xu and Kleckner 1995), amely akár 50%-os DSB-gyakoriságot is mutathat. Ez praktikusan 100%-os vágási hatékonyságot jelent, mivel mindig csak az egyik homológ kromoszóma törik el a rekombináció során.) Méréseink szerint az összes mutáns fiziológiás DSB-szintet produkált (amely gyakorisága a vad típusú mintákkal összevethető volt; 6-8% a görbék globális maximuma alapján), azonban az *asf1Δ* és *hst3/4Δ* mutánsokban a DSB-k módosult kinetikával jelentek meg és tűntek el, kb. 60 perces időbeli késéssel. Fontos kiemelni, hogy a mutánsokban azonosított meiotikus DSB-k erősen rekombinogének voltak (lásd az R1/R2

rekombinánsokat; **31/B-C ábra**), vagyis a *HIS4-LEU2* szakaszon keletkező DNS törések a "természetes" repair útvonalakon processzálódtak.

A modifikálható lizin 56 aminosav oldallánc (K56) közvetlen szerepének teszteléséhez egy funkcionális kísérletet végeztünk (plasmid shuffle assay), amelyben egy nem-módosítható H3K56A mutáns hiszton molekulát expresszáltunk a H3 hiszton fehérje egyedüli forrásaként a sejtekben (**32/A ábra**). A vad típusú sejtekhez hasonlóan a H3K56A mutánsok szinkron-meiózist és normál transzkriptom dinamikát mutattak (**32/B-C ábra**); a DSB-fehérjék, a kromoszóma tengely fehérjék és a DNS hibajavító fehérjék mind megfelelően íródtak át mRNS-é, beleértve például a Spo11-et vagy az Rfa1-et, amelyek nélkülözhetetlenek RPA ChIP-alapú DSB-térképezéshez (ld. később). A DSB-ket Southern blot-tal térképezve háromszoros csökkenését figyeltünk meg a H3K56A mutánsban a *BUD23-ARE1* hotspot régióban (**32/D ábra**), amit ezután független RPA ChIP-qPCR módszerrel is validáltunk (**32/E ábra**).

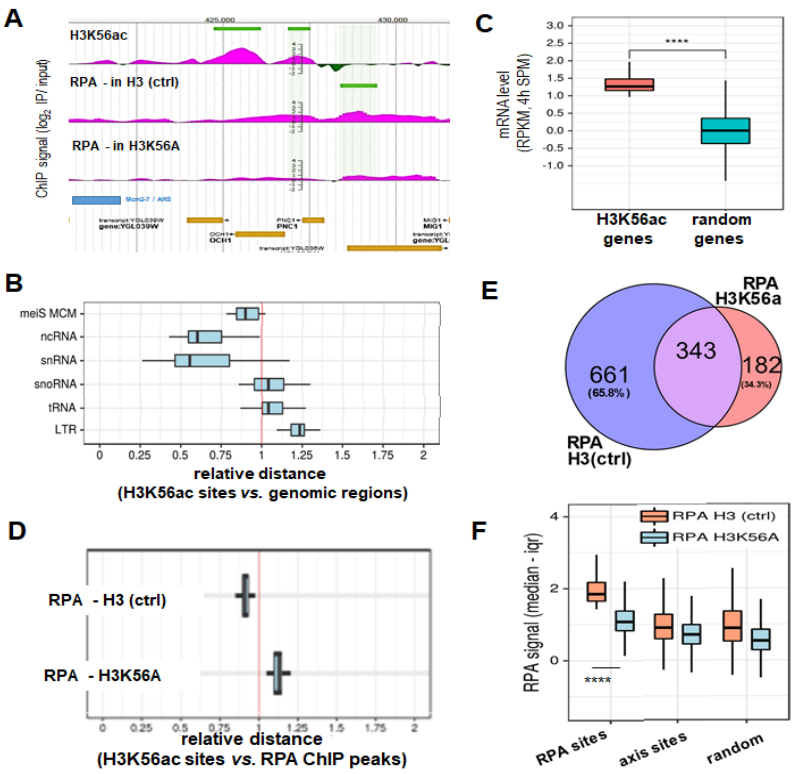


32. ábra Meiotikus DSB-k vizsgálata H3K56A mutánsban. *A.* A plazmid shuffle (csere) módszer elve. A hiszton gén klasztereket kiütjük és plazmidról expresszált hiszton génekről komplementáljuk a mutáns fenotípust. Ezután a vad típusú hisztonokat kódoló plazmidokat kicseréljük a mutáns H3K56A-t hordozó plazmidra. *B.* A H3K56A sejtek FACS profilja a kontroll sejtekhez hasonló kinetikát mutat, amely szerint a meiózis progressziója nem sérült. *C.* Meiotikus transzkriptom vizsgálata 8x15K Agilent microarray-vel. 0-6h SPM időpontokban megmértük az mRNS-ek mennyiségét kontroll és H3K56A mutáns sejtekben. A hőtérkép a 0h SPM időponthoz képest mért relatív mRNS változást mutatja a különféle géncsoportokban. Az alsó panelen külön kiemeltük a *SPO11* és *RFA1* gének transzkripciós dinamikáját. *D.* DSB térképezés

Southern blot-tal a *BUD23-ARE1* hotspot-ban. A baloldali gél a vad típusú sejteket hasonlítja össze a kontroll hiszton shuffle sejtekkel (amelyek vad típusú H3 hisztont expresszálnak) Nincs eltérés a DSB fenotípusban, vagyis a plazmid shuffle rendszer fiziológiásan viselkedik. A jobb oldali gél a kontroll és H3K56A mutáns sejteket hasonlítja össze. A DSB-k mennyisége csökkenést mutat a mutánsban. *E.* RPA-ChIP qPCR mérés a *BUD23* hotspot és *rDNA* coldspot régiókban 0h SPM és 5h SPM időpontokban. Az RPA jel mennyisége (antitest-nélküli ()nAb) kontrollra normálva) megnő a DSB keletkezés időpontjában (5h SPM), viszont a H3K5A mutánsban szignifikáns csökkenés figyelhető meg a kontroll sejtekhez képest. Hibasáv: SEM.

(Az RPA-ChIP a DSB végek visszavágása során keletkező ssDNS-t borító Rfa1 immunprecipitációján alapul.) RPA ChIP-qPCR mérésünkben a kontroll sejtekben szignifikánsan emelkedett RPA szintet detektáltunk a DSB képződés tipikus időpontjában (5h SPM) a *BUD23-ARE1* hotspot-ban, amely jelentősen lecsökkent a H3K56A mutánsban. Ezzel szemben a rekombinációsan "hideg" rDNS szakaszon nem detektáltunk RPA változást (**32/E ábra**). Mindez a modifikálható H3 lizin 56 oldallánc közvetlen szerepére utal a meiotikus DSB-k keletkezésében.

A K56 acetiláció és a DSB hotspotok kapcsolatát H3K56ac ChIP-Chip és RPA ChIP-Chip módszerrel vizsgáltuk a teljes genomban, amelyekben anti-H3K56ac és anti-H3 antitestek felhasználásával immunprecupitáltuk a hiszton fehérjéket, majd meghatároztuk a H3K56ac nukeloszómák dúsulását az össz H3-hoz képest, illetve ugyanezt elvégeztük anti-Rfa1 antitestekkel kontroll és H3K56A mutáns sejtekben, amellyel a meiotikus DSB-ket jelöltük meg (**33/A ábra**). Eredményeink szerint a H3K56 acetilált hisztonok szignifikánsan dúsultak a meiotikus replikációs origók (meiS MCM), ncRNS-ek és snRNS-ek környezetében (**33/B ábra**), illetve a H3K56 acetilált gének átlagos magasabb expressziós szintje magasabb volt a véletlenhez képest (**33/C ábra**). A kapott összefüggések összecsengenek a H3K56 acetiláció mitotikus DNS replikációban és transzkripcióban betöltött szerepével. A meiotikus DSB-k tekintetében H3K56ac hisztonok preferenciálisan dúsultak az RPA ChIP peak-ek közvetlen közelében (a DSB-k mellett) a kontroll sejtekben, míg nem mutattak asszociációt a H3K56A mutánsban azonosított ("megmaradó") RPA ChIP peak-ekkel (**33/D ábra**). Az RPA ChIP peak-ek száma szignifikánsan lecsökkent a H3K56A mutánsban (**33/E ábra**), mivel a mutáció a kontroll sejtekben detektált 1004 RPA peak 65,8%-át eliminálta (661 peak), 343 RPA-helyet nem érintett (metszet halmaz), míg 182 esetben "ektópikus" RPA peak-ek keletkeztek, amelyek nem voltak jelen a vad típusú sejtekben. Fontos kiemelni, hogy az RPA szignál K56A mutánsban tapasztalt csökkenését nem detektáltuk kromoszóma tengely (Mer2) kötőhelyeken és véletlenszerűen kiválasztott (random) genomi pozíciókban (**33/F ábra**), amely az RPA kötőhelyeken kapott különbség specificitását mutatja. Az RPA peak-ek számának jelentős genomi csökkenése összecseng a *BUD23-ARE1* régióban kapott Southern blot eredményekkel (**32/D ábra**), amely összességében H3K56 oldallánc közvetlen szerepét bizonyítja a rekombináció iniciációjában.

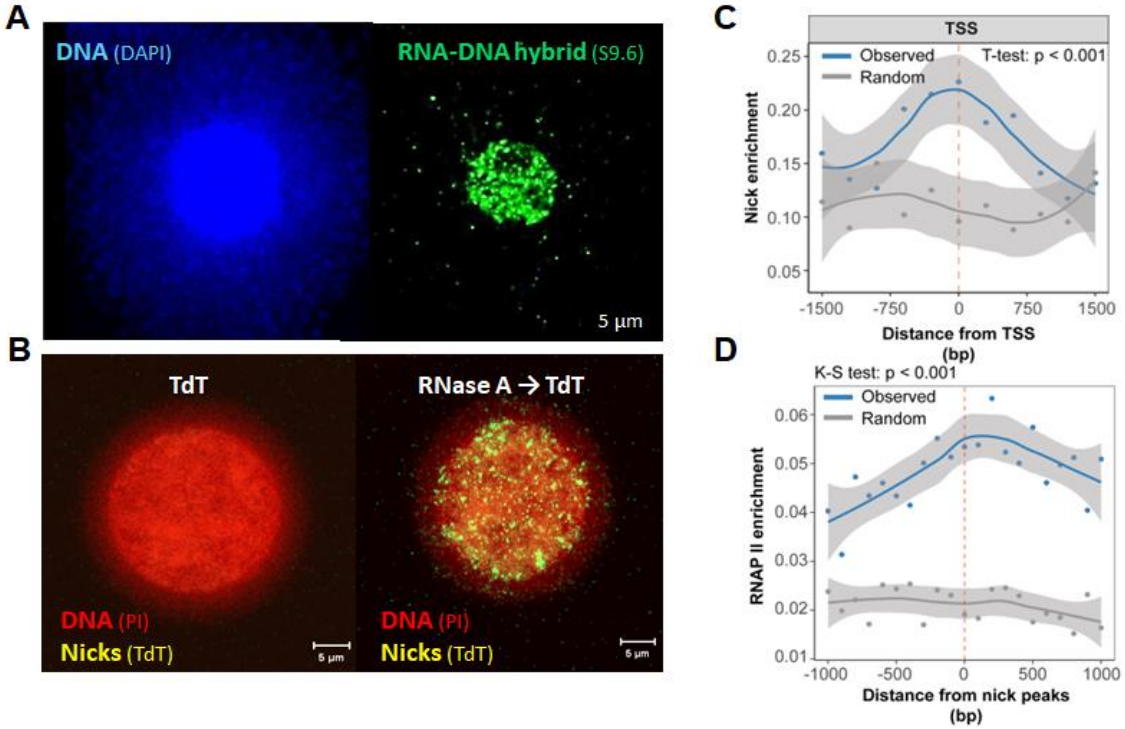


33. ábra H3K56ac és RPA ChIP-chip térképezés az élesztő genomban. *A.* JBrowse pillanatkép, amely a H3K56 acetiláció (H3-ra normálva) és a kontroll és H3K56A mutáns sejtekben mért RPA ChIP profilt mutatja. Vízszintes zöld vonalak jelzik az azonosított peak-eket. *B.* A H3K56ac ChIP peak-ek relatív távolsága a jelzett annotációs elemekhez képest. Függőleges piros vonal mutatja az átlagos random távolságot, az ettől "balra" elhelyezkedő eloszlás szignifikáns asszociációt, míg a "jobbra" tolódó eloszlás szignifikáns távolságot jelez a H3K56ac és a vizsgált genomi elemek között. *C.* A H3K56ac hisztonokat tartalmazó gének magasabb transzkripciós aktivitást mutatnak a véletlenszerűen kiválasztott génekhez képest *D.* Ugyanaz, mint a 'B' ábrán, csak itt a H3K56ac peak-ek relatív távolságát vizsgáltuk az RPA ChIP peak-ekhez képest. *E.* Venn diagram, amely a kontroll és H3K56A mutánsban azonosított RPA peak-ek átfedését mutatja. *F.* Az RPA peak-ek, Mer2 kötőhelyek (axis) és véletlenszerűen generált peak-ek felett mért RPA ChIP szignál eloszlása. Az RPA jel szignifikánsan csökken a H3K56A mutánsban az azonosított RPA peak-ek felett, viszont nem különbözik a kontroll sejttől a kromoszóma tengely (axis) és random kötőhelyek felett. Ez az RPA ChIP módszer specificitását bizonyítja.

A H3K56-acetilezett hisztonok tehát fiziológiás szintre állítják meiotikus DNS törések mennyiségét, mintegy finomhangolva a rekombinációs forrópontok működését. A pontos molekuláris mechanizmus azonban még tisztázásra vár. Adataink alapján úgy gondoljuk, hogy a H3K56ac módosítás a DSB hotspotokat szegélyező nukleoszómák stabilitását vagy turnover-ét befolyásolja (Kaplan et al. 2008) amely jelentős megváltozása (pl. a H3K56A mutánsban) hatással van a H3K4me4-Spp1-Mer2 hurok-tengely komplex keletkezésére és a Spo11 reakció hatékonyságára. Fontos kiemelni, hogy a H3K56A mutánsban nem tapasztaltunk differenciális génexpressziót a kontroll tözshöz viszonyítva, ezért a K56 oldallánc közvetlen hatást gyakorol a DSB képződésre, amely független a transzkripciótól. A H3K56ac tehát valószínűleg közvetlenül elősegíti a H3K4me3, Spp1 és Mer2 kölcsönhatását és a hurok-tengely kapcsolat kialakulását, amelyet nagyban alátámaszt a H3K56ac és H3K4me3 közelmúltban azonosított episztatikus kapcsolata (Voichek et al. 2018). Feltételezzük, hogy e komplex funkcionális interakciók és génkölcsönhatások a 3D kromatin szerkezeteben tettenérhetők, amely genomkonformáció leképezéssel és mutáns analízissel tesztelhető. A Hi-C-alapú módszerek genetikai analízissel történő kombinációja ma már egy reális lehetőség a fenti kérdések megválaszolására.

Kromoszómális R-hurkok és R-hurok regulátorok molekuláris vizsgálata

Az R-hurok struktúrák rekombinogén és mutagén hatása régóta ismert, azonban a DNS törésekkel való közvetlen fizikai kapcsolatukat nem sikerült bizonyítani. Egyszál DNS törésekkel (nick-ekkel) történő asszociációjuk egy korai fluoreszcens mikroszkópos vizsgálatunkban merült fel (Székvölgyi et al. 2007), amikor az S9.6 anti-RNS-DNS hibrid antitesttel *in situ* R-hurok struktúrákat mutattunk ki különféle emberi és élesztő sejtekben. Kísérletünkben az R-hurkok jellegzetes, punktált jelölési mintázatot mutattak (**34/A ábra**), amely hasonló volt a terminális transzferázzal (TdT), *E.coli* DNS polimeráz I holoenzimmel, vagy Klenow fragmentummal jelölt nick-ek mintázatához (**34/B ábra**).

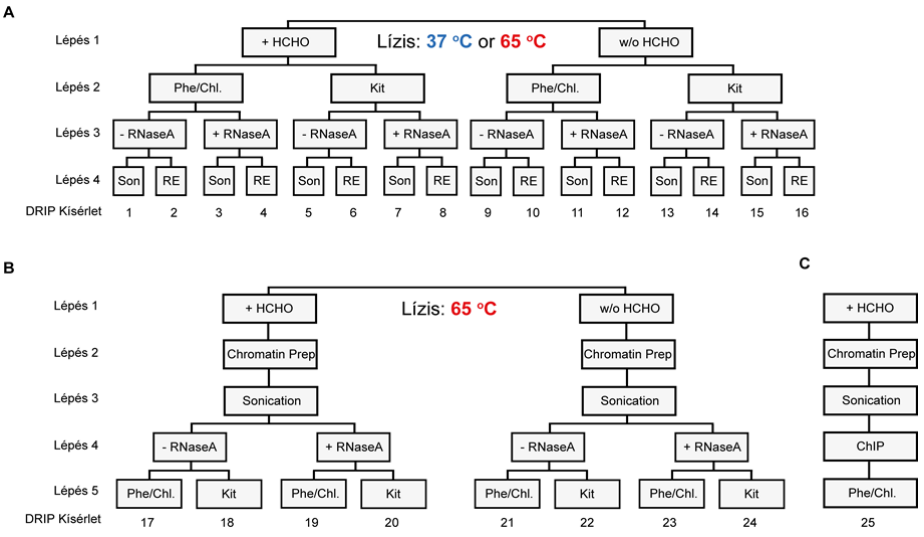


34. ábra RNS-DNS hibridek és nick-ek detektálása *in situ* fluoreszcens mikroszkópiával és microarray-vel. *A.* Sejtmag "halo" preparátumon az RNS-DNS hibridek S9.6 antitesttel jelölhetőek (zöld), amelyek jellegzetes punktált mintázatot adnak a halo központi részén, amelyet "mátrixnak" neveznek. *B. In situ* nick jelölés terminális transzferáz (TdT) enzimmel, amely a DNS törések 3'OH csoportjához épít egy homopolimer biotin-dUTP farkat. (A Klenow és DNAP I jelöléseket ezen az ábrán nem mutatjuk). A TdT jelölés önmagában nem hatékony (bal panel), a hatékony nick jelöléshez RNázA előkezelés szükséges (jobb panel). A jelölt nick-ek hasonló punktált mintázatot mutatnak a "mátrixon", mint az RNS-DNS hibridek. *C.* Nick-ChIP microarray kísérlet élesztőben. A detektált nick-ek szignifikánsan dúsulnak a transzkripciós start helyeken (TSS). Kék vonal: megfigyelt jel, szürke vonal: random szignál. A halványszürke sávok a 95%-os konfidencia intervallumot mutatják. *D.* A nick-ek genomi pozíciója szignifikáns asszociációt mutat az RNS polimeráz II-vel (RNAPII). A jelölés ugyanaz, mint a C panelen. (A fenti kísérleteket Prof. Szabó Gáborral kollaborációban (Debreceni Egyetem ÁOK, Biofizikai és Sejtbiológiai Intézet) végeztük.)

Fontos megfigyelés volt, hogy a TdT és Klenow enzimek kizárólag ribonukleolítikus vagy proteolitikus előkezelés után jelölték hatékonyan a nick-eket (pl. RNázA, RNázH, lúg, vagy proteináz K), amely a folytonossághiányok ribonukleopreotein struktúrák általi maszkírozottságára utalt. A

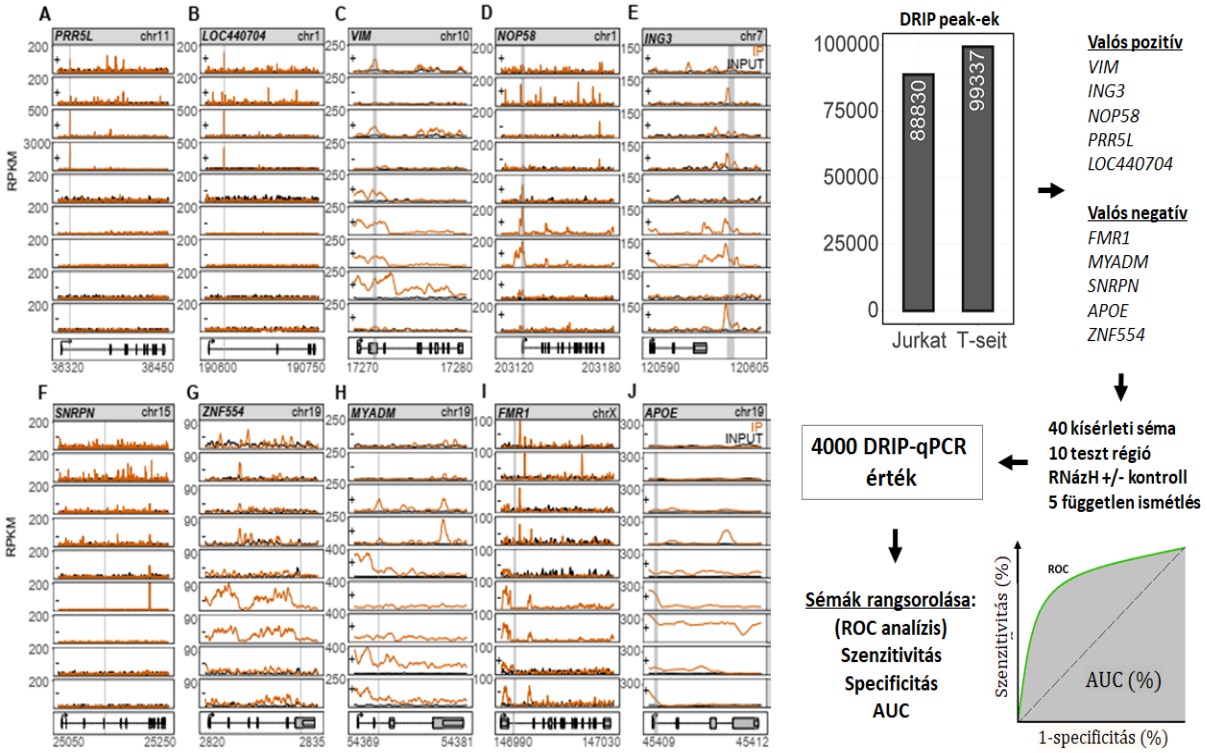
nick-ek genomi eloszlását az általunk bevezetett nick-ChIP módszerrel részletesen kielemeztük, amelyben limitált *in situ* nick transzlációval biotinilált nukleotidákat építettünk a nick-ek 3'OH csoportjához, melyeket anti-biotin antitestekkel immunprecipitálva teljes-genomi microarray-hez hibridizáltunk (Hegedüs et al. 2018). (A módszert élesztőre fejlesztettük ki, amit azóta humán sejtekre is adoptáltunk). Eredményeink szerint a nick-ek elsősorban a transzkripciós start helyeken (TSS) dúsultak és jelentős átfedést mutattak az RNS polimeráz II (RNAP II) enzim genomi kötőhelyeivel (**34/C-D ábra**). Mindez arra utal, hogy a transzkripció iniciációja során nagyszámú fiziológiás DNS törés (zömében nick) keletkezik, amelyeket valamilyen R-hurok-tartalmú ribonukleoprotein komplex hatékonyan elfed a DNS repair rendszer elől. Emiatt a fiziológiás nick-ek nem rekombinogének.

A fenti modell teszteléséhez és további funkcionális kísérletek elvégzéséhez azonban megbízható R-hurok térképezési módszer szükséges, amely hatékonyan képes az R-hurok struktúrákat azonosítani a genomban. Ez kulcsfontosságú a genom stabilitását szabályozó folyamatok és a DNS léziók kromatinszerkezeti aspektusainak megértéséhez. Az időközben bevezetett DRIP módszer (Ginno et al. 2012) rendre inkonzisztens adatokat generált, amely a publikált R-loop eredmények jelentős variabilitásában tettenérhető. Ezért célul tűztük ki egy univerzális genomikai térképezési módszer kidolgozását, amely az "alap" DRIP módszer torzításait kikerülve alkalmas az R-hurkok nagypontosságú azonosítására bármely modellorganizmusban.



35. ábra Osztályozó változók bevezetése az RNS-DNS hibrid immunprecipitáció torzításának becslésére. Nagyszámú publikus DRIP protokoll alapján egy sor osztályozó változót (kísérleti sémát) terveztünk, amelyben megvizsgáltuk a formaldehid-fixálás hatását (1. lépés), a nukleinsav izolálás hatását (2. lépés), a szabad RNS eltávolításának hatását (3. lépés) és a nukleinsav fragmentálás módját (4. lépés), illetve a 1-16. sémában vizsgáltuk még a hőmérséklet hatását a sejtlízis során (37°C vs. 65 °C). A hőmérsékleti változót az 'A' panelen nem tüntettük fel külön DRIP kísérletként a könnyebb áttekinthetőség kedvéért, ez azonban plusz 16 változót jelent. A 'B' panelen látható 17-24 DRIP sémákban nem nukleinsav preparátumon, hanem kromatinon végeztük a mérést. A sejtlízis hőmérséklete rendre 65 °C volt, amely a HCHO-keresztkötések revertálásához szükséges. Ez az elrendezés a klasszikus ChIP kísérlethez hasonló, amely a 25. sémán látható ('C' panel). Az összes kísérleti sémában alkalmaztunk *in vitro* RNázH-emésztést a S9.6 immunprecipitáció negatív kontrolljaként, azonban az ábrán ezt nem tüntettük fel az egyszerűség kedvéért.

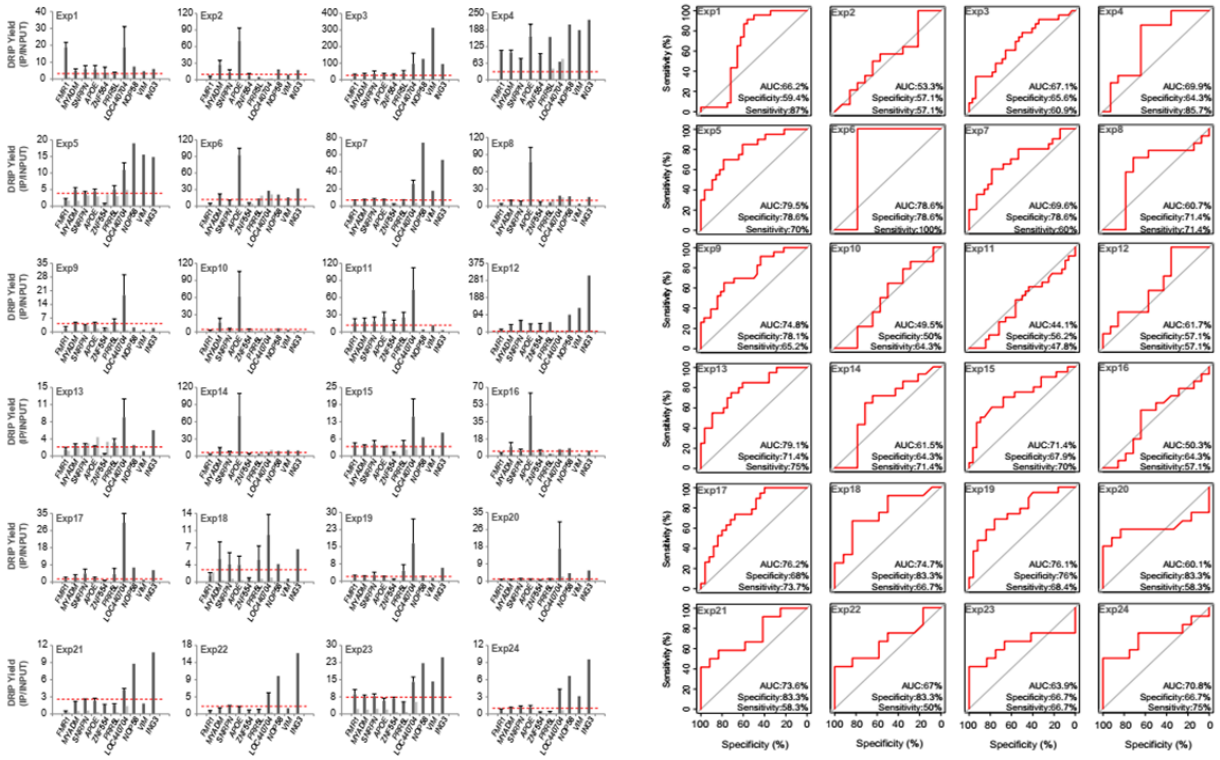
A publikált protokollokat áttanulmányozva meghatároztuk a legfontosabb kísérleti változókat (sejtfixálás, lízis hőmérséklete, nukleinsav izolálás, szabad RNS eltávolítása, genom-fragmentáció), amelyek kombinálásával 40 kísérleti sémát (osztályozó változót) terveztünk (**35. ábra**). Az osztályozó változók segítségével felmértük, hogy azok hogyan rangsorolnak különböző teszt lókuszokat R-hurok státuszuk alapján ("tréning szett"), amelyet publikus adatokból és saját mérésekből előzetesen megismertünk. Saját DRIP-seq kísérletünket kétféle humán mintán végeztük el (Jurkat T-sejt leukémia sejtvonal és CD4+ T limfociták), amelyben 88,830 és 99,337 R-hurok régiót azonosítottunk (**36. ábra**).



36. ábra Referencia R-hurok régiók meghatározása DRIP szekvenálással az osztályozó változók teljesítmény vizsgálatához. *A-J.* Genomböngésző példák Jurkat és CD4+ T sejtekben meghatározott DRIP-seq profilokra. Függőleges szürke vonal jelzi a qPCR mérések (ampikonok) pozícióját. Jobb panel: bioinformatikai elemzéssel azonosított DRIP-peak-ek a kétféle sejttípusból, amelyekből valós pozitív és valós negatív teszt szakaszokat választottunk DRIP-qPCR mérésekhez. A méréseket a 40 osztályozó változón, 10 teszt szakaszon, RNázH-kontroll mellett, 5 biológiai replikátumban végeztük el, amely kb. 4000 DRIP hozamot eredményezett. A DRIP hozamokból ROC (receiver operator characteristics) elemzéssel álnegatív és álpozitív arányokat számoltunk, amelyek alapján meghatároztuk a DRIP sémák szenzitivitását, specificitását, és a görbe alatti területet (AUC).

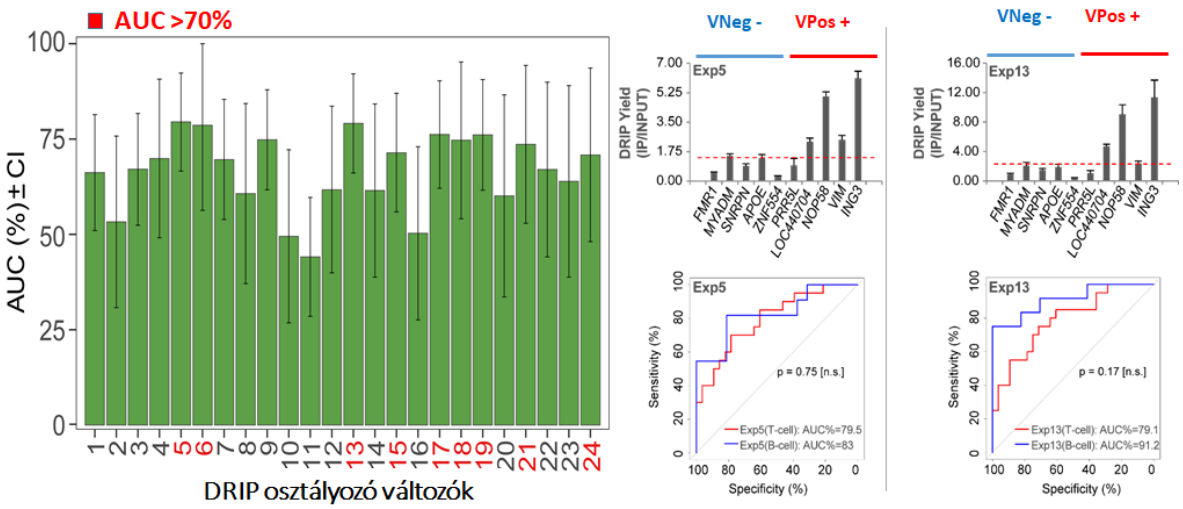
Az azonosított DRIP kötőhelyekből konszenzus R-hurok készletet állítottunk elő és valós-pozitív / valós-negatív teszt szakaszokat választottunk tréning szettként. A kiválasztott lókuszokon RNázH-kontroll mellett 5 független ismétlésben elvégeztük a 40-féle DRIP kísérletet, amely ~4000 DRIP-qPCR értéket (hozamot) eredményezett. A DRIP-qPCR adatokat ezután ROC elemzés független változójaként felhasználva meghatároztuk a 40 függő változónk (osztályozó változók) szenzitivitását

és specificitását (**36. ábra**). A **37. ábra** reprezentatív qPCR eredményeket és ROC görbéket mutat az 1-24 DRIP sémákra vonatkozóan. Azonos elemzéseket végeztük el az összes változó kapcsán, illetve eltérő sejtvonalakban is, amelyeket nem mutatunk az ábrán.



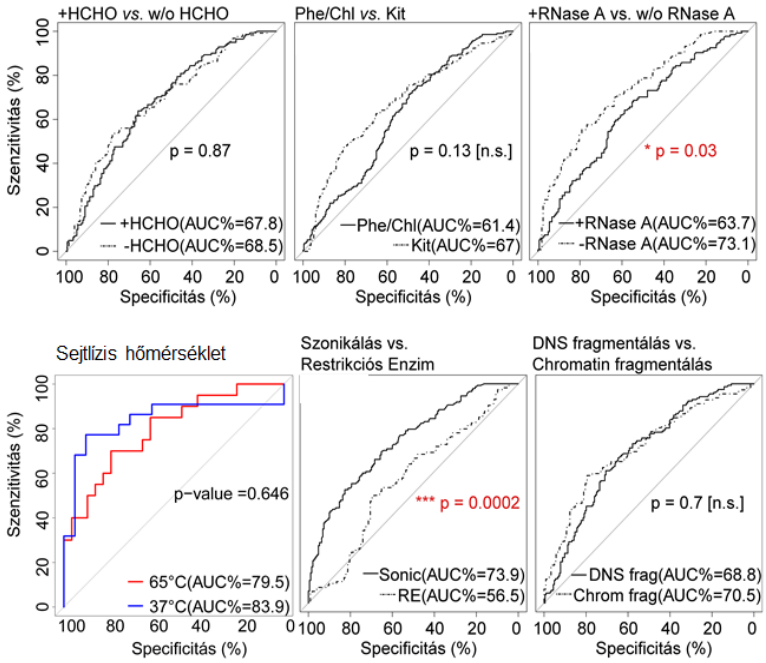
37. ábra Referencia szakaszok DRIP-qPCR és ROC analízise. Bal oldali panel: az 1-24 sémáknak megfelelő DRIP kísérletetek elvégzése után az R-hurok hozamot tíz referencia szakaszon mértük qPCR-ral. A 10 szakasz közül az első öt valós negatív régiónak, míg a második 5 valós pozitív DRIP régiónak tekinthető. Y tengely: Átlag DRIP hozam (IP/input). Vízszintes piros vonal: a ROC görbékből számított cutoff érték, amelynél a valós DRIP szignál elválik a háttértől. Jobb oldali panel: a qPCR mérésekből számított ROC görbék. AUC: görbe alatti terület. A fenti elemzést nemcsak az 1-24 változóra végeztük el, hanem az összes változóra, illetve más sejtvonalakban is, amelyeket nem tüntettünk fel az ábrán.

A ROC görbék összehasonlítása során tíz osztályozó változó esetén (5, 6, 13, 15, 17, 18, 19, 21 és 24) magas AUC értékeket kaptunk (>0,7), vagyis ezek a DRIP sémák nagy szenzitivitással és specificitással képesek az R-hurkokat azonosítani (**38. ábra**). Négy DRIP sémában 0,5 körüli AUC-t kaptunk (2, 1, 11, 16), ami arra utal, hogy ezek a kísérletek véletlenszerű válaszokat adnak, ezért alkalmazásuk feltétlenül kerülendő. A legjobban teljesítő két DRIP séma a következő volt: 5 és 13, 79.1-79.5%-os AUC-val, 70-75%-os érzékenységgel, és 72-80% specificitással (**38. ábra**, jobb oldali panel). Fontos megemlíteni, hogy független B-limfoblasztoid sejtvonallal elvégzett kísérletsorozatban a fentiekhez hasonló ROC paramétereket kaptunk 83-91.2%-os AUC-val (**38. ábra**, jobb alsó ROC görbék), ami bizonyítja, hogy a tesztelt DRIP sémák más sejttípusokban is megbízhatóan működnek.



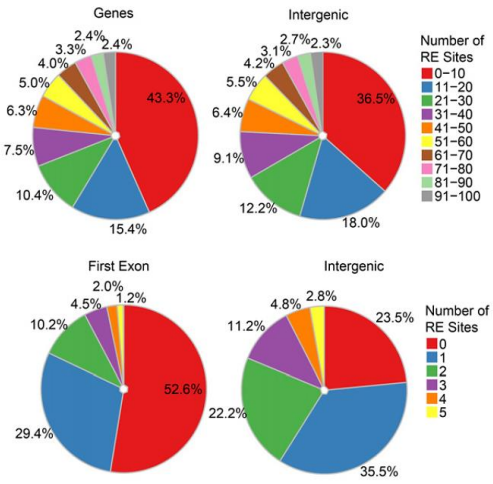
38. ábra A ROC elemzés eredménye. Bal oldali panel: az 1-24 DRIP osztályozó változók AUC eloszlása. Pisorral kiemelve a magas AUC-vel (>70%) rendelkező DRIP sémák. Jobb oldali panel: a két legjobban teljesítő DRIP séma (5, 13) qPCR eredményei (fent) és ROC görbéi (lent). A ROC görbéken összevetettük a T limfocitákon kapott DRIP-qPCR eredményeket (piros) B limfoblasztoid leukémia sejteken kapott független eredményekkel (kék), amelyek nem mutattak statisztikailag kimutatható eltérést.

A kísérleti változók páronkénti összehasonlítása nem mutatott szignifikáns különbséget i) formaldehid-fixált és nem-fixált minták között; ii) fenol-kloroform extrahált és szilika-membrán tisztított nukleinsav minták között, és iii) DNS-fragmentált (1-16) és kromatin-fragmentált (17-24) minták között (**39. ábra**). A sejtlízis hőmérséklete (37°C vs. 65 °C) szintén nem változtatta meg a DRIP specificitását és szenzitivitását. Ellenben statisztikailag szignifikáns különbséget találtunk az RNázA-val kezelt és nem-kezelt minták között ($p = 0,03$), valamint a szonikált és restrikciós enzim koktéllal fragmentált DRIP minták között ($p = 0,0002$). Az RNázA hozzáadása a szabad RNS-ek eltávolítása céljából tehát inkább ront a DRIP módszer hatékonyságán, amelyet az enzim nagyfokú DNS-kötő képességével magyarázunk, a DNS-kötött RNázA ugyanis nagymennyiségű DNS-t képes szelektíven eltávolítani a tisztítási lépések során (Benore-Parsons and Ayoub 1997; Dona and Houseley 2014). Az RNázA erős DNS kötését lineáris és szuperhelikális DNS templáttal végzett gél-shift vizsgálatokkal megerősítettük (Halász et al. 2017). Végül, a szonikált és restrikciós enzimmel fragmentált DRIP minták közötti szignifikáns különbség megmutatta, hogy az ultrahangos kezelés hatékonyabban különbözteti meg a valós pozitív DRIP jelet az álpozitív és álnegatív jelektől, legalábbis a vizsgált kísérleti körülmények között.



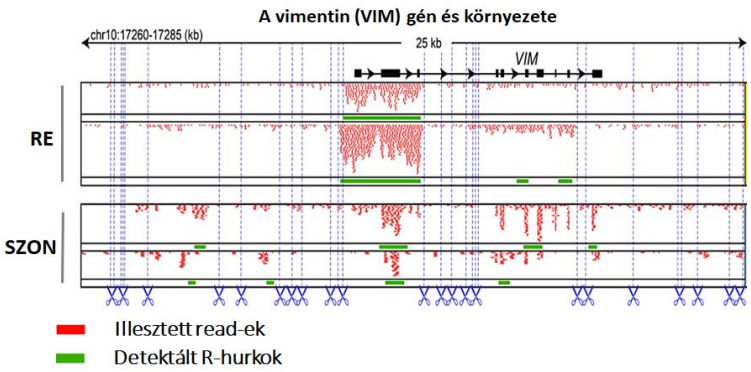
39. ábra DRIP kísérleti változók páronkénti összevetése. A tesztelt kísérleti változók hatását páronként összehasonlítva megállapíthatjuk, hogy az RNázA kezelés elhagyása és szonikálás alkalmazása (restrikciós enzimek helyett) szignifikánsan javítja a DRIP módszer megbízhatóságát. Ellenben a formaldehides keresztkötés, a nukleinsav izolálás módja, a templát típusa (kromatin vagy DNS), illetve a sejtlízis hőmérséklete nem befolyásolja a detektálás specificitását és szenzitivitását.

Fontos megfigyelésünk volt, hogy a DRIP kísérletekben rutinszerűen alkalmazott restrikciós enzim (RE) felismerési szekvenciák eloszlása nem véletlenszerű a genomban, a vágási helyek sűrűsége szignifikánsan magasabb volt az intergénikus régiókban, mint a fehérjekódoló ORF-ekben (**40. ábra**). Különösen az első exonokban alulreprezentáltak az RE felismerő motívumok, amely torzított mintavételt okoz. Az első exonok 82%-ában csak 0-1 megfelelő restrikciós hely található az intergénikus régiókhoz képest (59%) a publikált DRIP kísérletekben alkalmazott enzimkoktélt tesztelve (HindIII, EcoRI, BsrGI, XbaI, SspI).



40. ábra Restrikciós hasítóhelyek eloszlása génikus és intergénikus genomi régiókban. A publikált DRIP kísérletekben a Chédin labor által kifejlesztett HindIII, EcoRI, BsrGI, XbaI, SspI koktélt használják genom-fragmentációra. A kördiagrammok az RE koktél hasítóhelyeinek eloszlását mutatják a gének és intergénikus szakaszok felett. Mindkét genomi területből azonos hosszúságú és esetszámú mintákat hasonlítottunk össze.

A vágási helyek tehát alulreprezentáltak az első exonokon, így a restrikciós enzimekkel történő genom-fragmentálás és DRIP szekvenálás hosszú, restrikciós fragmentum méretű R-hurkokat azonosít a fehérjekódoló gének közelében. A restrikció enzim koktél emésztés hosszabb fragmenteket és szélesebb R-hurok régiókat eredményez, amely a kódoló ORF-ek első exonjánál okozza a legnagyobb torzítást. Szonikált mintáknál ellenben az R-hurkok genomi pozíciója pontosan behatárolható, mivel az ultrahang véletlenszerűen darabolja a DNS-t (**41. ábra**).



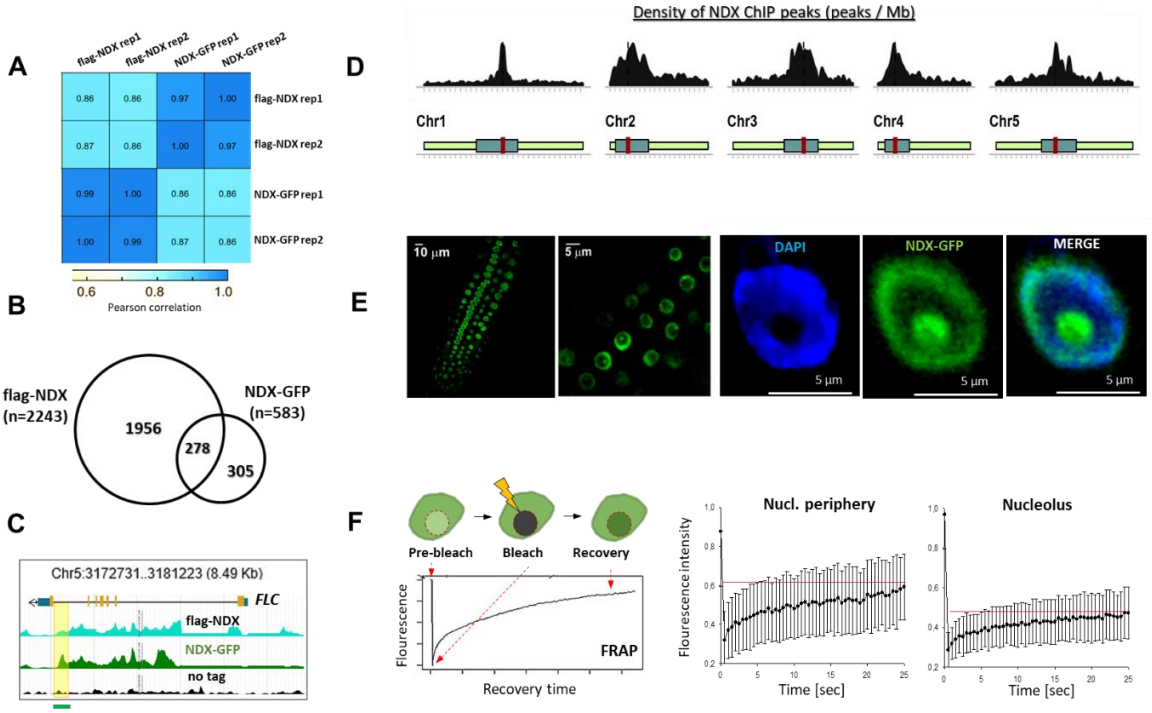
41. ábra DRIP-seq kísérlet restrikciós enzim koktéllal fragmentált genomon és ultrahanggal fragmentált genomon. Az ábra a Vimentin gén (*VIM*) környezetét mutatja be szemléltetésképpen. RE emésztéssel a gén első exonjára eső hosszú fragmentum lett R-hurokként definiálva (zöld vonal). Az R-hurok pontos helye szonikállással feloldható.

Összefoglalva a fentieket, a nukleinsav fragmentáció jelentősen befolyásolja az R-hurkok méreteloszlását és genomi térképezhetőségét, és ezáltal a biológiai funkciójuk pontos megítélését. A fenti torzítás, amelyet "first-exon bias"-ként írtunk le, előzetes *in silico* restrikciós elemzéssel tesztelhető és kiküszöbölhető (Roszik et al. 2017).

A fenti metodikai tapasztalatokra alapozva célul tűztük ki a kromoszómális R-hurok struktúrák nagypontosságú annotációját és új R-hurok regulátor fehérjék azonosítását különféle modell szervezetekben. Első kísérletünket lúdfűben (*Arabidopsis thaliana*) végeztük, mivel ebben a modellorganizmusban elsőként azonosítottak egy R-hurok-asszociált fehérjét, amely a virágzást szabályozó *FLOWERING LOCUS C* (*FLC*) gén területén specifikusan megkötött egy R-hurok struktúrát (Sun et al. 2013). Ez a fehérje a *Nodulin homeobox* (NDX / AT4G03090) transzkripciós faktor volt, az egyetlen ismert makromolekula kísérleteink kezdetéig, amely közvetlen szerepet játszott egy R-hurok felismerésében. A lúdfű kicsiny, teljesen annotált genomja (135 Mbp, 5 kromoszóma, ~25,000 gén) és viszonylag egyszerű genetikai manipulálhatósága lehetőséget kínált az NDX-el kapcsolatos hipotéziseink gyors és hatékony tesztelésére.

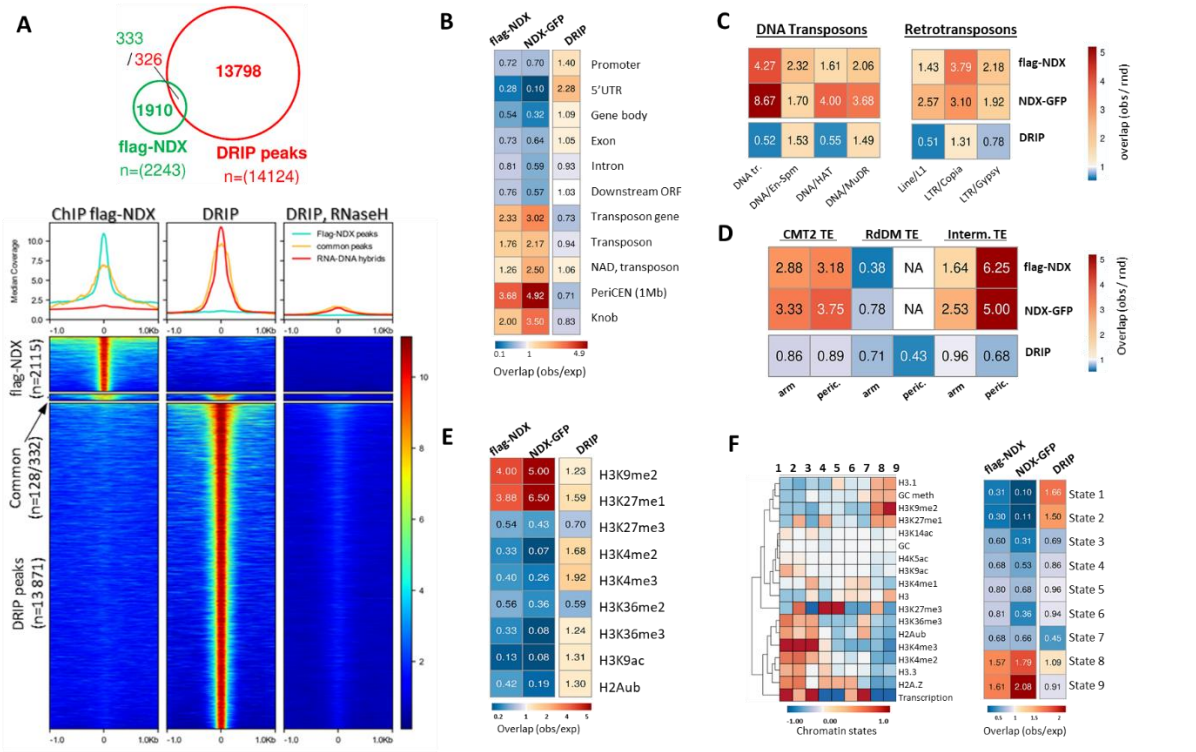
Az NDX-ről kimutatták, hogy az *FLC* gén 3' UTR régiójában antiszensz irányban átíródó nem-kódoló RNS (*COOLAIR*) expresszióját kontrollálja egy R-hurok stabilizálásán keresztül, ezáltal represszálva az *FLC* kifejeződését (Sun et al. 2013). Ennek a mérföldkőnek számító eredménynek a tükrében jogosan vetődött fel a kérdés, hogy vajon az NDX általános R-hurok regulátorként működik-e *Arabidopsis*-ban, vagy szerepe csupán az *FLC* lókuszra korlátozódik? A kérdés megválaszolásához párhuzamosan vizsgáltuk az NDX és R-hurkok kromoszómális eloszlását ChIP és DRIP szekvenálással. Az NDX genomi kötőhelyeit független transzgénikus vonalakban expresszált N-terminálisan jelölt flag-NDX és C-terminálisan jelölt NDX-GFP fúziós fehérjék kromatin immunprecipitációjával és NGS vizsgálattal határoztuk meg. Mindkét fúziós fehérje teljesen komplementálta a vad típusú *NDX* allél delécióját (ld. (Sun et al. 2013), genomi kötődésük pedig erős pozitív korrelációt mutatott (**42/A ábra**). A flag-NDX és NDX-GFP növényekben azonosított peak-ek szignifikáns átfedést mutattak a véletlenhez képest ($p<0.00001$; **42/B ábra**), dúsulásukat

megerősítettük az *FLC* gén területén (**42/C ábra** és (Sun et al. 2013). Az NDX kötőhelyek jellemzően a kromoszómák centromerikus és pericentrikus régióiban halmozódtak fel, ahol sűrűségük (84 peak/Mb) szignifikánsan nagyobb volt, mint a kromoszóma karok mentén (16 peak/Mb; **42/D ábra**). Előbbi régiók funkcionálisan az ún. perifériás heterokromatinnak felelnek meg, amely többnyire erősen kondenzált, génekben szegény és funkcionálisan teljesen vagy részlegesen inaktív, míg az utóbbi régiók eukromatikusak, tehát laza szerkezetűek és elsősorban aktívan átíródó géneket tartalmaznak. A perifériás heterokromatin NDX-asszociációját mikroszkópos mérésekkel is megerősítettünk (**42/E ábra**). FRAP vizsgálatunkban az NDX-GFP rendkívül lassú fluoreszcencia visszatérést és stabil kromatin-asszociációt mutatott (**42/F ábra**), amely mértéke a core hisztonok kinetikájával összevethető (Kimura and Cook 2001; Hetey et al. 2017). A fenti eredmények összességében azt jelzik, hogy az NDX egy heterokromatin-kötő fehérje, amely szorosan kapcsolódik a centromerikus és pericentrikus genomi régiókhoz.



42. ábra Az NDX kromoszómális kötődésének vizsgálata ChIP szekvenálással és FRAP mikroszkópiával. *A-B.* A flag-NDX és NDX-GFP növényekben azonosított kötőhelyek erős pozitív korrelációt és szignifikáns átfedést mutattak a véletlenhez képest ($p<0.00001$, prop.test). *C.* A flag-NDX és NDX-GFP kötődése az *FLC* gén 3' UTR régiójához. A sárgával kiemelt szakasz megegyezik a Sun Q et al. cikkben tesztelt qPCR amplikon pozíciójával. *D.* A flag-NDX peak-ek a pericentromérikus kromoszómális régiókban sűrűsödnek. Piros vonal jelzi a centromerek pozícióját. *E.* CLSM felvétel DAPI-val festett NDX-GFP csíranövényről (gyökércsúcs). A GFP szignál erős perifériás jelölődést és magvacska jelölődést mutat a sejtmagon belül. Sejtmagvacska jelet a vizsgált sejtek ~20%-ban detektáltunk. *F.* Az NDX-GFP mobilitásának vizsgálata FRAP mikroszkópiával. A kísérleti elrendezést a bal oldali panel mutatja. A jobb oldali panel a fluoreszcencia visszatérési görbéket mutatja 25 másodperces időskálán, külön a perifériára és a magvacskára. A vízszintes piros vonal feletti rész jelzi az ún. immobilis hányadot, amely nem mutat fluoreszcencia visszatérést és a kromatinba stabilan beépülő NDX-frakciónak felel meg. Hibasáv: SEM.

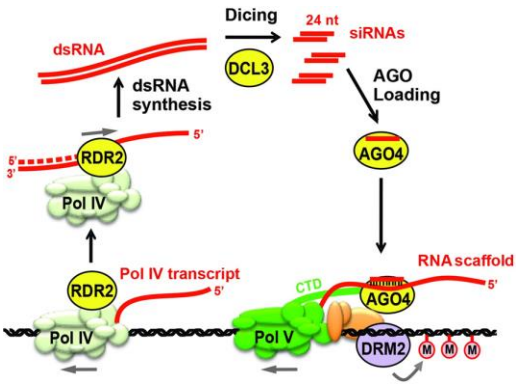
Ezt követően DRIP-seq kísérletekben elemeztük az *Arabidopsis thaliana* kromoszómáin elhelyezkedő R-hurkok számát és elhelyezkedését. Genomi annotációjukat az NDX kötőhelyekkel összevetve megállapítottuk, hogy az R-hurkok teljesen komplementer pozíciót foglalnak el az NDX kötőhelyekhez képest (**43. ábra**), 97%-uk ugyanis semmilyen átfedést nem mutat az NDX peak-ekkel (**43/A ábra**, felső panel), és fordítva, az NDX peak-ek 85%-a független az R-hurkoktól. A DRIP mérést RNázH-kezelt kontroll mellett végeztük, amely teljesen eliminálta az NGS szignált; ez az RNS-DNS hibrid immunprecipitáció specificitását jelzi (**43/A ábra**, alsó panel).



43. ábra Az NDX fehérje és az R-hurkok komplementer genomi eloszlást követnek. *A.* A flag-NDX kötőhelyek és DRIP peak-ek antagonisztikus kötődést mutatnak. Az azonosított kötőhelyek >95%-a nem mutat átfedést (Venn diagram). A hőtérképen (alsó panel) a különbség és metszet halmazok peak-jei felett mutatjuk a flag-NDX jelet (első oszlop), a DRIP jelet (második oszlop), illetve az RNázH-emésztett mintában detektált DRIP jelet (harmadik oszlop). *B.* A flag-NDX, NDX-GFP, és DRIP peak-ek genomi annotációja. Az annotációs kategóriák sorok szerint vannak rendezve. A cellák értékei és színezése arányos az annotációs kategóriával átfedő megfigyelt peak-ek és randomizált peak-ek arányával (obs/rnd). *C.* Az NDX ChIP és DRIP peak-ek annotációja a transzpozon szupercsaládokon (csak a szignifikáns találatokat mutatjuk). D. Ugyanaz, mint a 'C' panel, csak a TE családokat a transzpozon-csendesítő DNS metiláz rendszerek szerint csoportosítottuk. CMT2 TE: CMT2 kromometiláz által targettált TE. RdDM TE: az RNS-emdiált DNS metilá rendszer által targettált TE. Intermediate TE: CMT2/RdDM közös útvonalon metilált TE-k. *E.* Az NDX ChIP és DRIP peak-ek dúsulása a különféle aktiváló- és gátló hiszton modifikácók felett. *F.* A panel bal oldali része mutatja az *Arabidopsis* kromatinállapotokat (states 1-9), amelyeket Sequeira-Mendes munkája nyomán rekonstruáltunk klaszterelemzéssel. A DRIP peak-ek a state 1-2-vel mutatnak jelentős átfedést. A state 1-et aktív mRNS transzkripció, H3K4me3, H3K4me2, H3K36me3,H2Bub, és H3K9ac jellemzi. A state 2-ben szintén az aktív jelek dominálnak, azonban itt már megjelenik a represszív H3K27me3 is. A flag-NDX és NDX-GFP peak-ek a state 8-9-el mutatnak szignifikáns asszociációt. A state 8-9 főepigeneikai markere a H3K9me2, H3K27me3, és a DNS metiláció.

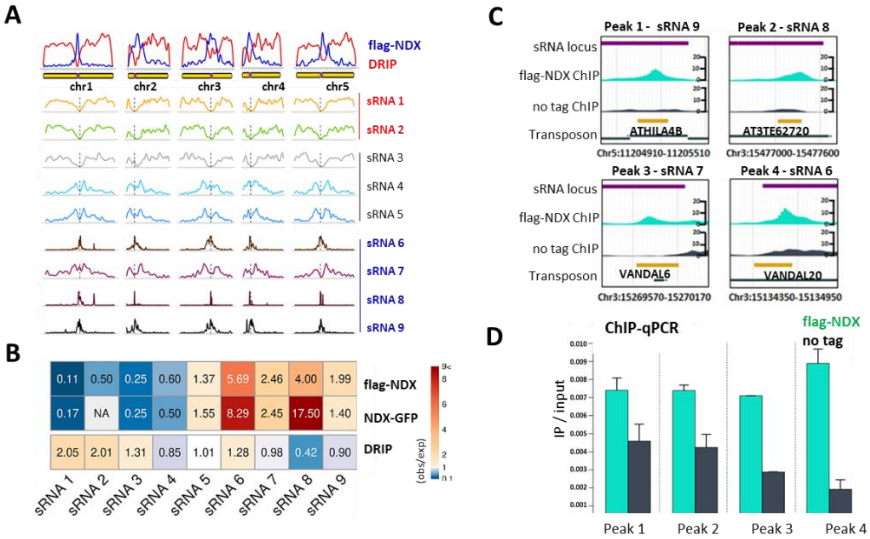
Az R-hurkok jellemzően a gének 5’ UTR régiójában dúsultak és mérsékelt együttálást mutattak bizonyos aktiváló hiszton modifikációkkal (H3K4me2, H3K4me3) (**43/B, E ábrák**). Az NDX kötőhelyek ezzel szemben a pericentromerikus régiókban dúsultak (és az ezzel analóg ún. heterokromatin „szigeten” (Knob), amely a kromoszóma karon található (Grob et al. 2013), valamint a transzpozonokonokon/transzpozon géneken és a represszív hiszton módosítások területén (H3K9me2, H3K27me1) (**43/B, E ábrák**). Az annotált transzpozábilis elemek (TE) közül néhány TE család szignifikáns NDX-asszociációt mutatott (**43/C ábra**), amelyek heterogén szerkezete, eloszlása és transzpozíciós mechanizmusa nem tette lehetővé a funkcionális kapcsolatok igazolását. Viszont a TE-ket a csendesítésükért felelős DNS metilációs útvonalak szerint csoportosítva (CMT2 TE, RdDM TE, intermediate TE) világosan látszik, hogy az NDX elsősorban a CMT2 útvonalon és a közös (CMT2/RdDM; „intermediate”) útvonalon szabályozott TE-khez kötődik, függetlenül azok pericentromerikus vagy kromoszóma kari elhelyezkedésétől (**43/D ábra**).

Ennek alapján az NDX funkciója a transzpozon metilációs útvonalakhoz kapcsolódhat, amely megértéséhez ki kell emelnünk a növényi DNS metilációs rendszer sajátosságait. Először is, a növényi DNS-metiláció három szekvencia-kontextusban történik (CG, CHG, CHH, ahol H: A, C vagy T) és azt különböző genetikai útvonalak vezérlik (Zhang and Zhu 2012). A CG metilációt a MET1, a CHG metilációt a CMT2/3 kromometilázok, míg a CHH metilációt a DRM2 metiltranszferázok katalizálják (Lindroth et al. 2001; Stroud et al. 2014). Kiemelendő, hogy a CHH/CHG metiláció elsősorban a transzpozonok aktivitását szabályozza a pericentromérikus régiókban (ahol az NDX jellemzően kötődik). A CHH/CHG metilációt az RdDM útvonal hozza létre *de novo*, amely 24nt het-siRNS-ek szintézisével indul specifikus sRNA lókuszokról (Matzke and Mosher 2014; Gao et al. 2010; Zheng et al. 2010; Zhong et al. 2014). Az RdDM iniciációja során a növényspecifikus RNS-polimeráz IV (Pol IV) prekurzor sRNS-transzkriptumokat generál, melyeket az RNS-függő RNS-polimeráz 2 (RDR2) és a DICER-Like 3 (DCL3) endoribonukleáz funkcionális 21nt het-siRNS-ekké processzál (Zhong et al. 2014). A het-siRNS-ek egyik szála ezután beépül az RNS-indukált transzkripciós csendesítő komplex (RITSC) központi egységébe (AGO4), amelyet egy másik növényspecifikus RNS-polimeráz, a Pol V által generált scaffold-RNS-ek rögzítenek a cél-lókuszokhoz. A target szekvenciákat végül a DRM2 citozin-metiltranszferáz metilálja *de novo* (**44. ábra**). A CHH/CHG metiláció fenntartásáért elsősorban a CMT2/3 kromometilázok felelősek, amelyek H3K9me2 represszív hiszton módosításokat felismerve és azokhoz kötődve végzik a transzpozonok metilálását.



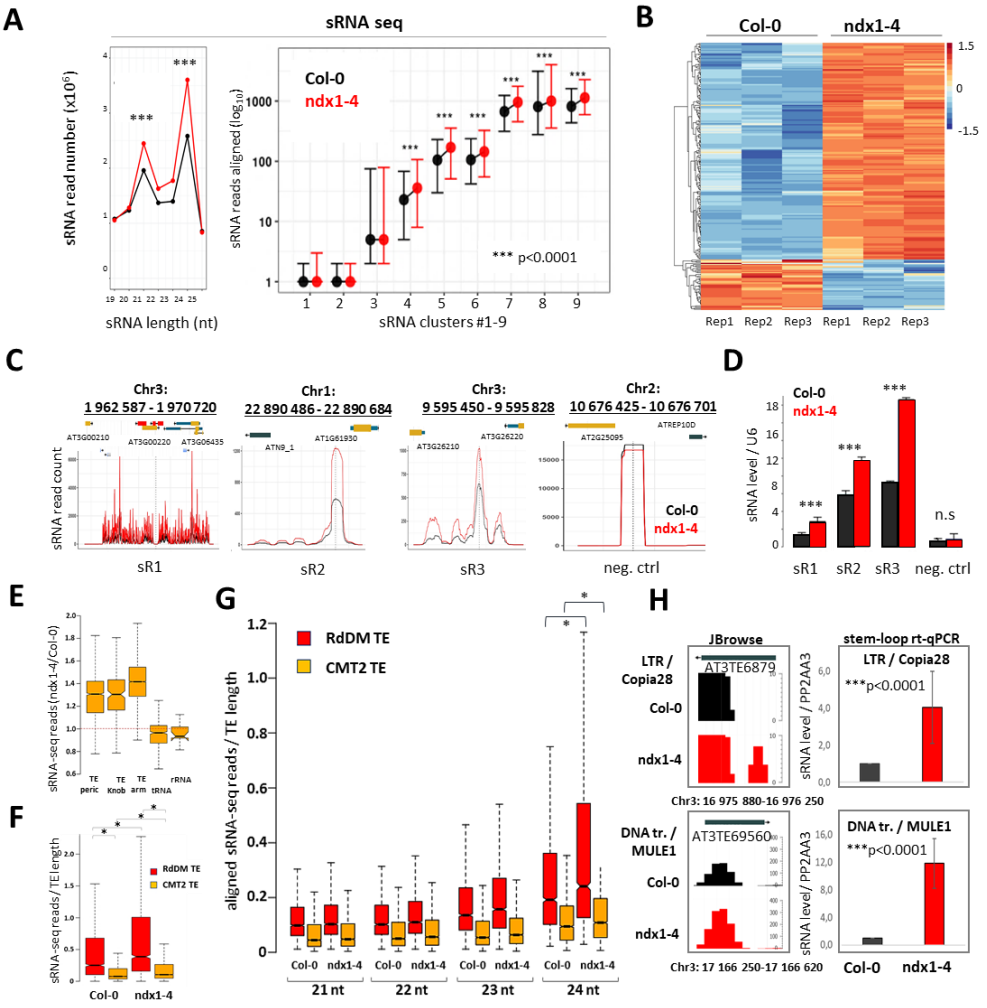
44. ábra Az RNS-mediált DNS metiláció (RdDM) útvonal egyszerűsített sémája lúdfűben. Az sRNS lókuszok egy csoportjához kötődve a növény-specifikus RNS polimeráz (Pol) IV és az RNS-dependens RNS polimeráz 2 (RDR2) együttműködése kettősszálú RNS-ek (dsRNS) szintéziséhez vezet, amelyeket a DICER-like 3 (DCL3) 24nt prekurzor-siRNS duplexekké processzál. A pre-siRNS-ek egyik szála az Argonaute 4-re (AGO4) kerül, amely a Pol V-által átírt speciális scaffold RNS-ekhez hibridizálódik a target transzpozonok felett és ezáltal a célszekvenciákhoz kötődik. Az ide rekrutált DRM2 citozin metiltranszferáz végül a Pol V által átírt DNS szakaszokon *de novo* citozin-metilációt eredményez. Forrás: (Blevins et al. 2015).

Az NDX fehérje és a CMT2/3 metiláz rendszer kapcsolatát (**43/D ábra**) tovább erősíti az a megfigyelés, hogy az *Arabidopsis*-ban azonosított kilenc funkcionális kromatinállapot közül (Sequeira-Mendes et al. 2014) az NDX a heterokromatikus *state 8-9*-el mutat szignifikáns átfedést, míg az R-hurkok az eukromatikus *state 1-2*-re voltak jellemzőek (**43/F ábra**). A legmarkánsabb együttállást azonban a közelmúltban azonosított nem-kódoló kisRNS lókuszok (Hardcastle et al. 2018) egy csoportja, a 'klaszter 6-9' sRNS-ek, és az NDX kötőhelyek között figyeltük meg, melyek a pericentrikus heterokromatin-asszociált siRNS-eket (het-siRNS) kódolják (**45/A-B ábra**). Az R-hurkok ebben az esetben is antagonisztikus eloszlást követtek az NDX kötőhelyekhez képest, mivel a 'klaszter 1-2' sRNS lókuszokon mutattak szignifikáns kötődést (**45/A-B ábra**). Néhány reprezentatív 'klaszter 6-9' sRNS lókusz NDX-kötődését ChIP-qPCR-ral validáltuk, amely a **45/C-D ábrán** látható.



45. ábra Az NDX és a het-siRNS lókuszok asszociációja. *A.* Felső panel: A DRIP (piros) és flag-NDX (kék) szignál ellentétes eloszlása az 5 kromoszóma mentén. Alsó panel: Hardcastle és munkatársai által azonosított kisRNS klaszterek (sRNS 1-9) sűrűsége az 5 kromoszóma mentén. A flag-NDX ChIP peak-ek eloszlása az sRNA 6-9 lókuszok (het-siRNS-ek) eloszlásával mutat jelentős hasonlóságot. *B.* Hőtérkép, amely az sRNS lókuszok átfedését mutatja a flag-NDX, NDX-GFP, és DRIP peak-ekkel (obs / expected). *C-D.* Reprezentatív flag-NDX kromatin kötőhelyek ChIP-qPCR validációja, amelyek az sRNA 6-9 heti-siRNS lókuszokhoz kötődnek.

A továbbiakban egy T-DNS inzerciós mutánsban (*ndx1-4*) teszteltük az NDX fehérje het-siRNS asszociációjának funkcionális következményeit. Kontroll (Col-0) és *ndx1-4* mutáns növényekben kisRNS szekvenálást (sRNS-seq) végeztünk, és megállapítottuk, hogy az NDX hiánymutánsban szignifikánsan több funkcionális kisRNS (21nt/24nt) szintetizálódik, mint NDX jelenlétében (**45/A ábra**, bal panel). A kilenc sRNS osztály közül a 4-9. csoport mutatott megnövekedett expressziós szintet az *ndx1-4* mutánsban, míg az 1-3. csoport nem különbözött a vad típustól (**46/A ábra**, jobb panel).



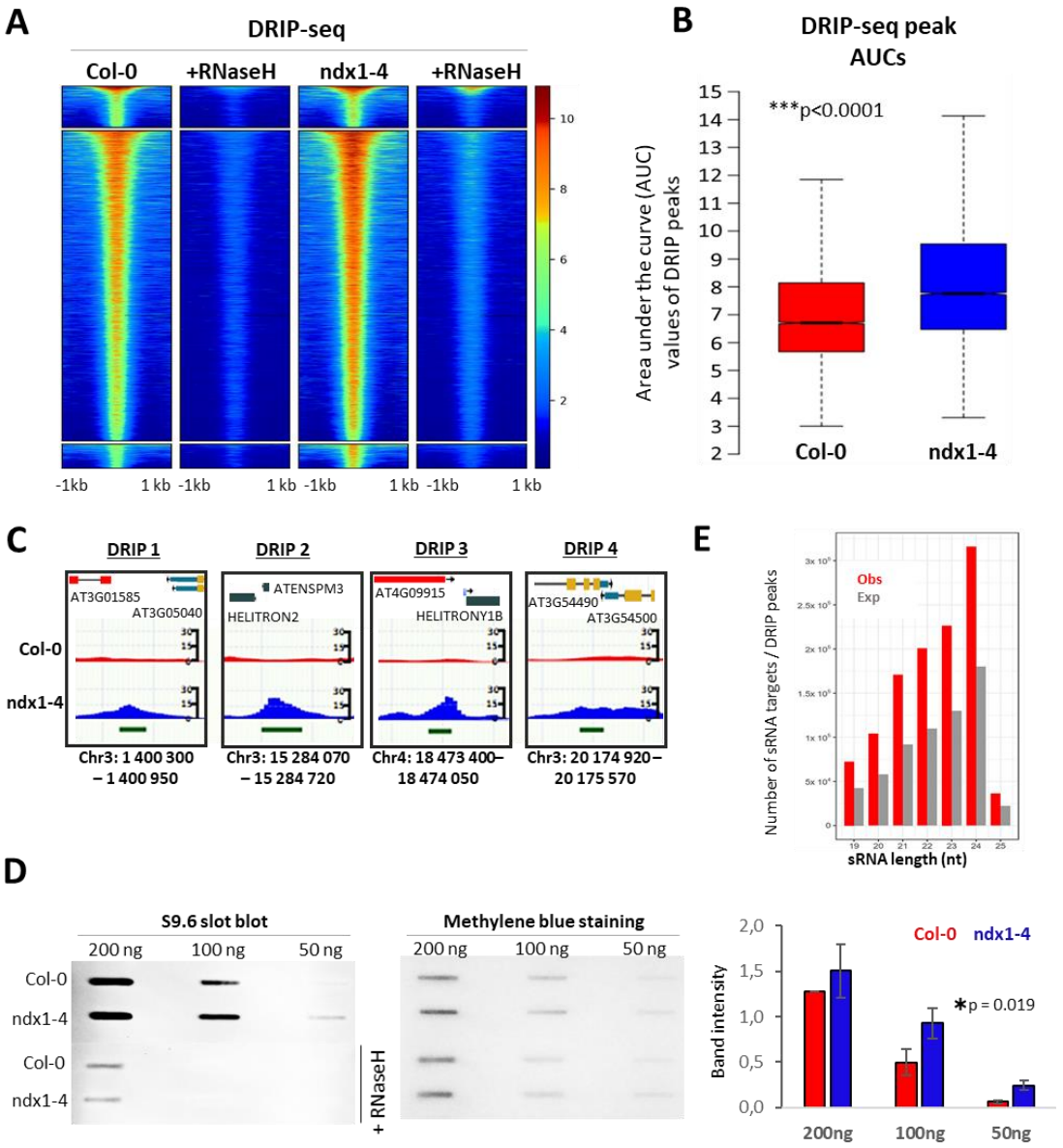
46. ábra A het-siRNS-ek indukciója NDX hiányában. *A.* Bal panel: a kisRNS szekvenálással (sRNS-seq) azonosított sRNS-ek számossága a 19-25nt méret-kategóriában. A két domináns csúcs a 24nt (prekurzor) és 21nt (érett) sRNS-nél látható, amely szignifikánan emelkedett mennyiséget mutat az *ndx1-4* mutánsban (piros vonal). Jobb panel: ugyanez, csak a kilenc funkcionális sRNS csoportra lebontva. Szignifikáns növekedés a 4-9. csoportban látható az *ndx1-4* mutánsban (piros vonal). *B.* Hőtérkép, amely az indukálódó (narancssárga) és represszált (kék) sRNS-lókuszokat mutatja az *ndx1-4* mutánsban (3-3 biológiai replikátum). *C.* Reprezentatív példák kisRNS indukcióra. Az sR1, sR1, sR3 indukálódó kisRNS-eket mutat, amely transzkripciója szignifikánsan magasabb az *ndx1-4* mutánsban (piros vonal). A 'no change' pedig negatív kontroll sRNS lókusz, amely nem mutat differenciális expressziót. *D.* Az sR1-sR3 és 'no change' sRNS lókuszok transzkripciós profiljának rt-qPCR validációja. *E.* A transzpozon sRNS expresszió növekedése NDX hiányában. Az illesztet sRNS read-ek aránya (*ndx1-4* / Col-0) szignifikánsan megnőtt a pericentromerikus, Knob, és kromoszóma karokon található TE-en a „negatív kontroll"

régiókhoz (tRNS és rRNS gének) és a genomi átlagértékhez képest (horizontális piros vonal). Statisztikai szignifikancia: ***p<0,0001 (Wilcoxon teszt, kétoldalas). *F.* A illesztett sRNS read-ek száma szignifikáns növekedést mutat az *ndx1-4* mutánsban mind az RdDM TE-ken mind a CTM2 TE-ken. Az sRNS read-ek számát a TE hosszára normáltuk. (Az RdDM TE-k siRNS szintjei szignifikánsan magasabbak, mint a CMT2 TE-ké (mind a Col-0-ban, mind az *ndx1-4*-ben)). Statisztikai szignifikancia: *p<0,05 (Wilcoxon teszt, kétoldalas). *G.* Ugyanaz, mint az 'F', de az sRNS-eket felbontottuk méretük szerint 21, 22, 23, 24 nt siRNS csoportokra. A 24 nt osztályban statisztikailag szignifikáns különbség van az *ndx1-4* és a Col-0 minták expressziója között (RdDM TE-k: p<2x10-16; CMT2 TE-k: p=0,014). *H.* A Copia28 és MULE1 transzpozonok megnövekedett sRNS expressziójának validálása stem-loop rt-qPCR-ral (PP2AA3 expresszióra normálva). 3 független kísérlet átlaga. Hibasáv: SEM.

Ez a transzkripciós heterogenitás összhangban van az sRNS osztályok elsődleges funkcionális felosztásával, amely pontosan az 1-3. és 4-9. csoportok között történik (előbbi a fehérjekódoló génekkel és a promóterekkel mutat szoros asszociációt, míg utóbbi az epigenetikailag aktivált siRNS-ekkel (easiRNS) és az RNS-függő DNS metilációs (RdDM) útvonallal (Matzke and Mosher 2014). Statisztikai elemzésünk szerint 451 sRNS lókusz mutatott szignifikánsan magasabb (n=286) vagy alacsonyabb (n=165) expressziót a kontroll növényhez képest (**46/B ábra**), melyek többsége a 'klaszter 6-9' sRNS csoportba esett. Néhány indukálódó sRNS lókuszt (**46/C ábra**) rt-qPCR-ral validáltunk, amely a **46/D ábrán** látható. A transzpozonokon méhető sRNS expressziót vizsgálva szignifikáns TE indukciót figyeltünk meg az *ndx1-4* mutánsban, mind a kromoszóma karokon és a pericentromerikus és 'Knob' régiókban (>1.5-fold change a genomi átlaghoz képest (vízszintes piros vonal) és negatív kontrol régiókhoz képest (tRNA. rRNA gének), **46/E ábra**). A transzpozonok fő funkcionális csoportjait összehasonlítva megállapítottuk, hogy mind az RdDM TE-k mind a CMT2 TE jelentősen magasabb sRNS expressziót mutattak az *ndx1-4* mutánsban (**46/F ábra**), azonban az sRNS-eket hosszuk szerint kategórizálva (21 nt, 22 nt, 23 nt, 24 nt) kiderült, hogy kizárólag 24 nt het-siRNS-ek indukálódtak (**46/G ábra**). (A 24 nt sRNS-eket a Pol IV szintetiázálja a pericentromerikus régiókban.) Eredményeinket stem-loop rt-qPCR kísérletekkel támasztottuk alá két jól karakterizált pericentromerikus transzpozonon (LTR/Copia28, DNA/MULE1; **46/H ábra**).

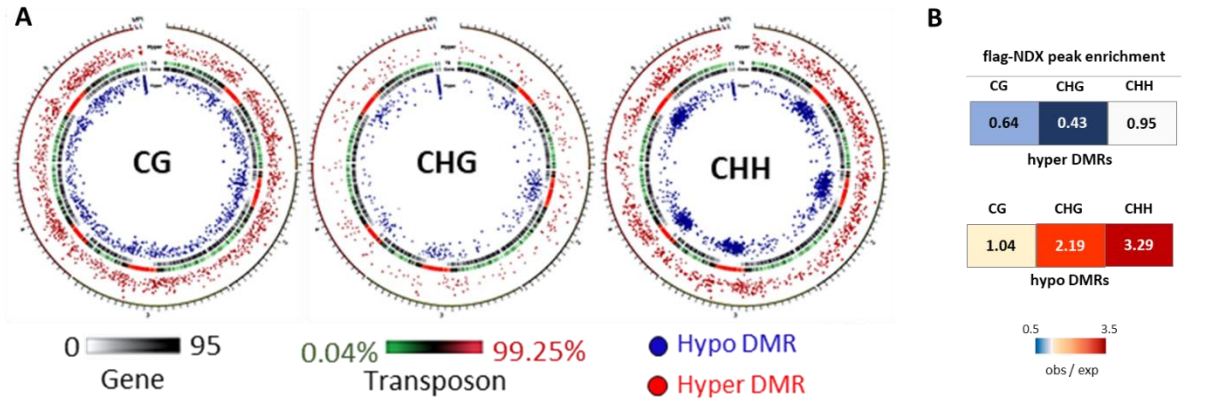
A továbbiakban az *ndx1-4* mutánsra jellemző R-hurok profilt DRIP szekvenálással térképeztük (**47/A ábra**), amely mérsékelt, de statisztikailag szignifikáns növekedést mutatott a kontroll növényhez képest (**47/B-C ábra**). Az *ndx1-4* hiánymutánsban tapasztalt emelkedett R-hurok szintet független Slot blot hibridizációs módszerrel validáltuk (**47/D ábra**), amely a DRIP-seq méréshez hasonló eredményt adott. Mivel a DRIP csúcsok és az NDX kötőhelyek reciprok genomi eloszlást követnek, az R-hurok szint mérsékelt növekedését valószínűleg az *NDX* mutáció indirekt hatása okozza. Feltételezésünk szerint az *ndx1-4* mutánsban felszaporodó het-siRNS-ek homológ RNS-DNS hibridként beépülhetnek a kromatinba, amely az R-hurkok mennyiségi növekedését eredményezi. Statisztikai elemzésünkben szignifikáns szekvencia-homológiát találtunk az indukálódó siRNS-ek és az R-hurkok között (**47/E ábra**), amely arra utal, hogy az *ndx1-4* mutánsban felszabályozott siRNS-

ek homológ DNS szekvenciákkal hibridizálva növelik a globális R-hurok szintet. Ez a modell azonban további kísérletes validációra vár.



47. ábra R-hurkok indukciója NDX hiányában. *A.* Hőtérkép, amely a Col-0 kontroll és *ndx1-4* mutáns növényben azonosított DRIP peak-ek feletti szignált mutatja. A mutánsban mérsékelt növekedés látszik. Mindkét mintát RNázH-kontroll mellett szekvenáltuk, amely teljesen eliminálta a DRIP jelet. *B.* Az R-hurkok indukcióját az *ndx1-4* mutánsban detektált peak-ek nagyobb intenzitása bizonyítja, amelyet AUC (görbe alatti terület) analízissel számoltunk ki (FC>1,5; p<0,0001). *C.* Reprezentatív példák *ndx1-4* mutánsban indukálódó DRIP csúcsokra. *D.* Slot blot kísérlet Col-0 kontroll és *ndx1-4* mutáns növényben. A bal oldali membrán mutatja az RNS-DNS hibrid specifikus S9.6 antitest jelet különböző DNS koncentrációknál és RNázH-kezelt kontroll mellett, míg a jobb oldali membrán a felcseppentett nukleinsav mennyiségét jelzi (metilénkék festés). Az oszlopdiagram a metilénkék szignálra normált S9.6 jel intenzitását mutatja. *E.* Az R-hurkok szekvenciákkal homológiát mutató, *ndx1-4* mutánsban indukálódó sRNS-ek eloszlása a 19-25 nt méret-kategóriában (piros). Szürke oszlopok jelzik ugyanezt random szakaszok felett. A véletlenhez képest szignifikánsan magasabb szekvencia homológia jellemző az R-hurkok ész *ndx1-4* mutánsban indukálódó sRNS-ek között. (3 független kísérlet átlaga). Hibasáv: SEM.

Mivel a növényekben het-siRNS-ek szabályozzák a *de novo* DNS metilációt és géncsendesítést (Matzke and Mosher 2014; Hardcastle et al. 2018), DNS-metilom vizsgálatokat végeztünk Col-0 és *ndx1-4* növényekben biszulfit-szekvenálással (BS-seq), amely a nem-metilált citozin bázisok timinné konvertálásán alapul. BS-seq kísérletünkben mindhárom szekvencia kontextusban (CG/CHG/CHH) meghatároztuk a metilált citozinok arányát Col-0 és *ndx1-4* növényekben, majd differenciálisan metilált régiókat (DMR) azonosítottunk, amelyek szignifikánsan alacsonyabb vagy magasabb metilációs szintet mutattak a mutánsban (előbbieket hipo-DMR-eknek, utóbbiakat hiper-DMR-eknek hívjuk). Elemzésünk 2449 hipometilált és 1597 hipermetilált régiót mutatott ki a CHH és CHG szekvenciáknál, amelyek jellemzően a pericentromérius régiókban (hipo-DMR-ek) és a kromoszóma karokon (hiper-DMR-ek) dúsultak (**48/A ábra**).

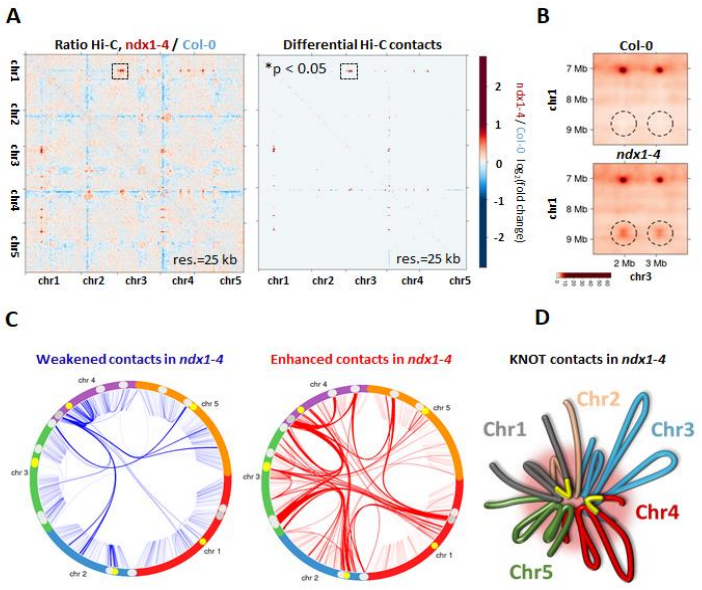


48. ábra DNS metiláció globális vizsgálata biszulfit szekvenálással (BS-seq). *A.* Circos ábra, amely koncentrikus gyűrűi különböző genomi kategóriákat és DNS metilációs változásokat mutatnak. A genomi kategóriák kívülről befelé haladva: i. *Arabidopsis* kromoszómák (1-5), ii. Transzpozonok, amelyek kromoszómális sűrűségét zöld-piros hőtérkép jelzi, iii. Fehérjekódoló gének, amelyek sűrűségét fekete-fehér hőtérkép jelzi. A pericentromerikus régiók könnyen beazonosíthatók, mivel transzpozon gazdagok és génekben szegények. Differenciális DNS metiláció: a kék és piros pontok a hipometilált (hipo-DMR) és hipermetilált (hipr-DMR) DNS szakaszokat mutatják, amelyeket bioinformatikai elemzéssel azonosítottunk. A hipo- és hiper-DMR-ek szignifikánsan alacsonyabb vagy magasabb metilációs szintet mutatnak az *ndx1-4* mutánsban a Col-0-hoz képest. A CG DMR-eket a bal oldali circos plot mutatja, a középső ábra a CHG DMR-eket mutatja, míg a jobboldali ábra a CHH DMR-eket mutatja. *B.* Hőtérkép, amely az NDX ChIP kötőhelyek dúsulását/deplécióját mutatja az azonosított hipo- és hiper-DMR-eken. A cellák értéke a megfigyelt peak-ek arányát mutatja a véletlenszerűen generált peak-ekhez képest (obs/exp). A melegebb színek nagyobb dúsulást jeleznek.

Az DMR-ek közül a hipo-CHH/CHG DMR-ek szignifikáns kolokalizálót mutattak az NDX kötőhelyekkel (amelyeket a vad típusú növényben azonosítottunk) (**48/B ábra**), amely az NDX fehérje direkt hatására utal. A CG dinukleotidok kapcsán 1353 hipo- és 1624 hiper-CG DMR-t azonosítottunk, amelyek egyenletes eloszlást követtek a kromoszómák mentén (**48/A ábra**) és nem mutattak statisztikai asszociációt az NDX peak-ekkel (**48/B ábra**). A fenti összefüggések az NDX kettős szerepére utalnak a DNS metiláció és a heterokromatin homeosztázis szabályozásában, mivel az eukromatikus kromoszóma karokon az NDX inkább gátolja a CHH/CHG metilációt (NDX hiányában

hipermetilálódnak), viszont a pericentrikus régiókban serkenti azt (*ndx1-4* hiánymutánsban ezek a szakaszok hipometilálódnak). A kromoszóma karokon az NDX valószínűleg közvetett szerepet játszik a DNS-metiláció gátlásában, mivel genomi kötőhelyei nem itt találhatók. A pericentrikus heterokromatinban viszont közvetlenül serkentheti a DNS-metilációt, mivel kötőhelyei jellemzően ebben a régióban találhatók. Mivel a pericentrikus heterokromatin rendkívül gazdag transzpozonokban, az NDX-nek szerepe lehet azok metilációjában és transzkripciós gátlásában.

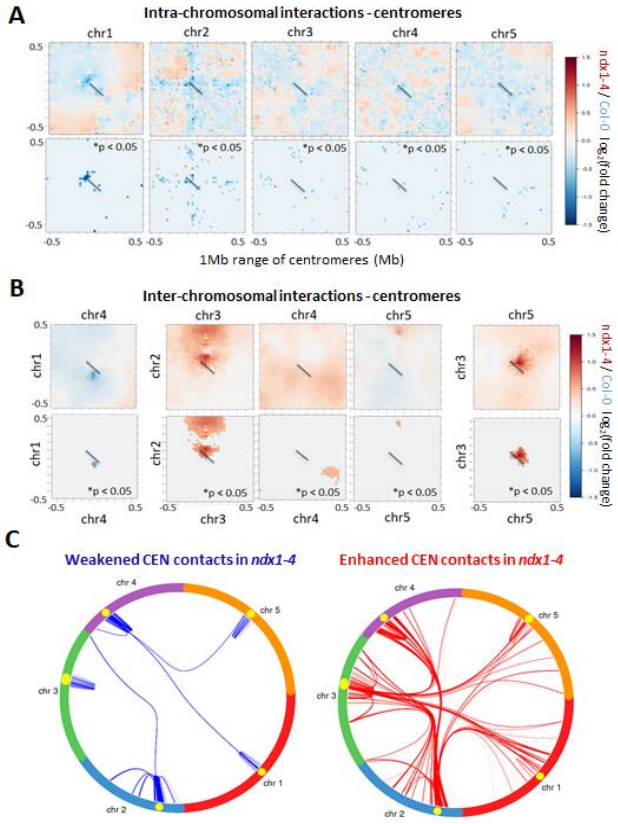
A fenti DNS metilációs, siRNS, és R-hurok változások kromatinszerkezeti hátterének vizsgálatához Hi-C kísérleteket végeztünk Col-0 és *ndx1-4* mutáns növényekben, hogy kiderítsük, a megfigyelt különbségek összefüggésbe hozhatók-e a genomszerkezet változásaival. Mindkét mintából ~200 millió paired-end read-et szekvenáltunk, amelyekből nagyfelbontású Hi-C interakciós mátrixokat generáltunk. A mátrixok kvantitatív elemzésével szignifikáns különbségeket azonosítottunk a Col-0 és *ndx1-4* minták között, amely számos genomi régióban megnövekedett (piros) vagy csökkent (kék) Hi-C értéket tárt fel (**49/A-B ábra**).



49. ábra 3D genomszerkezet vizsgálata Hi-C módszerrel. *A.* Differenciális Hi-C mátrix, amelyben az *ndx1-4* interakciós mintázatát viszonyítjuk a Col-0 mátrixhoz. Piros pixelek jelzik a mutánsban felerősödő Hi-C kontaktokat. A jobb oldali panel mutatja a statisztikailag szignifikáns differenciális kapcsolatokat. *B.* Reprezentatív példa *ndx1-4-ben* kierősödő Hi-C kapcsolatra. *C.* Circos plot, amely az *ndx1-4*-ben gyengülő (kék) és felerősödő (piros) Hi-C kapcsolatok hálózati topográfiáját mutatja. A színek intenzitása arányos a különbség mértékével (de mind statisztikailag szignifikáns, p<0.05, FC>1.5). *D.* Módosult KNOT struktúra az *ndx1-4* mutánsban. A sárgával jelzett régiók interkromoszómális kapcsolatokkal hozzáadódnak a vad típusú növényekre jellemző 10 alaprégióhoz.

E különbségek a kromatinszerkezet globális megváltozását jelzik az NDX fehérje hiányában. Az egyik legmarkánsabb változást az ún. KNOT (csomó) struktúrában figyeltük meg, amely a vad típusú növényekre jellemző tíz interkromoszómális régió helyett (Grob et al. 2014) új *de novo* elemekkel egészült ki az *ndx1-4* mutánsban (**49/D ábra**). A másik markáns Hi-C változás a centromerekre és pericentromerikus régiókra jellemző, amely szignifikánsan csökkent intrakromoszómális kapcsolatokat és felerősödő interkromoszómális interakciókat mutat az *ndx1-4* mutánsban (**50. ábra**). Az eltűnő intracentromerikus kölcsönhatások a kromatin fellazulására utalnak NDX hiányában (dekompakció), amely egybeesik az NDX jellemző genomi kötőhelyeivel és

erősen korrelál az *ndx1-4* mutánsban megfigyelhető sRNS-expressziós és CHH/CHG-metilációs változásokkal.



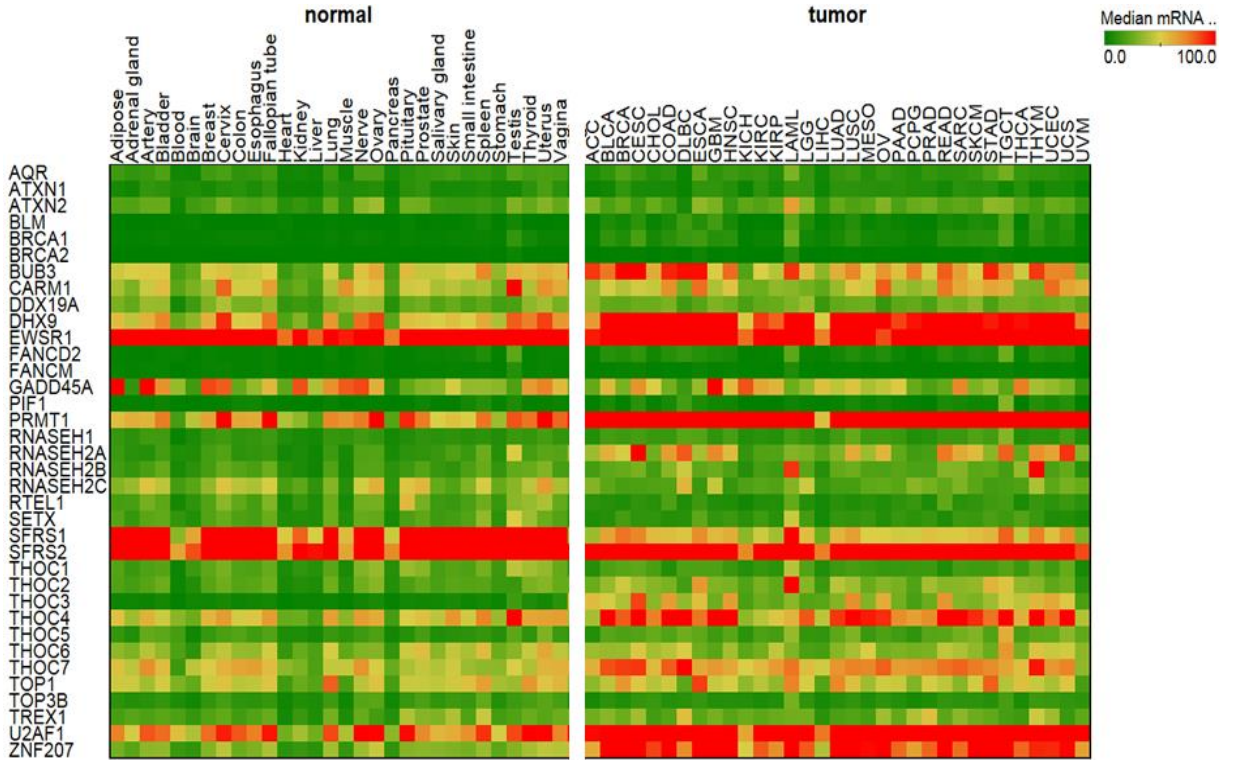
50. ábra A pericentromerikus régiók intra- és interkromoszómális Hi-C interakciói *A.* Differenciális Hi-C pile up plot, amely az összes centromer összes intrakromoszómális interakcióját mutatja. Piros pixelek jelzik a mutánsban felerősödö Hi-C kontaktokat, míg kék pixelek mutatják az eltűnő kapcsolatokat. Az átlón futú fekete vonal jelzi a centromerek pontos pozícióját. Alsó panel: szignifikáns pixelek. *B.* Ugyanaz, mint az 'A' panel, csak az interkromoszómális kapcsolatokra. *C.* Circos plot, amely az *ndx1-4* mutánsban gyengülő (kék) és felerősödő (piros) centromer-specifikus kapcsolatok topográfiáját mutatja. Az intrakromoszómális kapcsolatok gyengülnek, míg az interkromszomszómális kontaktok inkább erősödnek. A színek intenzitása arányos a különbség mértékével (a Col-0-hoz képest). Csak a statisztikailag szignifikáns változásokat mutatjuk (p<0.05, FC>1.5).

Adataink alapján azonban ok-okozati kapcsolatokat nem lehet felállítani, vagyis azt nem tudjuk meghatározni, hogy az NDX inaktiválása okozza-e a kromatin dekompakcióját a pericentromerikus genomi régiókban, amely ezután fokozott het-siRNS expresszióhoz és DNS metilációs változásokhoz vezet, vagy a felerősödött siRNS transzkripciós és DNS metilációs változások vezetnek a kromatinszerkezet fellazulásához. A megnövekedett globális R-hurok képződés viszont valószínűleg az aberráns siRNS-termelés és a kromatin-változások következményeként, azok mellékhatásként történik. A fentieket összefoglalva, különféle genomikai és sejtbiológiai módszerekkel az NDX fehérjéről bebizonyítottuk, hogy egy alapvető heterokromatin regulátor, amely a korábbi elképzelésekkel szemben nem az R-hurok-gazdag eukromatikus kromoszóma karokon, hanem a heterokromatikus centromer és pericentromerikus régiókban működik, ahol a het-siRNS-ek transzkripcióját és a represszív CHH/CHG metilációt kontrollálja. A fenti molekuláris útvonalak részletes genetikai analízise a siRNS-biogenézis mélyebb megértéséhez és a nem-kódoló kisRNS-ek pontosabb manipulálásához vezethet növényekben és más fajokban egyaránt.

További vizsgálatainkban figyelmünket a humánspecifikus R-hurok regulátor fehérjék felé fordítottuk, amelyeket egyáltalán nem ismertünk a 2018-as évig. Ekkor jelent meg egy tanulmány (Wang et al. 2018), amelyben tömegspektrometriás vizsgálattal és immunprecitációval több száz

potenciális R-hurok-kötő fehérjét azonosítottak *in vitro*, amely megnyitotta az utat az R-hurok regulátor gének vizsgálatához különféle fiziológiás és patológiás folyamatokban. A rákos megbetegedések kapcsán például az elmúlt 10-12 év kutatásaban világossá vált, hogy az R-hurkok kontrollálatlan keletkezése és a tumorokra jellemző genom instabilitás között van valamilyen mechanikai kapcsolat, azonban a résztvevő faktorokat és funkcionális kölcsönhatásokat nem tudták azonosítani. A legtöbb onkogén fehérje a sejtciklus beindítása mellett a transzkripciós gépezetet is erősen aktiválja, beleértve pédául mindhárom RNS polimerázt és az általános transzkripciós faktorokat (Kotsantis et al. 2016). Ez fokozott R-hurok képződéshez és replikációs stresszhez vezet (a repliszóma és az R-hurkok összeütközése révén), amely a replikációs apparátus szétesése miatt nagymennyiségű DSB-t generál (Gan et al. 2011). A DSB-k rekombinációs folyamatokat indítanak el, amelyek kromoszóma aberrációt és hibás testvérkromatid szegregációt okoznak. E folyamatokat a tumorsejtek meglehetősen jól tolerálják (pl. az apoptózis gátlása révén), ami összességében elősegíti a kontrollálatlan sejtburjánzást és a metasztázisok keletkezését.

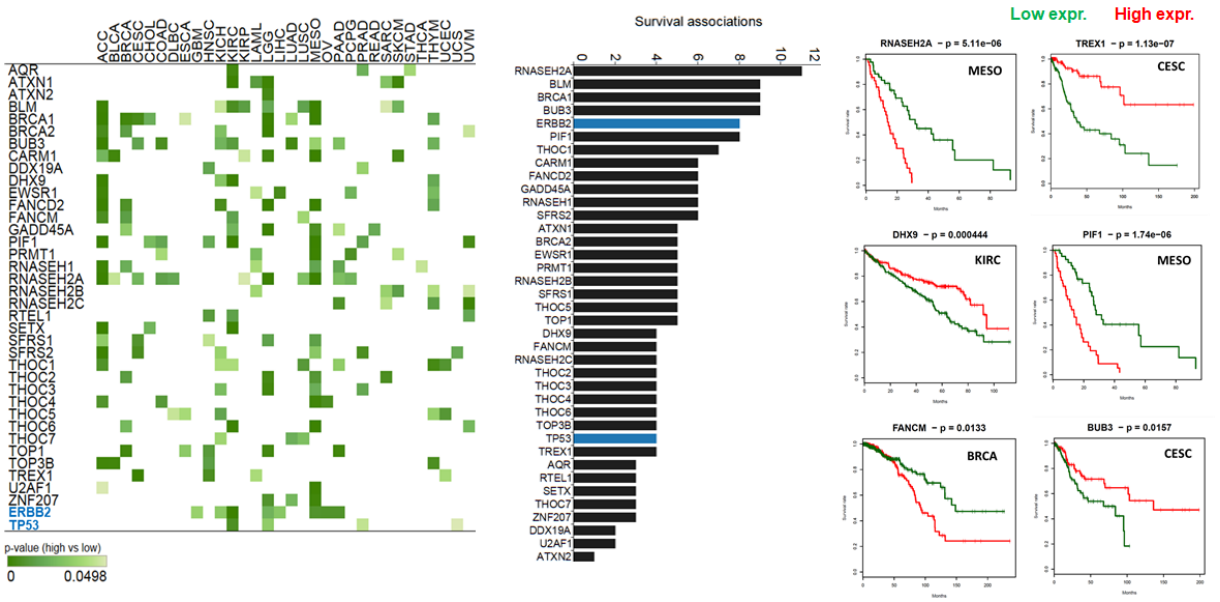
A fenti összefüggések mélyebb megértéséhez farmakogenomikai vizsgálatokat végeztünk a szakirodalomban azonosított (és kísérletesen validált) R-hurok regulátor fehérjékkel, amely azon az elképzelésen alapult, hogy a daganatok permanensen magas R-hurok szintje kihasználható a rákos sejtek kemoterápiás kezelésekkel szembeni érzékennyé tételére. A tumorsejtek R-hurok struktúrákra adott endogén válaszát felerősítve ugyanis sebezhetőbbé tehetők bizonyos citosztatikumokkal szemben (Yun et al. 2017; Tumini et al. 2018; Casado et al. 2008; Monk et al. 2016; Gorthi et al. 2018; Drew et al. 2017; Marinello et al. 2016; Chakraborty et al. 2018; Wan et al. 2015; Sorrells et al. 2018; Flanagan et al. 2009), amely végsősoron új R-hurok alapú kombinációs terápiák kifejlesztéséhez vezethet. Vizsgálatainkhoz olyan R-hurok regulátor géneket választottunk (**2. táblázat**), amelyek egyértelmű R-hurok asszociációval rendelkeznek a kísérletes adatok alapján, különféle molekuláris útvonalakon működnek (pl. DNS hibajavítás, splicing, hiszton modifikáció, stb.), és bizonyítottan szerepet játszanak a rák kialakulásában. Ezen R-hurok regulátorok mRNS szignatúráit 33 primer tumorban vizsgáltuk (**3. táblázat**) a Cancer Genome Atlas (TCGA) adatbázis mRNS-seq adatai alapján (The Cancer Genome Atlas Research Network et al. 2013), amelyet egészséges kontroll szövetekhez viszonyítva számos tumor-specifikus génexpressziós különbségeket azonosítottunk (**51. ábra**). Például a BUB3, DHX9, PRMT1, THOC4, THOC7, U2AF1 és ZNF207 (BUGZ) gének expressziója számos primer tumorban szignifikáns növekedést mutatott a normál szövetekhez képest, míg az SRSF1 (ASF/SF2) splicing faktor az összes ráktípusban represszálódott (a LAML kivételével).



51. ábra Az R-hurok regulátor gének mRNS expressziós szintje egészséges szövetekben és primer tumorokban. Az adatok több ezer betegből illetve szövetmintából meghatározott mRNS-seq értékek (medián), amelyet a Cancer Genome Atlas (TCGA) adatbázis alapján számítottunk ki. A zöld és piros cellák intenzitása arányos az mRNS expresszió mértékével. Az R-hurok gének sorok szerint, míg a szövet illetve tumor típusok oszlopok szerint vannak rendezve.

Ezt követően a primer tumorokban mért génexpressziós értékeket nagyszámú rákos betegcsoport túlélési adataival kapcsoltuk össze, majd megvizsgáltuk, hogy az R-hurok regulátorok transzkripciós szintje korrelál-e a betegek túlélési arányával. Elemzésünkben 12,862 Kaplan-Meier túlélési görbét állítottunk elő és 179 olyan tumor típust azonosítottunk, amely szignifikáns R-hurok regulátor-függő túlélés asszociációt mutatott (**52. ábra, bal panel**). Az R-hurok regulátorokat magasan expresszáló és alacsonyan expresszáló betegcsoportok túlélési esélyei szignifikánsan és egyértelműen elváltak egymástól, és az esetek 70%-ában (n=123) az R-hurok regulátorok alacsony expressziós szintje volt előnyös a betegség lefolyása szempontjából. További megfigyelés, hogy bizonyos gének, pl. az RNASEH2A, BLM, és BRCA1 (top 3), számos ráktípusban mutattak szignifikáns túlélés-asszociációt, amely ezen fehérjék általános szerepére utal, így potenciális biomarkerként tekinthetünk rájuk. Ezzel szemben más gének, pl. az ATXN2, DDX19A és U2AF1 (utolsó 3), hatása csak bizonyos tumor típusokra korlátozódott, amely nagyfokú szövetspecificitásukra utal és csökkenti diagnosztikai vagy terápiás jelentőségüket (**52. ábra, középső panel**). Néhány jellegzetes Kaplan-Meier görbe az **52. ábrán** látható (jobb oldali panel), amely mesothelioma (MESO), cervix carcinoma (CESC), és emlő

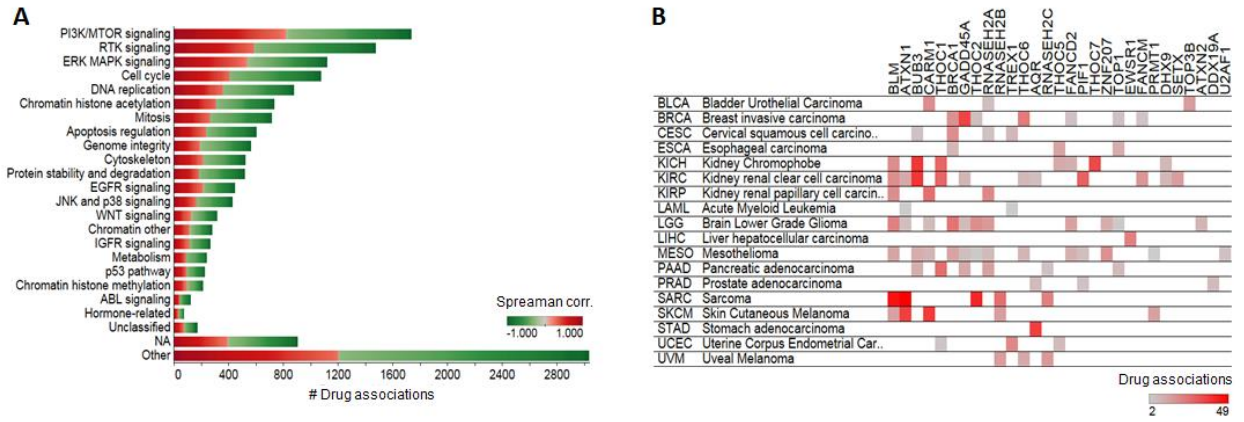
carcinoma (BRCA) tumorokban mutatja a túlélési esélyeket az RNASEH2A, TREX1, DHX9, PIF1, FANCM, és BUB3 gének expressziójának függvényében.



52. ábra Az R-hurok regulátorok túlélés asszociációi primer tumorokban. Bal panel: Hőtérkép, amely a szignifikáns túlélés asszociációt mutató R-hurok regulátorokat mutatja (zöld színű cellák). A zöld szín árnyalata fordítottan arányos a p értékkel. Az R-hurok gének sorok szerint, míg a tumor típusok oszlopok szerint vannak rendezve. Középső panel: Az R--hurok regulátorok túlélés asszociációinak száma, csökkenő sorrendben. Elemzésünkbe pozitív kontrollként R-hurok asszociációval nem rendelkező onkogént (ERBB2 - HER2) és tumorszuppresszort (TP53) is bevontunk (kék oszlopok), amelyekhez képest az R-hurok regulátorok hasonlítva meg lehet ítélni a számításaink biológiai relevanciáját. Az R-hurok regulátorok (fekete oszlopok) hasonló mértékű túlélés asszociációt mutattak, mint a pozitív kontroll gének. Jobb panel: Reprezentatív Kaplan-Meier görbék, amelyek 6 R-hurok regulátor túlélés asszociációját mutatják génexpresszió szerint csoportosítva az eseteket. Magas expressziós szint (piros), alacsony expressziós szint (zöld). A két csoportot az adott tumorban mért medián mRNS-seq érték alapján választottuk el egymástól. A tumor típusát a jobb felső sarokban jeleztük. A rövidítések a 3. táblázatban találhatóak (ld. Anyagok és Módszerek).

A következő lépésben szinergisztikus gyógyszer kölcsönhatásokat kerestünk a túlélés asszociációt mutató primer tumorokkal szövettanilag és genetikailag egyező tumorsejtvonalakban. Vizsgálatunkban 267 FDA (US Food and Drug Administration) által engedélyezett terápiás készítmény érzékenységi interakcióit határoztuk meg 1065 tumorsejtvonal farmakogenomikai adatai alapján (GDSC, (Yang et al. 2013). Számításainkban azonosítottuk azokat a tumor típusokat, amelyek érzékenyebben reagálnak (vagy rezisztenssé válnak) bizonyos terápiákra egy adott R-hurok regulátor expressziójának függvényében. Például, a CX-5461 nevű kemoterápiás szerrel (RNS polimeráz I inhibitor, G4 ligand) kezelt sarcoma sejtvonalban a THOC2 R-hurok gén magas expressziója magas IC_{50} értékekkel (félhalálos dózis) korrelált, vagyis a CX-5461 nem hatékony THOC2-pozitív sarcoma-k kezelésére, használata kerülendő. Ezzel szemben a CX-5461-el kezelt melanoma és endometriális tumorsejtekben az ATXN1 és TREX1 R-hurok gének magas expressziója

alacsony IC_{50} értékekkel korrelált, vagyis a ATXN1-pozitív és TREX1-pozitív tumorok érzékenyen reagálnak a CX-5461 terápiára, használatuk javasolt. A vizsgált terápiás szerek számos biológiai útvonalat felöleltek (a protein kináz és MAPK szignalizációtól a citoszkeletonon át a DNS hibajavításig), amelyek közül az PI3K/mTOR, MAPK/sejtciklus és DNS replikáció útvonalakra ható szerek mutatták a legtöbb génkölcsönhatást (**53/A ábra**). Összességében 508 olyan R-hurok regulátor-függő gyógyszer kölcsönhatást detektáltunk, amely egyben szignifikáns túlélés asszociációt is mutatott a megfelelő primer tumoros populációban (**53/B ábra**).



53. ábra Az R-hurok regulátorok gyógyszer kölcsönhatásai. Az R-hurok regulátorok génexpresszió-függő drug-szenzitivitás asszociációit 276 FDA által engedélyezett szerre számítottuk ki 1065 tumorsejtvonal adatai alapján, amelyet a a GDSC adatbázisból töltöttünk le. A drug asszociációkat az IC_{50} értékek (félhalálos dózis) és az adott R-hurok regulátor gén medián mRNS-seq expressziójának Spearman rangkorrelációja alapján határoztuk meg. *A.* Az R-hurok regulátorok gyógyszerkölcsönhatásai a terápiás szerek molekuláris útvonalai szerint rendezve. Zöld és piros színek jelzik a pozitív és negatív gyógyszer kölcsönhatásokat, amely az adott R-hurok regulátor magas (piros) vagy alacsony (zöld) expressziójával korrelál. *B.* A szignifikáns túlélés-asszociációt mutató tumorok gyógyszerkölcsönhatásai. Sorok: tumor típusok. Oszlopok: R-hurok regulátor gének. A cellák a szignifikáns drug interakciókat mutatják. A piros szín árnyalata arányos az adott tumorban adott R-hurok expresszió esetén mért drug interakciók számával.

Az R-hurok regulátorok expressziós státusza tehát jelentősen befolyásolja a betegek túlélési esélyeit valamint a gyógyszeres kezelésekre adott választ is módosíthatja. Ezen farmakogenomikai összefüggések klinikai felhasználása még várat magára, azonban megfelelő preklinikai vizsgálatok elvégzése után alkalmas lehet a betegcsoportok prioritizálására. A rákos sejtekben zajló R-hurok képződési folyamatok és az azokat kontrolláló fehérjék tehát potenciális biomarkerek és/vagy terápiás célpontok, amelyek alkalmasak lehetnek egyes daganattípusok kemoterápiás szerekkel szembeni érzékenyítésére. Eredményeink, az alapkutatási jelentőségükön túl, új utat nyithatnak az epigenetikai terápiák továbbfejlesztéséhez, amely a daganatok R-hurok szintjének módosításán alapul.

ÚJ MEGÁLLAPÍTÁSOK

A meiotikus homológ rekombinációval kapcsolatos munkánkban megmutattuk, hogy a Set1C hiszton metiláz komplex által katalizált H3K4me3 okozati szerepet játszik a DNS duplaszáltörések kialakulásában. Igazoltuk, hogy a DSB-k kialakulása függ a módosítható lizin 4 oldallánc (H3K4) jelenlététől, a Set1 alegység katalitikus aktivitásától, és az Spp1 alegység jelenlététől. Bizonyítottuk, hogy az Spp1 a DSB képződés során i) fizikai kölcsönhatásba lép a DSB-komplex egyik tagjával (Mer2), amely egy cinkujjszerű Mer2-interakciós doménen keresztül történik (CXXC Mer2-ID), ii) felismeri és megköti a Set1-által katalizált H3K4me3 szignált, amely a PHD-finger doménje által történik. Bizonyítottuk, hogy a Mer2-ID és PHD-domének delációja a DSB-szint jelentős csökkenését okozza. Egy új rekombinációs modellt vezettünk be, amely egyrészt kiterjesztette a régebbi hurok-tengely modellt (Blat et al. 2002), másrészt az Spp1-Mer2 kölcsönhatás alapján molekuláris szinten megmagyarázta a meiotikus DNS-törések kapcsolatát a Set1C-katalizált H3K4me3 módosítással. A modell szerint a meiotikus homológ rekombináció iniciációjához szükség van H3K4me3 modifikációra, amelyet a Set1C katalizál, szükség van Mer2 DSB-fehérjére, amely a Spo11-kompexet rögzíti a kromoszóma tengelyhez (Panizza et al. 2011), és szükség van Spp1-re, amely a PHD-finger és Mer2-ID doménjén keresztül fizikai kapcsolatot teremt az epigenetikai jel és a DNS törések között. Megmutattuk, hogy az Spp1 diffúziós tulajdonságai és dinamikus turnover rátája alapvető fontosságú a rekombinációt megelőző kromatinváltozások kialakulásához. Ennek kapcsán bizonyítottuk, hogy az Spp1 Set1C/COMPASS-független működésre vált a meiózis során, amely elősegíti a rekombinációs iniciációs helyek aktiválását.

A Set1C mellett egyéb epigenetikai modifikációk szerepét is bizonyítottuk a meiotikus rekombináció során. Megmutattuk, hogy a központi hiszton-fold doménre koncentrálódó H3 lizin 56 acetiláció (H3K56ac) fiziológiás szintre állítja be meiotikus DSB-k mennyiségét és ezzel finomhangolja a rekombinációs forrópontok működését.

Az R-hurok struktúrákkal kapcsolatos munkánk során kidolgoztunk egy nagypontosságú analitikai eljárást, amely az R-hurkok azonosítására és genomi térképezésére alkalmas bármely modellorganizmusban. Felismertük az ún. “első-exon torzítás” jelenségét, amely bizonyos genomfragmentációs eljárásoknál torzítja az R-hurkok méreteloszlását és genomi térképezhetőségét, és ezáltal a biológiai funkciójuk megítélését. E metodikára alapozva R-hurok térképezést végeztünk *Arabidopsis thaliana*-ban, hogy az addig egyetlen ismert R-hurok-kötő fehérje, a *Nodulin homeobox* (NDX) működését molekuláris szinten megértsük. Az NDX fehérjéről korábban azt feltételezték, hogy egy általános R-hurok regulátor, amely kötődésével stabilizálja e struktúrákat és ezzel bizonyos gének aktivitását befolyásolja (Sun et al. 2013). Ezzel szemben mi megmutattuk,

hogy az NDX egy alapvető heterokromatin regulátor, amely a korábbi elképzelésekkel szemben nem az R-hurok-gazdag eukromatikus kromoszóma karokon, hanem a heterokromatikus régiókban működik. Bizonyítottuk, hogy az NDX kromoszómális eloszlása ellentétes az R-hurkokhoz képest és az RNS-függő DNS metilációs rendszerrel (RdDM) együttműködve a heterokromatin homeosztázisát szabályozza a centromerikus és pericentrometikus régiókban. Az NDX hatásmechanizmusával összefüggő genomszerkezet változások detektálásához genomkonformáció leképezés alapú módszereket implementáltunk a laboratóriumomban. NDX hiánymutánsban végzett Hi-C vizsgálatainkban jelentős genomszerkezet változásokat detektáltunk az NDX kötőhelyekben gazdag genomi régiókban (pericentromerek), amely erősen korrelált a megfigyelt het-siRNS expresszió és DNS metilációs változásokkal. E molekuláris változások fenotípusos következményeit további funkcionális vizsgálatok elvégzésével próbáljuk feltérképezni.

Humánspecifikus R-hurok regulátor fehérjékkel végzett farmakogenomikai vizsgálatainkban megmutattuk, hogy R-hurok regulátorok expressziós szintje kihasználható a rákos sejtek kemoterápiás kezelésekkel szembeni érzékennyé tételére. Emellett számos olyan tumor típust azonosítottunk, amely szignifikáns R-hurok regulátor-függő túlélés asszociációt mutatott, vagyis az R-hurok regulátorokat magasan expresszáló és az alacsonyan expresszáló betegcsoportok túlélési esélyei szignifikánsan elváltak egymástól. A fenti összefüggések miatt az R-hurok regulátorokat mindenképp érdemes potenciális diagnosztikai markerként vagy terápiás célpontként tekintetbe venni.

KITEKINTÉS

A meiotikus rekombinációval kapcsolatos eredményeinknek lehetséges olyan hosszútávú kimenete, amellyel a terméketlenség kockázata a jövőben csökkenthető lesz új biomarkerek vagy epigenetikai terápiák révén. Ha egy női petesejt például 35-40 éves kor felett termékenyül meg, ez azt jelenti, hogy az éretlen petesejt évtizedekig volt olyan állapotban, amikor reparálatlan DNS-törések és éretlen crossover intermedierek voltak benne (nőkben a meiózis szakaszosan és időben rendkívül elnyújtva történik; születéskor a meiózis I profázisában megáll a folyamat, amely az ovuláció során folytatódik). A magzati kor óta törött kromoszómák tehát hosszú évekig ebben az állapotban vannak és ezalatt a kromoszómák között kialakult fizikai kapcsolatok meggyengülnek vagy elromlanak. Emiatt a 35 év feletti anyáknál exponenciálisan nő a kromoszóma rendellenességek gyakorisága az utódokban, mint amilyen a Down-kór. A DNS-törések helyének, létrejöttük mechanizmusának, és epigenetikai szabályozásának megismerése ezért valóban komoly orvosi jelentőségű. Az tényként kezelhető, hogy eredményeink alapján már most meg tudjuk érteni a női meddőség elvi alapjait. Sokszor az a fő probléma, hogy elmaradnak vagy rossz helyen keletkeznek a DNS-törések, és így nincs homológ rekombináció sem; a crossover rekombinánsok keletkezése pedig "kötelező" feltétele a homológok szétválásának az első meiotikus osztódás során. Emellett ha sérülnek, vagy mutációt szenvednek a DSB-ket közvetlenül szabályozó enzimek (ilyen például, a Spo11-komplex vagy a COMPASS), akkor infertilitás lehet a genetikai következmény. A DSB fehérjék azonosítása tehát a miénkhez hasonló élesztő modellkísérletekben lehetővé teszi a humán funkcionális ortológok megtalálását és későbbiekben ezek klinikai szűrését.

A következő fontos megválaszolatlan kérdés, hogy mi történik azokkal a DSB-kkel, amelyek rendkívül kis számban, de mégis létrejönnek a Set1C/COMPASS hiányában. Ezek vajon hasonlóak a fiziológiás DNS törésekhez és ugyanolyan rekombinációs útvonalakon processzálódnak? Vagy pedig ezek patológiás DSB-k, egészen más élettani következményekkel? Jelenleg azt gondoljuk, hogy az ektópikus DSB-k rekombinogén hatása valamennyire fajfüggő. Egérben például Prdm9 KO mutánsok teljes mértékben infertilisek (Hayashi et al. 2005), mivel a DSB-k géngazdag területek felé irányítódnak és többségük nem vesz részt a klasszikus interhomológ rekombinációs folyamatokban (Brick et al. 2012). Prdm9 hiszton metiláz nélkül tehát egérben nem fejlődnek egészséges ivarsejtek. Élesztő esetében viszont meglepő módon azt tapasztaltuk, hogy a Set1C inaktiválása nem befolyásolta a halpoid spórák keletkezését és szaporodási képességét, vagyis ebben az organizmusban az ektópikus DSB-k valószínűleg rekombinogének voltak. Ezeket a kérdéseket teljes-

genom rekombinációs térképezéssel lehetne tisztázni, amely metodikáját élesztőben már beállítottuk (tetrad-seq).

A rekombinációval és R-hurkokkal kapcsolatos eredményeink egyaránt előrelépést hozhatnak a malignus sejttranszformáció kutatásában. A rákos folyamatra a genom instabilitása jellemző, amely során törés-rekombináció ciklusok történnek kontrollálatlanul. Az R-hurkok hipermutagének és rekombinogének, ám azt jelenleg nem tudjuk, hogy pontosan milyen útvonalon és mely rekombinációs enzimek által processzálódnak. A rekombinációs útvonalak feltérképezése és terápiás modulálása hatékonyabb tumor eradikációhoz vezethet. A tumorok R-hurkok szintjét pedig az R-hurok regulátorokon keresztül módosíthatjuk, amelyek expressziója a tumoros szövet gyógyszerérzékenységét is befolyásolhatja. Ez viszont a terápia hatékonyságára és a betegek túlélésére lehet hatással. Ezek az összefüggések utat nyithatnak az R-hurok regulátorok klinikai kiaknázásához diagnosztikai markerként vagy terápiás célpontként, amely hosszútávon új R-hurok alapú kombinációs terápiák kifejlesztéséhez vezethet.

A fenti epigenetikai folyamatoknak van egy metabolomikai aspektusa is, amelyet az utóbbi években kezdtünk el részletesebben vizsgálni az indolamin-2,3-dioxigenáz (IDO) enzim kapcsán (Hornyák et al. 2018)). Minden olyan anyagcsere útvonal, amely NAD-ot, ATP-t, és aktív metionint (SAM) fogyaszt vagy termel, befolyásolja a hisztonmódosító enzimek aktivitását és ezáltal az epigenetikai mintázatok kialakulását. Az indolamin-2,3-dioxigenáz (IDO) kinurenin termelése révén jelentős immunmoduláló és immunszuppresszív hatással rendelkezik humán tumorokban, azonban nagymennyiségű NAD-ot és ATP-t is generál, amely a hiszton deacetilázok működéséhez szükséges. Az IDO epigenetikai szerepének és a hozzá hasonló onkometabolit-termelő enzimek működésének molekuláris vizsgálata komplex megközelítéseket igényel, amely a genomikai és molekuláris biológiai módszerek mellett magába foglalja a metabolomika eszköztárát is.

Fontos kihangsúlyozni, hogy a fenti eredmények gyakorlati alkalmazása egyelőre csupán elvi lehetőség, spekuláció. Jelenleg ott tartunk, hogy bizonyos molekuláris mechanizmusokat megértettük, bizonyos fehérjéket karakterizáltunk, és bizonyos kölcsönhatásokat validáltunk. Alapkutatási eredményeinknek tehát nincs közvetlen gyakorlati haszna, azonban klinikai vagy egyéb alkalmazásukat mindenképp érdemes továbbgondolni.

KÖSZÖNETNYILVÁNÍTÁS

Az értekezésben bemutatott eredmények a Genomszerkezet és Rekombináció Kutatócsoport munkájából származnak, amelyért elsősorban Nekik tartozom köszönettel: Karányi Zsolt, Miskei Márton, Feró Orsolya, Nagy Éva, Horváth Adrienn, Boros-Oláh Beáta, Fillér Csaba, Hornyák Lilla, Fazekas-Bálint Ágnes, Varga Dóra, Mosolygó-Lukács Ágnes, Nagy Dénes, Sipos Éva, Hetey Szabolcs, Halász László, Fürtös Ibolya, Petrucz Anita.

A kutatócsoport nem jöhetett volna létre a Magyar Tudományos Akadémia Lendület programjának támogatása nélkül, amelyért köszönettel tartozom az MTA döntéshozóinak és a bírálóknak.

Hálás vagyok a Debreceni Egyetem ÁOK Biokémiai és Molekuláris Biológiai Intézetnek, hogy 2015-ben befogadta a csoportomat, lehetőséget és teret biztosítva a kutatásainknak.

Végül, köszönettel tartozom szeretett családomnak: hálás vagyok végtelen türelmetekért és bátorításotokért.

IRODALOMJEGYZÉK

Abshiru N, Ippersiel K, Tang Y, Yuan H, Marmorstein R, Verreault A, Thibault P. 2013. Chaperone-mediated acetylation of histones by Rtt109 identified by quantitative proteomics. *J Proteomics* **81**: 80–90. http://dx.doi.org/10.1016/j.jprot.2012.09.026.

Acquaviva L, Székvölgyi L, Dichtl B, Dichtl SB, de La Roche Saint André C, Nicolas A, Géli V. 2013. The COMPASS subunit Spp1 links histone methylation to initiation of meiotic recombination. *Science (80-)* **339**: 215–8. http://www.ncbi.nlm.nih.gov/pubmed/23160953 (Accessed February 28, 2013).

Adolph KW, Cheng SM, Laemmli UK. 1977. Role of nonhistone proteins in metaphase chromosome structure. *Cell* **12**: 805–816. http://www.cell.com/article/0092867477902793/fulltext (Accessed October 15, 2021).

Agarwal N, Hardt T, Brero A, Nowak D, Rothbauer U, Becker A, Leonhardt H, Cardoso MC, Delbru M. 2007. MeCP2 interacts with HP1 and modulates its heterochromatin association during myogenic differentiation. *Nucleic Acids Res* **35**: 5402–5408.

Aguilera A, García-Muse T. 2012. R loops: from transcription byproducts to threats to genome stability. *Mol Cell* **46**: 115–24. http://www.ncbi.nlm.nih.gov/pubmed/22541554 (Accessed March 7, 2013).

Alecki C, Chiwara V, Sanz LA, Grau D, Pérez OA, Armache K, Chédin F, Francis NJ. 2020. RNA-DNA strand exchange by the Drosophila Polycomb complex PRC2. *Nat Commun* **11**: 1781. http://dx.doi.org/10.1038/s41467-020-15609-x.

Allshire RC, Madhani HD. 2017. Ten principles of heterochromatin formation and function. *Nat Rev Mol Cell Biol* **19**: 229–244. https://www.nature.com/articles/nrm.2017.119 (Accessed August 5, 2021).

Alzu A, Bermejo R, Begnis M, Lucca C, Piccini D, Carotenuto W, Saponaro M, Brambati A, Cocito A, Foiani M, et al. 2012. Senataxin associates with replication forks to protect fork integrity across RNA-polymerase-II-transcribed genes. *Cell* **151**: 835–46. http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3494831&tool=pmcentrez&rendertype=abstract (Accessed March 10, 2013).

Arab K, Karaulanov E, Musheev M, Trnka P, Schäfer A, Grummt I, Niehrs C. 2019. GADD45A binds R-loops and recruits TET1 to CpG island promoters. *Nat Genet* **51**: 217–223.

Ariel F, Lucero L, Christ A, Mammarella MF, Jegu T, Veluchamy A, Mariappan K, Latrasse D, Blein T, Liu C, et al. 2020. R-Loop Mediated trans Action of the APOLO Long Noncoding RNA. *Mol Cell* **77**: 1055-1065.e4. https://doi.org/10.1016/j.molcel.2019.12.015.

Arora C, Kee K, Maleki S, Keeney S. 2004. Antiviral protein Ski8 is a direct partner of Spo11 in meiotic DNA break formation, independent of its cytoplasmic role in RNA metabolism. *MolCell* **13**: 549–559.

Arudchandran a, Cerritelli S, Narimatsu S, Itaya M, Shin DY, Shimada Y, Crouch RJ. 2000. The absence of ribonuclease H1 or H2 alters the sensitivity of Saccharomyces cerevisiae to hydroxyurea, caffeine and ethyl methanesulphonate: implications for roles of RNases H in DNA replication and repair. *Genes Cells* **5**: 789–802. http://www.ncbi.nlm.nih.gov/pubmed/11029655.

Auton A, Fledel-Alon A, Pfeifer S, Venn O, Ségurel L, Street T, Leffler EM, Bowden R, Aneas I, Broxholme J, et al. 2012. A fine-scale chimpanzee genetic map from population sequencing. *Science* **336**: 193–8. http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3532813&tool=pmcentrez&rendertype=abstract (Accessed March 21, 2013).

Auton A, Rui Li Y, Kidd J, Oliveira K, Nadel J, Holloway JK, Hayward JJ, Cohen PE, Greally JM, Wang J, et al. 2013. Genetic recombination is targeted towards gene promoter regions in dogs. *PLoS Genet* **9**: e1003984. http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3861134&tool=pmcentrez&rendertype=abstract (Accessed March 24, 2014).

Bae HJ, Dubarry M, Jeon J, Soares LM, Dargemont C, Kim J, Geli V, Buratowski S. 2020. The Set1 N-terminal domain and Swd2 interact with RNA polymerase II CTD to recruit COMPASS. *Nat Commun* **11**: 2181. http://dx.doi.org/10.1038/s41467-020-16082-2.

Balk B, Maicher A, Dees M, Klermund J, Luke-Glaser S, Bender K, Luke B. 2013. Telomeric RNA-DNA hybrids affect telomere-length dynamics and senescence. *Nat Struct Mol Biol* **20**: 1199–1206.

Battulin N, Fishman VS, Mazur AM, Pomaznoy M, Khabarova AA, Afonnikov DA, Prokhortchouk EB, Serov OL. 2015. Comparison of the three-dimensional organization of sperm and fibroblast genomes using the Hi-C approach. *Genome Biol* **16**. http://dx.doi.org/10.1186/s13059-015-0642-0.

Baudat F, Nicolas A. 1997. Clustering of meiotic double-strand breaks on yeast chromosome III. *ProcNatlAcadSciUSA* **94**: 5213–5218.

Becherel OJ, Sun J, Yeo AJ, Nayler S, Fogel BL, Gao F, Coppola G, Criscuolo C, De Michele G, Wolvetang E, et al. 2015. A new model to study neurodegeneration in ataxia oculomotor apraxia type 2. *Hum Mol Genet* **24**: 5759–5774.

Benore-Parsons M, Ayoub MA. 1997. Presence of RNase a causes aberrant DNA band shifts. *Biotechniques* **23**: 128–131.

Berezney R, Coffey DS. 1974. Identification of a nuclear protein matrix. *Biochem Biophys Res Commun* **60**: 1410–1417.

Berg IL, Neumann R, Sarbajna S, Odenthal-Hesse L, Butler NJ, Jeffreys AJ. 2011. Variants of the protein PRDM9 differentially regulate a set of human meiotic recombination hotspots highly active in African populations. *Proc Natl Acad Sci U S A* **108**: 12378–83. http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3145720&tool=pmcentrez&rendertype=abstract (Accessed March 21, 2013).

Bhatia V, Barroso SI, García-Rubio ML, Tumini E, Herrera-Moyano E, Aguilera A. 2014. BRCA2 prevents R-loop accumulation and associates with TREX-2 mRNA export factor PCID2. *Nature* **511**: 362–5. http://www.nature.com/doifinder/10.1038/nature13374 (Accessed June 4, 2014).

Biswas M, Voltz K, Smith JC, Langowski J. 2011. Role of histone tails in structural stability of the nucleosome. *PLoS Comput Biol* **7**: 1–12. http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3240580&tool=pmcentrez&rendertype=abstract (Accessed March 4, 2013).

Blackledge NP, Klose RJ. 2021. The molecular principles of gene regulation by Polycomb repressive

complexes. *Nat Rev Mol Cell Biol* **0123456789**. http://dx.doi.org/10.1038/s41580-021-00398-y.

Blat Y, Protacio RU, Hunter N, Kleckner N. 2002. Physical and functional interactions among basic chromosome organizational features govern early steps of meiotic chiasma formation. *Cell* **111**: 791–802. http://www.ncbi.nlm.nih.gov/pubmed/12526806.

Blevins T, Podicheti R, Mishra V, Marasco M, Wang J, Rusch D, Tang H, Pikaard CS. 2015. Identification of Pol IV and RDR2- dependent precursors of 24 nt siRNAs guiding de novo DNA methylation in Arabidopsis. *Elife* **4**: e09591.

Bojcsuk D, Nagy G, Balint BL. 2017. Inducible super-enhancers are organized based on canonical signal-specific transcription factor binding elements. *Nucleic Acids Res* **45**: 3693–3706.

Bolzer A, Kreth G, Solovei I, Koehler D, Saracoglu K, Fauth C, Müller S, Eils R, Cremer C, Speicher MR, et al. 2005. Three-dimensional maps of all chromosomes in human male fibroblast nuclei and prometaphase rosettes. *PLoS Biol* **3**: 0826–0842.

Boque-Sastre R, Soler M, Oliveira-Mateos C, Portela A, Moutinho C, Sayols S, Villanueva A, Esteller M, Guil S. 2015. Head-to-head antisense transcription and R-loop formation promotes transcriptional activation. *Proc Natl Acad Sci* **112**: 201421197. http://www.pnas.org/lookup/doi/10.1073/pnas.1421197112.

Borde V, Goldman ASH, Lichten M. 2000. Direct coupling between meiotic DNA replication and recombination initiation. *Science (80-)* **290**: 806–809.

Borde V, Robine N, Lin W, Bonfils S, Geli V, Nicolas A. 2009. Histone H3 lysine 4 trimethylation marks meiotic recombination initiation sites . *EMBO J* **28**: 99–111.

Boulé J-B, Zakian V a. 2007. The yeast Pif1p DNA helicase preferentially unwinds RNA DNA substrates. *Nucleic Acids Res* **35**: 5809–18. http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2034482&tool=pmcentrez&rendertype=abstract (Accessed March 21, 2013).

Brick K, Smagulova F, Khil P, Camerini-Otero RD, Petukhova G V. 2012. Genetic recombination is directed away from functional genomic elements in mice. *Nature* **485**: 642–5. http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3367396&tool=pmcentrez&rendertype=abstract (Accessed March 4, 2013).

Buenrostro JD, Giresi PG, Zaba LC, Chang HY, Greenleaf WJ. 2013. Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. *Nat Methods* **10**: 1213–1218.

Buhler C, Borde V, Lichten M. 2007. Mapping meiotic single-strand DNA reveals a new landscape of DNA double-strand breaks in Saccharomyces cerevisiae. *PLoS Biol* **5**: e324. http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2121111&tool=pmcentrez&rendertype=abstract (Accessed March 18, 2013).

Buning R, Van Noort J. 2010. Single-pair FRET experiments on nucleosome conformational dynamics. *Biochimie* **92**: 1729–1740. http://www.ncbi.nlm.nih.gov/pubmed/20800089 (Accessed March 6, 2013).

Busslinger GA, Stocsits RR, Van Der Lelij P, Axelsson E, Tedeschi A, Galjart N, Peters JM. 2017. Cohesin is positioned in mammalian genomes by transcription, CTCF and Wapl. *Nature* **544**:

503–507. http://dx.doi.org/10.1038/nature22063.

Calladine C. 2004. *Understanding DNA: The Molecule and How it Works*.

Camerini-Otero RD. 2014. DNA recombination. Recombination initiation maps of individual human genomes. *Science (80-)* **346**: 1256442.

Casado JA, Rio P, Marco E, Garcia-Hernandez V, Domingo A, Perez L, Tercero JC, Vaquero JJ, Albella B, Gago F, et al. 2008. Relevance of the Fanconi anemia pathway in the response of human cells to trabectedin. *Mol Cancer Ther* **7**: 1309–1318. http://mct.aacrjournals.org/cgi/doi/10.1158/1535-7163.MCT-07-2432.

Celic I, Masumoto H, Griffith WP, Meluh P, Cotter RJ, Boeke JD, Verreault A. 2006. The Sirtuins Hst3 and Hst4p Preserve Genome Integrity by Controlling Histone H3 Lysine 56 Deacetylation. *Curr Biol* **16**: 1280–1289.

Chakraborty P, Huang JTJ, Hiom K. 2018. DHX9 helicase promotes R-loop formation in cells with impaired RNA splicing. *Nat Commun* **9**: 4346. http://www.ncbi.nlm.nih.gov/pubmed/30341290%0Ahttp://www.nature.com/articles/s41467-018-06677-1.

Chen L, Chen J-Y, Zhang X, Gu Y, Xiao R, Shao C, Tang P, Qian H, Luo D, Li H, et al. 2017. R-ChIP Using Inactive RNase H Reveals Dynamic Coupling of R-loops with Transcriptional Pausing at Gene Promoters. *Mol Cell* 1–13. http://linkinghub.elsevier.com/retrieve/pii/S1097276517307578.

Chen Y, Zhang Y, Wang Y, Zhang L, Brinkman EK, Adam SA, Goldman R, Steensel B van, Ma J, Belmont AS. 2018. Mapping 3D genome organization relative to nuclear compartments using TSA-Seq as a cytological ruler. *J Cell Biol* jcb.201807108. http://jcb.rupress.org/content/early/2018/08/27/jcb.201807108.

Cheng H, He X, Moore C. 2004. The Essential WD Repeat Protein Swd2 Has Dual Functions in RNA Polymerase II Transcription Termination and Lysine 4 Methylation of Histone H3. *Mol Cell Biol* **24**: 2932–2943.

Choi K, Zhao X, Kelly KA, Venn O, Higgins JD, Yelina NE, Hardcastle TJ, Ziolkowski PA, Copenhaver GP, Franklin FCH, et al. 2013. Arabidopsis meiotic crossover hot spots overlap with H2A.Z nucleosomes at gene promoters. *Nat Genet* **45**: 1327–1338. https://www.nature.com/ng/journal/v45/n11/abs/ng.2766.html.

Chon H, Sparks JL, Rychlik M, Nowotny M, Burgers PM, Crouch RJ, Cerritelli SM. 2013. RNase H2 roles in genome integrity revealed by unlinking its activities. *Nucleic Acids Res* **41**: 3130–3143.

Chua E, Bates A, Maxwell A. 2005. *DNA Topology*.

Chua EYD, Vogirala VK, Inian O, Wong ASW, Nordenskiöld L, Plitzko JM, Danev R, Sandin S. 2016. 3.9 Å structure of the nucleosome core particle determined by phase-plate cryo-EM. *Nucleic Acids Res* **44**: 8013–8019.

Cohen S, Puget N, Lin Y-L, Clouaire T, Aguirrebengoa M, Rocher V, Pasero P, Canitrot Y, Legube G. 2018. Senataxin resolves RNA:DNA hybrids forming at DNA double-strand breaks to prevent translocations. *Nat Commun* **9**: 533. http://www.nature.com/articles/s41467-018-02894-w.

Costantino L, Hsieh THS, Lamothe R, Darzacq X, Koshland D. 2020. Cohesin residency determines chromatin loop patterns. *Elife* **9**: 1–31.

Cristiano S, Leal A, Phallen J, Fiksel J, Adleff V, Bruhm DC, Jensen SØ, Medina JE, Hruban C, White JR, et al. 2019. Genome-wide cell-free DNA fragmentation in patients with cancer. *Nature* **570**: 385–389.

Cristini A, Ricci G, Britton S, Salimbeni S, Huang S yin N, Marinello J, Calsou P, Pommier Y, Favre G, Capranico G, et al. 2019. Dual Processing of R-Loops and Topoisomerase I Induces Transcription-Dependent DNA Double-Strand Breaks. *Cell Rep* **28**: 3167-3181.e6.

Crossley MP, Bocek M, Cimprich KA. 2019. R-Loops as Cellular Regulators and Genomic Threats. *Mol Cell* **73**: 398–411.

Dai X, Zhuang Z, Zhao PX. 2018. PsRNATarget: A plant small RNA target analysis server (2017 release). *Nucleic Acids Res* **46**: W49–W54.

Dehe P-M, Geli V. 2006. The multiple faces of Set1 1. *Biochem Cell Biol* **548**: 536–548.

Dekker J, Mirny L. 2016. The 3D Genome as Moderator of Chromosomal Communication. *Cell* **164**: 1110–1121. http://dx.doi.org/10.1016/j.cell.2016.02.007.

Di Pierro M, Cheng RR, Aiden EL, Wolynes PG, Onuchic JN. 2017. De novo prediction of human chromosome structures: Epigenetic marking patterns encode genome architecture. *Proc Natl Acad Sci U S A* **114**: 12126–12131.

Dixon JR, Jung I, Selvaraj S, Shen Y, Antosiewicz-Bourget JE, Lee AY, Ye Z, Kim A, Rajagopal N, Xie W, et al. 2015. Chromatin architecture reorganization during stem cell differentiation. *Nature* **518**: 331–336. http://www.nature.com/doifinder/10.1038/nature14222.

Domínguez-Sánchez MS, Barroso S, Gómez-González B, Luna R, Aguilera A. 2011. Genome instability and transcription elongation impairment in human cells depleted of THO/TREX. *PLoS Genet* **7**: e1002386. http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3228816&tool=pmcentrez&rendertype=abstract (Accessed March 11, 2013).

Dona F, Houseley J. 2014. Unexpected DNA loss mediated by the DNA binding activity of ribonuclease A. *PLoS One* **9**: 1–11.

Drew AE, Moradei O, Jacques SL, Rioux N, Boriack-Sjodin AP, Allain C, Scott MP, Jin L, Raimondi A, Handler JL, et al. 2017. Identification of a CARM1 Inhibitor with Potent In Vitro and In Vivo Activity in Preclinical Models of Multiple Myeloma. *Sci Rep* **7**: 17993. http://www.nature.com/articles/s41598-017-18446-z (Accessed February 5, 2019).

Drolet M, Phoenix P, Menzel R, Massé E, Liu LF, Crouch RJ. 1995. Overexpression of RNase H partially complements the growth defect of an Escherichia coli delta topA mutant: R-loop formation is a major problem in the absence of DNA topoisomerase I. *Proc Natl Acad Sci U S A* **92**: 3526–30. http://www.ncbi.nlm.nih.gov/pubmed/7536935 (Accessed January 29, 2019).

Dumelie JG, Jaffrey SR. 2017. Defining the location of promoter-associated R-loops at near-nucleotide resolution using bisDRIP-seq. *Elife* **6**. https://elifesciences.org/articles/28306.

Duquette ML, Handa P, Vincent J a., Taylor AF, Maizels N. 2004. Intracellular transcription of G-rich DNAs induces formation of G-loops, novel structures containing G4 DNA. *Genes Dev* **18**: 1618–1629.

Durand NC, Shamim MS, Machol I, Rao SSP, Huntley MH, Lander ES, Aiden EL. 2016. Juicer Provides a One-Click System for Analyzing Loop-Resolution Hi-C Experiments. *Cell Syst* **3**: 95–98.

http://dx.doi.org/10.1016/j.cels.2016.07.002.

Ernst J, Kellis M. 2012. ChromHMM: automating chromatin-state discovery and characterization. *Nat Methods* **9**: 215–6. http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3577932&tool=pmcentrez&rendertype=abstract (Accessed December 14, 2013).

Eskeland R, Leeb M, Grimes GR, Kress C, Boyle S, Sproul D, Gilbert N, Fan Y, Skoultchi AI, Wutz A, et al. 2010. Ring1B Compacts Chromatin Structure and Represses Gene Expression Independent of Histone Ubiquitination. *Mol Cell* **38**: 452–464.

Farrelly LA, Thompson RE, Zhao S, Lepack AE, Lyu Y, Bhanu N V., Zhang B, Loh YHE, Ramakrishnan A, Vadodaria KC, et al. 2019. Histone serotonylation is a permissive modification that enhances TFIID binding to H3K4me3. *Nature* **567**: 535–539. http://dx.doi.org/10.1038/s41586-019-1024-7.

Feng S, Cokus SJ, Schubert V, Zhai J, Pellegrini M, Jacobsen SE. 2014. Genome-wide Hi-C Analyses in Wild-Type and Mutants Reveal High-Resolution Chromatin Interactions in Arabidopsis. *Mol Cell* **55**: 694–707. http://linkinghub.elsevier.com/retrieve/pii/S1097276514006017 (Accessed August 18, 2014).

Fillér C, Hornyák L, Roszik J. 2018. Commentary: Nuclear dynamics of the Set1C subunit Spp1 prepares meiotic recombination sites for break formation. *Front Genet* **9**: 496.

Flanagan JM, Funes JM, Henderson S, Wild L, Carey N, Boshoff C. 2009. Genomics screen in transformed stem cells reveals RNASEH2A, PPAP2C, and ADARB1 as putative anticancer drug targets. *Mol Cancer Ther* **8**: 249–260. http://mct.aacrjournals.org/cgi/doi/10.1158/1535-7163.MCT-08-0636.

Francis NJ, Kingston RE, Woodcock CL. 2004. Chromatin compaction by a polycomb group protein complex. *Science (80-)* **306**: 1574–1577.

Fudenberg G, Imakaev M, Lu C, Goloborodko A, Abdennur N, Mirny LA. 2016. Formation of Chromosomal Domains by Loop Extrusion. *Cell Rep* **15**: 2038–2049. http://dx.doi.org/10.1016/j.celrep.2016.04.085.

Gan W, Guan Z, Liu J, Gui T, Shen K, Manley JL, Li X. 2011. R-loop-mediated genomic instability is caused by impairment of replication fork progression. *Genes Dev* **25**: 2041–56. http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3197203&tool=pmcentrez&rendertype=abstract (Accessed March 6, 2013).

Gao Z, Liu HL, Daxinger L, Pontes O, He X, Qian W, Lin H, Xie M, Lorkovic ZJ, Zhang S, et al. 2010. An RNA polymerase II-and AGO4-associated protein acts in RNA-directed DNA methylation. *Nature* **465**: 106–109.

García-Muse T, Aguilera A. 2019a. R Loops: From Physiological to Pathological Roles. *Cell* **179**: 604–618.

García-Muse T, Aguilera A. 2019b. R Loops: From Physiological to Pathological Roles. *Cell* **179**: 604–618.

Gasser SM, Laroche T, Falquet J, Boy de la Tour E, Laemmli UK. 1986. Metaphase chromosome structure: Involvement of topoisomerase II. *J Mol Biol* **188**: 613–629.

Gerton JL, DeRisi J, Shroff R, Lichten M, Brown PO, Petes TD. 2000. Global mapping of meiotic

recombination hotspots and coldspots in the yeast Saccharomyces cerevisiae. *Proc Natl Acad Sci U S A* **97**: 11383–90. http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=17209&tool=pmcentrez&rendertype=abstract (Accessed July 21, 2014).

Ginno PA, Lott PL, Christensen HC, Korf I, Chédin F. 2012. R-Loop Formation Is a Distinctive Characteristic of Unmethylated Human CpG Island Promoters. *Mol Cell* **45**.

Gorthi A, Romero JC, Loranc E, Cao L, Lawrence LA, Goodale E, Iniguez AB, Bernard X, Masamsetti VP, Roston S, et al. 2018. EWS-FLI1 increases transcription to cause R-Loops and block BRCA1 repair in Ewing sarcoma. *Nature* **555**: 387–391. http://dx.doi.org/10.1038/nature25748.

Govin J, Dorsey J, Gaucher J, Rousseaux S, Khochbin S, Berger SL. 2010. Systematic screen reveals new functional dynamics of histones H3 and H4 during gametogenesis. *Genes Dev* **24**: 1772–1786.

Grob S, Schmid MW, Grossniklaus U. 2014. Hi-C Analysis in Arabidopsis Identifies the KNOT, a Structure with Similarities to the flamenco Locus of Drosophila. *Mol Cell* **55**: 678–93. http://linkinghub.elsevier.com/retrieve/pii/S1097276514006029 (Accessed August 15, 2014).

Grob S, Schmid MW, Luedtke NW, Wicker T, Grossniklaus U. 2013. Characterization of chromosomal architecture in Arabidopsis by chromosome conformation capture. *Genome Biol* **14**: 1–19.

Groh M, Albulescu LO, Cristini A, Gromak N. 2017. Senataxin: Genome guardian at the interface of transcription and neurodegeneration. *J Mol Biol* **429**: 3181–3195. http://dx.doi.org/10.1016/j.jmb.2016.10.021.

Grunseich C, Wang IX, Watts JA, Crain B, Fischbeck KH, Cheung VG. 2018. Senataxin Mutation Reveals How R-Loops Promote Transcription by Blocking DNA Methylation at Gene Promoters. *Mol Cell* **69**: 426-437.e7. https://doi.org/10.1016/j.molcel.2017.12.030.

Guccione E, Bassi C, Casadio F, Martinato F, Cesaroni M, Schuchlautz H, Lüscher B, Amati B. 2007. Methylation of histone H3R2 by PRMT6 and H3K4 by an MLL complex are mutually exclusive. *Nature* **449**: 933–937.

Guo Y, Xu Q, Canzio D, Krainer AR, Maniatis T, Guo Y, Xu Q, Canzio D, Shou J, Li J, et al. 2015. CRISPR Inversion of CTCF Sites Alters Genome Topology and Enhancer / Promoter Function. *Cell* **162**: 900–910. http://dx.doi.org/10.1016/j.cell.2015.07.038.

Gurard-levin ZA, Ugalde V, Li J, Forn I, Imhof A, Loyola A. 2015. Methylation of histone H3 lysine 9 occurs during translation. *Nucleic Acids Res* **43**: 9097–9106.

Haarhuis JHI, van der Weide RH, Blomen VA, Yáñez-Cuna JO, Amendola M, van Ruiten MS, Krijger PHL, Teunissen H, Medema RH, van Steensel B, et al. 2017. The Cohesin Release Factor WAPL Restricts Chromatin Loop Extension. *Cell* **169**: 693-707.e14.

Halász L, Karányi Z, Boros-oláh B, Kuik-rózsa T, Sipos É, Nagy É, Mosolygó-l Á, Mázló A, Rajnavölgyi É, Halmos G, et al. 2017. RNA-DNA hybrid (R-loop) immunoprecipitation mapping: an analytical workflow to evaluate inherent biases. *Genome Res* **27**: 1063–1073.

Hardcastle TJ, Müller SY, Baulcombe DC. 2018. Towards annotating the plant epigenome: The Arabidopsis thaliana small RNA locus map. *Sci Rep* **8**: 1–15.

Hatchi E, Goehring L, Landini S, Skourti-Stathaki K, DeConti DK, Abderazzaq FO, Banerjee P, Demers

TM, Wang YE, Quackenbush J, et al. 2021. BRCA1 and RNAi factors promote repair mediated by small RNAs and PALB2–RAD52. *Nature* **591**: 665–670. http://dx.doi.org/10.1038/s41586-020-03150-2.

Hatchi E, Skourti-Stathaki K, Ventz S, Pinello L, Yen A, Kamieniarz-Gdula K, Dimitrov S, Pathania S, McKinney KM, Eaton ML, et al. 2015. BRCA1 recruitment to transcriptional pause sites is required for R-loop-driven DNA damage repair. *Mol Cell* **57**: 636–647. http://dx.doi.org/10.1016/j.molcel.2015.01.011.

Hayashi K, Yoshida K, Matsui Y. 2005. A histone H3 methyltransferase controls epigenetic events required for meiotic prophase. *Nature* **438**: 374–378.

Hegedüs É, Kókai E, Nánási P, Imre L, Halász L, Jossé R, Antunovics Z, Webb MR, Hage A El, Pommier Y, et al. 2018. Endogenous single-strand DNA breaks at RNA polymerase II promoters in saccharomyces cerevisiae. *Nucleic Acids Res* **46**: 10649–10668.

Hetey S, Boros-Oláh B, Kuik-rózsa T, Li Q, Karányi Z, Szabó Z, Roszik J, Szalóki N, Vámosi G, Tóth K, et al. 2017. Biophysical characterization of histone H3.3 K27 M point mutation. *Biochem Biophys Res Commun* **490**: 868–875.

Hodroj D, Recolin B, Serhal K, Martinez S, Tsanov N, Abou Merhi R, Maiorano D. 2017a. An ATR-dependent function for the Ddx19 RNA helicase in nuclear R-loop metabolism. *EMBO J* **36**: 1182–1198. http://emboj.embopress.org/lookup/doi/10.15252/embj.201695131.

Hodroj D, Serhal K, Maiorano D. 2017b. Ddx19 links mRNA nuclear export with progression of transcription and replication and suppresses genomic instability upon DNA damage in proliferating cells. *Nucleus* **8**: 489–495. https://doi.org/10.1080/19491034.2017.1348448.

Hoencamp C, Dudchenko O, Elbatsh AMO, Brahmachari S, Raaijmakers JA, van Schaik T, Cacciatore ÁS, Contessoto VG, van Heesbeen RGHP, van den Broek B, et al. 2021. 3D genomics across the tree of life reveals condensin II as a determinant of architecture type. *Science (80-)* **372**: 984–989.

Hondele M, Sachdev R, Heinrich S, Wang J, Vallotton P, Fontoura BMA, Weis K. 2019. DEAD-box ATPases are global regulators of phase-separated organelles. *Nature* **573**: 144–148. http://dx.doi.org/10.1038/s41586-019-1502-y.

Hornyák L, Dobos N, Koncz G, Karányi Z, Páll D, Szabó Z, Halmos G, Székvölgyi L. 2018. The Role of Indoleamine-2,3-Dioxygenase in Cancer Development, Diagnostics, and Therapy. *Front Immunol* **9**: 1–8. http://journal.frontiersin.org/article/10.3389/fimmu.2018.00151/full.

Hsieh T-H, Fudenberg G, Goloborodko A, Rando O. 2016. Micro-C XL: assaying chromosome conformation at length scales from the nucleosome to the entire genome. *Nat Methods* **53**: 1689–1699. http://biorxiv.org/lookup/doi/10.1101/071357.

Hu Z, Zhang A, Storz G, Gottesman S, Leppla SH. 2006. An antibody-based microarray assay for small RNA detection. *Nucleic Acids Res* **34**: 1–7.

Huertas P, Aguilera A. 2003. Cotranscriptionally formed DNA:RNA hybrids mediate transcription elongation impairment and transcription-associated recombination. *Mol Cell* **12**: 711–721.

Iberg AN, Espejo A, Cheng D, Kim D, Michaud-Levesque J, Richard S, Bedford MT. 2008. Arginine methylation of the histone H3 tail impedes effector binding. *J Biol Chem* **283**: 3006–3010.

Jeon J, McGinty RK, Muir TW, Kim JA, Kim J. 2018. Crosstalk among Set1 complex subunits involved

in H2B ubiquitylation-dependent H3K4 methylation. *Nucleic Acids Res* **46**: 11129–11143.

Jerkovic´ I, Cavalli G. 2021. Understanding 3D genome organization by multidisciplinary methods. *Nat Rev Mol Cell Biol* **22**: 511–528. http://dx.doi.org/10.1038/s41580-021-00362-w.

Jin X, Fudenberg G, Pollard KS. 2021. Genome-wide variability in recombination activity is associated with meiotic chromatin organization. *Genome Res* **31**: 1561–1572.

Kannan A, Jiang ÃX, He L, Ahmad S, Gangwani L. 2019. ZPR1 prevents R-loop accumulation, upregulates SMN2 expression and rescues spinal muscular atrophy. *Brain* pii: awz373.

Kaplan T, Liu CL, Erkmann JA, Holik J, Grunstein M, Kaufman PD, Friedman N, Rando OJ. 2008. Cell cycle- and chaperone-mediated regulation of H3K56ac incorporation in yeast. *PLoS Genet* **4**.

Karányi Z, Halász L, Acquaviva L, Jónás D, Hetey S, Boros-Oláh B, Peng F, Chen D, Klein F, Géli V, et al. 2018. Nuclear dynamics of the Set1C subunit Spp1 prepares meiotic recombination sites for break formation. *J Cell Biol* **217**: 3398–3415. http://jcb.rupress.org/content/early/2018/07/20/jcb.201712122 (Accessed July 23, 2018).

Karányi Z, Hornyák L, Székvölgyi L. 2020. Histone H3 Lysine 56 Acetylation Is Required for Formation of Normal Levels of Meiotic DNA Breaks in S. cerevisiae. *Front Cell Dev Biol* **7**: 364.

Kasahara M, Clikeman JA, Bates DB, Kogoma T. 2000. RecA protein-dependent R-loop formation in vitro. *Genes Dev* **14**: 360–365.

Kim J, Guermah M, Mcginty RK, Lee J, Tang Z, Milne TA, Shilatifard A, Muir TW, Roeder RG. 2009. RAD6-Mediated Transcription-Coupled H2B Ubiquitylation Directly Stimulates H3K4 Methylation in Human Cells. *Cell* **137**: 459–471. http://dx.doi.org/10.1016/j.cell.2009.02.027.

Kimura H, Cook PR. 2001. Kinetics of core histones in living human cells: little exchange of H3 and H4 and some rapid exchange of H2B. *J Cell Biol* **153**: 1341–53. http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2150718&tool=pmcentrez&rendertype=abstract.

Klein HL. 2008. The consequences of Rad51 overexpression for normal and tumor cells. *DNA Repair (Amst)* **7**: 686–693.

Kotsantis P, Silva LM, Irmscher S, Jones RM, Folkes L, Gromak N, Petermann E. 2016. Increased global transcription activity as a mechanism of replication stress in cancer. *Nat Commun* **7**: 1–13. http://dx.doi.org/10.1038/ncomms13087.

Kralemann LEM, Liu S, Trejo-Arellano MS, Muñoz-Viana R, Köhler C, Hennig L. 2020. Removal of H2Aub1 by ubiquitin-specific proteases 12 and 13 is required for stable Polycomb-mediated gene repression in Arabidopsis. *Genome Biol* **21**: 1–19. https://genomebiology.biomedcentral.com/articles/10.1186/s13059-020-02062-8 (Accessed July 23, 2021).

Krebs JE, Goldstein ES, Kilpatrick ST. 2018. *Lewin's GENES XII*. Amazon.

Lam I, Keeney S. 2015. Mechanism and regulation of meiotic recombination initiation. *Cold Spring Harb Perspect Biol* **7**.

Landt SG, Marinov GK, Kundaje A, Kheradpour P, Pauli F, Batzoglou S, Bernstein BE, Bickel P, Brown JB, Cayting P, et al. 2012. ChIP-seq guidelines and practices of the ENCODE and modENCODE consortia. *Genome Res* **22**: 1813–1831.

Lazar-Stefanita L, Scolari VF, Mercy G, Muller H, Guérin TM, Thierry A, Mozziconacci J, Koszul R. 2017. Cohesins and condensins orchestrate the 4D dynamics of yeast chromosomes during the cell cycle. *EMBO J* **36**: e201797342. http://emboj.embopress.org/lookup/doi/10.15252/embj.201797342.

Li X, Manley JL. 2005. Inactivation of the SR protein splicing factor ASF/SF2 results in genomic instability. *Cell* **122**: 365–378.

Liang Z, Liang F, Teng Y, Chen X, Liu J, Longerich S, Rao T, Green AM, Collins NB, Xiong Y, et al. 2019. Binding of FANCI-FANCD2 Complex to RNA and R-Loops Stimulates Robust FANCD2 Monoubiquitination. *Cell Rep* **26**: 564-572.e5. https://doi.org/10.1016/j.celrep.2018.12.084.

Lichten M. 2015. Putting the breaks on meiosis. *Science (80-)* **350**: 913.

Lickwar CR, Mueller F, Hanlon SE, McNally JG, Lieb JD. 2012. Genome-wide protein–DNA binding dynamics suggest a molecular clutch for transcription factor function. *Nature* **484**: 251–255. http://dx.doi.org/10.1038/nature10985.

Lickwar CR, Mueller F, Lieb JD. 2013. Genome-wide measurement of protein-DNA binding dynamics using competition ChIP. *Nat Protoc* **8**: 1337–1353. http://eutils.ncbi.nlm.nih.gov/entrez/eutils/elink.fcgi?dbfrom=pubmed&id=23764940&retmode=ref&cmd=prlinks%5Cnpapers2://publication/doi/10.1038/nprot.2013.077.

Lieberman-aiden E, Berkum NL Van, Williams L, Imakaev M, Ragoczy T, Telling A, Amit I, Lajoie BR, Sabo PJ, Dorschner MO, et al. 2009. Comprehensive Mapping of Long-Range Interactions Reveals Folding Principles of the Human Genome. *Science (80-)* **33292**: 289–293.

Lim YW, Sanz LA, Xu X, Hartono SR, Chédin F. 2015. Genome-wide DNA hypomethylation and RNA:DNA hybrid accumulation in Aicardi-Goutières syndrome. *Elife* **4**: e08007. http://elifesciences.org/content/early/2015/07/16/eLife.08007.abstract.

Lindroth AM, Cao X, Jackson JP, Zilberman D, McCallum CM, Henikoff S, Jacobsen SE. 2001. Requirement of CHROMOMETHYLASE3 for maintenance of CpXpG methylation. *Science (80-)* **292**: 2077–2080.

Lockhart A, Pires VB, Bento F, Yakoub G, Ulrich HD, Luke B, Lockhart A, Pires VB, Bento F, Kellner V, et al. 2019. RNase H1 and H2 Are Differentially Regulated to Process RNA-DNA Hybrids Article RNase H1 and H2 Are Differentially Regulated to Process RNA-DNA Hybrids. *Cell Rep* **29**: 2890-2900.e5. https://doi.org/10.1016/j.celrep.2019.10.108.

Luger K, Rechsteiner TJ, Richmond TJ. Expression and Purification of Recombinant Histones and Nucleosome Reconstitution. **119**.

Maleki S, Neale MJ, Arora C, Henderson KA, Keeney S. 2007. Interactions between Mei4, Rec114, and other proteins required for meiotic DNA double-strand break formation in Saccharomyces cerevisiae. *Chromosoma* **116**: 471–486.

Mancera E, Bourgon R, Brozzi A, Huber W, Steinmetz LM. 2008. High-resolution mapping of meiotic crossovers and non-crossovers in yeast. *Nature* **454**: 479–85. http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2780006&tool=pmcentrez&rendertype=abstract (Accessed July 14, 2014).

Manning GS. 2006. The persistence length of DNA is reached from the persistence length of its null isomer through an internal electrostatic stretching force. *Biophys J* **91**: 3607–3616.

http://dx.doi.org/10.1529/biophysj.106.089029.

Manzo SG, Hartono SR, Sanz LA, De S, Cossarizza A, Capranico G, Chedin F. 2018. DNA Topoisomerase I differentially modulates R-loops across the human genome. *Genome Biol* 1–18.

Marinello J, Bertoncini S, Aloisi I, Cristini A, Tagliazucchi GM, Forcato M, Sordet O, Capranico G. 2016. Dynamic effects of topoisomerase i inhibition on R-loops and short transcripts at active promoters. *PLoS One* **11**: 1–18.

Matzke MA, Mosher RA. 2014. RNA-directed DNA methylation: An epigenetic pathway of increasing complexity. *Nat Rev Genet* **15**: 394–408.

McCulloch SD, Kunkel TA. 2008. The fidelity of DNA synthesis by eukaryotic replicative and translesion synthesis polymerases. *Cell Res* **18**: 148–161.

Messiaen S, Guiard J, Aigueperse C, Fliniaux I, Tourpin S, Barroca V, Allemand I, Fouchet P, Livera G, Vernet M. 2016. Loss of the histone chaperone ASF1B reduces female reproductive capacity in mice. *Reproduction* **151**: 477–489.

Meyer A, Schloissnig S, Franchini P, Du K, Woltering JM, Irisarri I, Wong WY, Nowoshilow S, Kneitz S, Kawaguchi A, et al. 2021. Giant lungfish genome elucidates the conquest of land by vertebrates. *Nature* **590**: 284–289. http://dx.doi.org/10.1038/s41586-021-03198-8.

Migliori V, Müller J, Phalke S, Low D, Bezzi M, Mok WC, Sahu SK, Gunaratne J, Capasso P, Bassi C, et al. 2012. Symmetric dimethylation of H3R2 is a newly identified histone mark that supports euchromatin maintenance. *Nat Struct Mol Biol* **19**: 136–44. http://www.ncbi.nlm.nih.gov/pubmed/22231400 (Accessed March 7, 2013).

Mihola O, Trachtulec Z, Vlcek C, Schimenti JC, Forejt J. 2009. A Mouse Speciation Gene Encodes a Meiotic Histone H3 Methyltransferase. *Science (80-)* **323**: 373–5. http://www.ncbi.nlm.nih.gov/pubmed/19074312.

Miller T, Krogan NJ, Dover J, Erdjument-Bromage H, Tempst P, Johnston M, Greenblatt JF, Shilatifard A. 2001. COMPASS: a complex of proteins associated with a trithorax-related SET domain protein. *Proc Natl Acad Sci U S A* **98**: 12902–7.

Mischo HE, Gómez-González B, Grzechnik P, Rondón AG, Wei W, Steinmetz L, Aguilera A, Proudfoot NJ. 2011. Yeast Sen1 helicase protects the genome from transcription-associated instability. *Mol Cell* **41**: 21–32. http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3314950&tool=pmcentrez&rendertype=abstract (Accessed March 10, 2013).

Mizuguchi T, Fudenberg G, Mehta S, Belton J-M, Taneja N, Folco HD, FitzGerald P, Dekker J, Mirny L, Barrowman J, et al. 2014. Cohesin-dependent globules and heterochromatin shape 3D genome architecture in S. pombe. *Nature* **516**: 432–435. http://www.nature.com/doifinder/10.1038/nature13833 (Accessed October 13, 2014).

Monk BJ, Lorusso D, Italiano A, Kaye SB, Aracil M, Tanović A, D'Incalci M. 2016. Trabectedin as a chemotherapy option for patients with BRCA deficiency. *Cancer Treat Rev* **50**: 175–182. http://dx.doi.org/10.1016/j.ctrv.2016.09.009.

Montavon T, Shukeir N, Erikson G, Engist B, Onishi-seebacher M, Ryan D, Musa Y, Mittler G, Meyer AG, Genoud C, et al. 2021. Complete loss of H3K9 methylation dissolves mouse

heterochromatin organization. *Nat Commun* **12**: 4359. http://dx.doi.org/10.1038/s41467-021-24532-8.

Muller H, Scolari VF, Agier N, Piazza A, Thierry A, Mercy G, Descorps-Declere S, Lazar-Stefanita L, Espeli O, Llorente B, et al. 2018. Characterizing meiotic chromosomes' structure and pairing using a designer sequence optimized for Hi-C. *Mol Syst Biol* **14**: e8293. http://www.ncbi.nlm.nih.gov/pubmed/30012718.

Murakami H, Keeney S. 2008. Regulating the formation of DNA double-strand breaks in meiosis. *Genes Dev* **22**: 286–292.

Murakami H, Lam I, Huang PC, Song J, van Overbeek M, Keeney S. 2020. Multilayered mechanisms ensure that short chromosomes recombine in meiosis. *Nature* **582**: 124–128. http://dx.doi.org/10.1038/s41586-020-2248-2.

Murton BL, Chin WL, Ponting CP, Itzhaki LS. 2010. Characterising the Binding Specificities of the Subunits Associated with the KMT2/Set1 Histone Lysine Methyltransferase. *J Mol Biol* **398**: 481–488. http://dx.doi.org/10.1016/j.jmb.2010.03.036.

Myers S, Bowden R, Tumian A, Bontrop RE, Freeman C, MacFie TS, McVean G, Donnelly P. 2010. Drive against hotspot motifs in primates implicates the PRDM9 gene in meiotic recombination. *Science* **327**: 876–9. http://science.sciencemag.org/content/327/5967/876.abstract.

Nadel J, Athanasiadou R, Lemetre C, Wijetunga NA, Ó Broin P, Sato H, Zhang Z, Jeddeloh J, Montagna C, Golden A, et al. 2015. RNA:DNA hybrids in the human genome have distinctive nucleotide characteristics, chromatin composition, and transcriptional relationships. *Epigenetics and Chromatin* **8**.

Nagano T, Lubling Y, Yaffe E, Wingett SW, Dean W, Tanay A, Fraser P. 2015. Single-cell Hi-C for genome-wide detection of chromatin interactions that occur simultaneously in a single cell. *Nat Protoc* **10**: 1986–2003. http://www.nature.com/doifinder/10.1038/nprot.2015.127.

Nagy G, Czipa E, Steiner L, Nagy T, Pongor S, Nagy L, Barta E. 2016. Motif oriented high-resolution analysis of ChIP-seq data reveals the topological order of CTCF and cohesin proteins on DNA. *BMC Genomics* **17**: 1–9. http://dx.doi.org/10.1186/s12864-016-2940-7.

Nakanishi S, Sanderson BW, Delventhal KM, Bradford WD, Staehling-Hampton K, Shilatifard A. 2008. A comprehensive library of histone mutants identifies nucleosomal residues required for H3K4 methylation. *Nat Struct Mol Biol* **15**: 881–8. http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2562305&tool=pmcentrez&rendertype=abstract (Accessed November 27, 2013).

Naughton C, Avlonitis N, Corless S, Prendergast JG, Mati IK, Eijk PP, Cockroft SL, Bradley M, Ylstra B, Gilbert N. 2013. Transcription forms and remodels supercoiling domains unfolding large-scale chromatin structures. *Nat Struct Mol Biol* **20**: 387–395.

Neale MJ, Pan J, Keeney S. 2005. Endonucleolytic processing of covalent protein-linked DNA double-strand breaks. *Nature* **436**: 1053–7. http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=1262668&tool=pmcentrez&rendertype=abstract (Accessed July 21, 2014).

Ngan CY, Wong CH, Tjong H, Wang W, Goldfeder RL, Choi C, He H, Gong L, Lin J, Urban B, et al. 2020. Chromatin interaction analyses elucidate the roles of PRC2-bound silencers in mouse development. *Nat Genet* **52**: 264–272. http://dx.doi.org/10.1038/s41588-020-0581-x.

Ngo GHP, Grimstead JW, Baird DM. 2021. UPF1 promotes the formation of R loops to stimulate DNA double-strand break repair. *Nat Commun* **12**: 1–15. http://dx.doi.org/10.1038/s41467-021-24201-w.

Nichols MH, Corces VG. 2015. A CTCF Code for 3D Genome Architecture. *Cell* **162**: 703–705. http://dx.doi.org/10.1016/j.cell.2015.07.053.

Niehrs C, Luke B. 2020. Regulatory R-loops as facilitators of gene expression and genome stability. *Nat Rev Mol Cell Biol* **21**: 167–178. http://dx.doi.org/10.1038/s41580-019-0206-3.

Noristani HN, Sabourin JC, Gerber YN, Teigell M, Sommacal A, dM Vivanco M, Weber M, Perrin FE. 2015. Brca1 is expressed in human microglia and is dysregulated in human and animal model of ALS. *Mol Neurodegener* **10**: 34. http://www.molecularneurodegeneration.com/content/10/1/34.

Ohle C, Tesorero R, Schermann G, Dobrev N, Sinning I, Fischer T. 2016. Transient RNA-DNA Hybrids are Required for Efficient Double-Strand Break Repair. *Cell* **167**: 1001–1013. https://www.ncbi.nlm.nih.gov/pubmed/?term=10.1016%2Fj.cell.2016.10.001.

Ohta K, Shibata T, Nicolas A. 1994. Changes in chromatin structure at recombination initiation sites during yeast meiosis. *EMBO J* **13**: 5754–5763.

Olins AL, Olins DE. 1974. Spheroid chromatin units (v bodies). *Science (80-)* **183**: 330–2. http://www.ncbi.nlm.nih.gov/pubmed/4128918.

Pan J, Sasaki M, Kniewel R, Murakami H, Blitzblau HG, Tischfield SE, Zhu X, Neale MJ, Jasin M, Socci ND, et al. 2011. A hierarchical combination of factors shapes the genome-wide topography of yeast meiotic recombination initiation. *Cell* **144**: 719–31. http://dx.doi.org/10.1016/j.cell.2011.02.009 (Accessed March 12, 2013).

Panizza S, Mendoza M a, Berlinger M, Huang L, Nicolas A, Shirahige K, Klein F. 2011. Spo11-accessory proteins link double-strand break sites to the chromosome axis in early meiotic recombination. *Cell* **146**: 372–83. http://www.ncbi.nlm.nih.gov/pubmed/21816273 (Accessed March 8, 2013).

Parpanov ED, Petkov PM, Paigen K. 2010. Prdm9 Controls Activation of Mammalian Recombination Hotspots. *Science (80-)* **327**: 2010.

Parvanov ED, Petkov PM, Paigen K. 2010. Prdm9 controls activation of mammalian recombination hotspots. *Science (80-)* **327**: 835.

Perego MGL, Taiana M, Bresolin N, Comi GP, Corti S. 2018. R-Loops in Motor Neuron Diseases. *Mol Neurobiol* 1–11. http://link.springer.com/10.1007/s12035-018-1246-y.

Pommier Y, Sun Y, Huang SYN, Nitiss JL. 2016. Roles of eukaryotic topoisomerases in transcription, replication and genomic stability. *Nat Rev Mol Cell Biol* **17**: 703–721.

Radzioch D. 2013. *Chromatin Remodelling*. ed. Danuta Radzioch.

Rao SSP, Huang SC, Glenn St Hilaire B, Engreitz JM, Perez EM, Kieffer-Kwon KR, Sanborn AL, Johnstone SE, Bascom GD, Bochkov ID, et al. 2017. Cohesin Loss Eliminates All Loop Domains. *Cell* **171**: 305-320.e24. https://doi.org/10.1016/j.cell.2017.09.026.

Rao SSP, Huntley MH, Durand NC, Stamenova EK, Bochkov ID, Robinson JT, Sanborn AL, Machol I, Omer AD, Lander ES, et al. 2014. A 3D Map of the Human Genome at Kilobase Resolution

Reveals Principles of Chromatin Looping. *Cell* **159**: 1665–1680. http://linkinghub.elsevier.com/retrieve/pii/S0092867414014974 (Accessed December 11, 2014).

Razin S V., Iarovaia O V., Vassetzky YS. 2014. A requiem to the nuclear matrix: From a controversial concept to 3D organization of the nucleus. *Chromosoma* **123**: 217–224.

Recht J, Tsubota T, Tanny JC, Diaz RL, Berger JM, Zhang X, Garcia BA, Shabanowitz J, Burlingame AL, Hunt DF, et al. 2006. Histone chaperone Asf1 is required for histone H3 lysine 56 acetylation, a modification associated with S phase in mitosis and meiosis. *Proc Natl Acad Sci* **103**: 6988–6993. http://www.pnas.org/cgi/doi/10.1073/pnas.0601676103.

Reddy K, Schmidt MHM, Geist JM, Thakkar NP, Panigrahi B, Wang Y, Pearson CE. 2014. Processing of double-R-loops in (CAG) · (CTG) and C9orf72 (GGGGCC) · (GGCCCC) repeats causes instability. *Nucleic Acids Res* **42**: 10473–10487.

Renaudin X, Lee M, Shehata M, Surmann E-M, Venkitaraman AR. 2021. BRCA2 deficiency reveals that oxidative stress impairs RNaseH1 function to cripple mitochondrial DNA maintenance. *Cell Rep* **36**: 109478. https://doi.org/10.1016/j.celrep.2021.109478.

Ribeiro de Almeida C, Dhir S, Dhir A, Moghaddam AE, Sattentau Q, Meinhart A, Proudfoot NJ. 2018. RNA Helicase DDX1 Converts RNA G-Quadruplex Structures into R-Loops to Promote IgH Class Switch Recombination. *Mol Cell* **70**: 650-662.e8. https://doi.org/10.1016/j.molcel.2018.04.001.

Rivero-Hinojosa S, Pugacheva EM, Kang S, Méndez-Catalá CF, Kovalchuk AL, Strunnikov A V., Loukinov D, Lee JT, Lobanenkov V V. 2021. The combined action of CTCF and its testis-specific paralog BORIS is essential for spermatogenesis. *Nat Commun* **12**: 1–17. http://dx.doi.org/10.1038/s41467-021-24140-6.

Roberts RW, Crothers DM. 1992. Stability and Properties of Double and Triple Helices: Dramatic Effects of RNA or DNA Backbone Composition. *Science (80-)* **258**: 1463–1466.

Roszik J, Fenyőfalvi G, Halász L, Karányi Z, Székvölgyi L. 2017. In Silico Restriction Enzyme Digests to Minimize Mapping Bias in Genomic Sequencing. *Mol Ther - Methods Clin Dev* **6**: 66–67.

Rousova D, Nivsarkar V, Altmannova V, Funk SK, Raina VB, Liedtke D, Müller F, Reichle H, Vader G, Weir JR. 2021. Novel mechanistic insights into the role of Mer2 as the keystone of meiotic DNA break formation. *Elife* **10**: e72330.

Roy D, Zhang Z, Lu Z, Hsieh C-L, Lieber MR. 2010. Competition between the RNA transcript and the nontemplate DNA strand during R-loop formation in vitro: a nick can serve as a strong R-loop initiation site. *Mol Cell Biol* **30**: 146–59. http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2798282&tool=pmcentrez&rendertype=abstract (Accessed March 21, 2013).

Rufiange A, Jacques PE, Bhat W, Robert F, Nourani A. 2007. Genome-wide replication-independent histone H3 exchange occurs predominantly at promoters and implicates H3 K56 acetylation and Asf1. *MolCell* **27**: 393–405.

Sakasai R, Isono M, Wakasugi M, Hashimoto M, Sunatani Y, Matsui T, Shibata A, Matsunaga T, Iwabuchi K. 2017. Aquarius is required for proper CtIP expression and homologous recombination repair. *Sci Rep* **7**: 13808. http://www.ncbi.nlm.nih.gov/pubmed/29061988 (Accessed February 5, 2019).

Santos-pereira JM, Aguilera A. 2015. R loops: new modulators of genome dynamics and function. *Nat Rev Genet* **16**: 583–597.

Santos-Rosa H, Kirmizis A, Nelson C, Bartke T, Saksouk N, Cote J, Kouzarides T. 2009. Histone H3 tail clipping regulates gene expression. *NatStructMolBiol* **16**: 17–22.

Sanz LA, Chédin F. 2019. High-resolution, strand-specific R-loop mapping via S9.6-based DNA–RNA immunoprecipitation and high-throughput sequencing. *Nat Protoc* **14**: 1734–1755.

Sanz LA, Hartono SR, Lim YW, Ginno PA, Sanz LA, Hartono SR, Lim YW, Steyaert S, Rajpurkar A, Ginno PA, et al. 2016. Prevalent , Dynamic , and Conserved R-Loop Structures Associate with Specific Epigenomic Signatures in Mammals. *Mol Cell* **63**: 167–178.

Schalbetter SA, Fudenberg G, Baxter J, Pollard KS, Neale MJ. 2019. Principles of meiotic chromosome assembly revealed in S. cerevisiae. *Nat Commun* **10**: 4795. http://dx.doi.org/10.1038/s41467-019-12629-0.

Schoenfelder S, Furlan-Magaril M, Mifsud B, Tavares-Cadete F, Sugar R, Javierre BM, Nagano T, Katsman Y, Sakthidevi M, Wingett SW, et al. 2015. The pluripotent regulatory circuitry connecting promoters to their long-range interacting elements. *Genome Res* **25**: 582–597.

Schwab RA, Nieminuszczy J, Shah F, Langton J, Lopez Martinez D, Liang CC, Cohn MA, Gibbons RJ, Deans AJ, Niedzwiedz W. 2015. The Fanconi Anemia Pathway Maintains Genome Stability by Coordinating Replication and Transcription. *Mol Cell* **60**: 351–361. http://dx.doi.org/10.1016/j.molcel.2015.09.012.

Schwacha A, Kleckner N. 1997. Interhomolog Bias during Meiotic Recombination: Meiotic Functions Promote a Highly Differentiated Interhomolog-Only Pathway. *Cell* **90**: 1123–1135.

Sequeira-Mendes J, Aragüez I, Peiró R, Mendez-Giraldez R, Zhang X, Jacobsen SE, Bastolla U, Gutierrez C. 2014. The Functional Topography of the Arabidopsis Genome Is Organized in a Reduced Number of Linear Motifs of Chromatin States. *Plant Cell* **26**: 2351–2366. http://www.plantcell.org.gate1.inist.fr/content/26/6/2351.long.

Sessions SK. 2013. *Genome Size*. Academic Press.

Shaw NN, Xi H, Arya DP. 2008. Molecular recognition of a DNA:RNA hybrid: Sub-nanomolar binding by a neomycin-methidium conjugate. *Bioorganic Med Chem Lett* **18**: 4142–4145.

Simon M, North JA, Shimko JC, Forties RA, Ferdinand MB, Manohar M, Zhang M, Fishel R, Ottesen JJ, Poirier MG. 2011. Histone fold modifications control nucleosome unwrapping and disassembly. *Proc Natl Acad Sci U S A* **108**: 12711–12716. http://www.ncbi.nlm.nih.gov/pubmed/21768347%5Cnhttp://www.ncbi.nlm.nih.gov/pmc/articles/PMC3150920/pdf/pnas.1106264108.pdf.

Singhal S, Leffler EM, Sannareddy K, Turner I, Venn O, Hooper DM, Strand AI, Li Q, Raney B, Balakrishnan CN, et al. 2015. Stable recombination hotspots in Birds. *Science (80-)* **350**.

Skourti-Stathaki K, Proudfoot NJ. 2014. A double-edged sword: R loops as threats to genome integrity and powerful regulators of gene expression. *Genes Dev* **28**: 1384–1396.

Skourti-Stathaki K, Proudfoot NJ, Gromak N. 2011. Human senataxin resolves RNA/DNA hybrids formed at transcriptional pause sites to promote Xrn2-dependent termination. *Mol Cell* **42**: 794–805. http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3145960&tool=pmcentrez&ren

dertype=abstract (Accessed March 10, 2013).

Skourti-Stathaki K, Triglia ET, Warburton M, Voigt P, Bird A, Pombo A. 2019. R-Loops Enhance Polycomb Repression at a Subset of Developmental Regulator Genes. *Mol Cell* **73**: 930–945. https://doi.org/10.1016/j.molcel.2018.12.016.

Smallwood A, Estève P, Pradhan S, Carey M. 2007. Functional cooperation between HP1 and DNMT1 mediates gene silencing. *Genes Dev* **21**: 1169–1178.

Smith JM, Eörs Szathmáry. 1995. *The Major Transitions in Evolution*.

Sofueva S, Yaffe E, Chan W-C, Georgopoulou D, Vietri Rudan M, Mira-Bontenbal H, Pollard SM, Schroth GP, Tanay A, Hadjur S. 2013. Cohesin-mediated interactions organize chromosomal domain architecture. *EMBO J* **32**: 3119–29. http://www.ncbi.nlm.nih.gov/pubmed/24185899.

Sollier J, Cimprich K a. 2015. Breaking bad: R-loops and genome integrity. *Trends Cell Biol* 1–9. http://linkinghub.elsevier.com/retrieve/pii/S0962892415000938.

Sollier J, Lin W, Soustelle C, Suhre K, Nicolas A, Géli V, de La Roche Saint-André C. 2004. Set1 is required for meiotic S-phase onset, double-strand break formation and middle gene expression. *EMBO J* **23**: 1957–67. http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=404324&tool=pmcentrez&rendertype=abstract (Accessed November 27, 2013).

Sollier J, Stork CT, García-Rubio ML, Paulsen RD, Aguilera A, Cimprich KA. 2014. Transcription-Coupled Nucleotide Excision Repair Factors Promote R-Loop-Induced Genome Instability. *Mol Cell* **56**: 777–785. http://linkinghub.elsevier.com/retrieve/pii/S1097276514008302.

Song C, Hotz-Wagenblatt A, Voit R, Grummt I. 2017. SIRT7 and the DEAD-box helicase DDX21 cooperate to resolve genomic R loops and safeguard genome stability. *Genes Dev* **31**: 1370–1381.

Sorrells S, Nik S, Casey M, Cameron RC, Truong H, Toruno C, Gulfo M, Lowe A, Jette C, Stewart RA, et al. 2018. Spliceosomal components protect embryonic neurons from R-loop-mediated DNA damage and apoptosis. *Dis Model Mech* **11**: pii: dmm031583.

Stroud H, Do T, Du J, Zhong X, Feng S, Johnson L, Patel DJ, Jacobsen SE. 2014. Non-CG methylation patterns shape the epigenetic landscape in Arabidopsis. *Nat Struct Mol Biol* **21**: 64–72.

Sue Mei Tan-Wong, Proudfoot NJ. 2013. Rad51, friend or foe? *Elife* 2:e00914.

Sugawara N, Wang X, Haber JE. 2003. In vivo roles of Rad52, Rad54, and Rad55 proteins in Rad51-mediated recombination. *Mol Cell* **12**: 209–219.

Sun Q, Csorba T, Skourti-Stathaki K, Proudfoot NJ, Dean C. 2013. R-loop stabilization represses antisense transcription at the Arabidopsis FLC locus. *Science* **340**: 619–21. http://www.ncbi.nlm.nih.gov/pubmed/23641115 (Accessed November 17, 2013).

Sureka R, Mishra R. 2021. Identification of Evolutionarily Conserved Nuclear Matrix Proteins and Their Prokaryotic Origins. *J Proteome Res* **20**: 518–530.

Symmons O, Uslu VV, Tsujimura T, Ruf S, Nassari S, Schwarzer W, Ettwiller L, Spitz F. 2014. Functional and topological characteristics of mammalian regulatory domains. *Genome Res* **24**: 390–400.

Szekvolgyi L, Nicolas A. 2010. From meiosis to postmeiotic events: homologous recombination is obligatory but flexible. *FEBS J* **277**: 571–589.

Székvölgyi L, Nicolas A. 2010. From meiosis to postmeiotic events: homologous recombination is obligatory but flexible. *FEBS J* **277**: 571–89. http://www.ncbi.nlm.nih.gov/pubmed/20015080 (Accessed March 7, 2013).

Székvölgyi L, Ohta K, Nicolas A. 2015a. Initiation of meiotic homologous recombination: Flexibility, impact of histone modifications, and chromatin remodeling. *Cold Spring Harb Perspect Biol* **7**: 1–18.

Székvölgyi L, Ohta K, Nicolas A. 2015b. Initiation of Meiotic Homologous Recombination: Flexibility, Impact of Histone modifications, and chromatin remodelling. *Cold Spring Harb Perspect Biol* **7**: 447–462.

Székvölgyi L, Rákosy Z, Bálint BL, Kókai E, Imre L, Vereb G, Bacsó Z, Goda K, Varga S, Balázs M, et al. 2007. Ribonucleoprotein-masked nicks at 50-kbp intervals in the eukaryotic genomic DNA. *Proc Natl Acad Sci U S A* **104**: 14964–9. http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=1986596&tool=pmcentrez&rendertype=abstract.

Takahashi Y, Westfield GH, Oleskie AN, Trievel RC, Shilatifard A, Skiniotis G. 2011. Structural analysis of the core COMPASS family of histone H3K4 methylases from yeast to human. *Proc Natl Acad Sci U S A* **108**: 20526–31. http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3251153&tool=pmcentrez&rendertype=abstract (Accessed March 26, 2013).

Tang Z, Luo OJ, Li X, Zheng M, Zhu JJ, Szalaj P, Trzaskoma P, Magalska A, Wlodarczyk J, Ruszczycki B, et al. 2015. CTCF-Mediated Human 3D Genome Architecture Reveals Chromatin Topology for Transcription. *Cell* **163**: 1611–1627. http://dx.doi.org/10.1016/j.cell.2015.11.024.

The Cancer Genome Atlas Research Network, Weinstein JN, Eric A Collisson, Gordon B Mills, Kenna R Mills Shaw, Chris Sander, Ellrott K, Joshua M Stuart, Ozenberger BA, Shmulevich I. 2013. The Cancer Genome Atlas Pan-Cancer analysis project. *Nat Genet* **45**: 1113–1120. http://dx.doi.org/10.1038/ng.2764.

Thomas M, White RL, Davis RW. 1976. Hybridization of RNA to double stranded DNA: Formation of R loops. *Proc Natl Acad Sci U S A* **73**: 2294–2298.

Thurman RE, Rynes E, Humbert R, Vierstra J, Maurano MT, Haugen E, Sheffield NC, Stergachis AB, Wang H, Vernot B, et al. 2012. The accessible chromatin landscape of the human genome. *Nature* **489**: 75–82. http://dx.doi.org/10.1038/nature11232.

Torres IO, Fujimori DG. 2015. Functional coupling between writers, erasers and readers of histone and DNA methylation. *CurrOpinStructBiol* 68–75.

Tran PLT, Pohl TJ, Chen C-F, Chan A, Pott S, Zakian VA. 2017. PIF1 family DNA helicases suppress R-loop mediated genome instability at tRNA genes. *Nat Commun* **8**: 15025. http://www.nature.com/doifinder/10.1038/ncomms15025.

Trésaugues L, Dehé P-M, Guérois R, Rodriguez-Gil A, Varlet I, Salah P, Pamblanco M, Luciano P, Quevillon-Cheruel S, Sollier J, et al. 2006. Structural characterization of Set1 RNA recognition motifs and their role in histone H3 lysine 4 methylation. *J Mol Biol* **359**: 1170–81. http://www.ncbi.nlm.nih.gov/pubmed/16787775 (Accessed January 5, 2015).

Tsubota T, Berndsen CE, Erkmann JA, Smith CL, Yang L, Freitas MA, Denu JM, Kaufman PD. 2007. Histone H3-K56 Acetylation Is Catalyzed by Histone Chaperone-Dependent Complexes. *Mol Cell* **25**: 703–712.

Tuduri S, Crabbé L, Conti C, Tourrière H, Holtgreve-Grez H, Jauch A, Pantesco V, De Vos J, Thomas A, Theillet C, et al. 2009. Topoisomerase I suppresses genomic instability by preventing interference between replication and transcription. *Nat Cell Biol* **11**: 1315–24. http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2912930&tool=pmcentrez&rendertype=abstract (Accessed March 15, 2013).

Tumini E, Herrera-Moyano E, San Martín-Alonso M, Barroso S, Galmarini CM, Aguilera A. 2018. The antitumor drugs trabectedin and lurbinectedin induce transcription-dependent replication stress and genome instability. *Mol Cancer Res* molcanres.0575.2018. http://mcr.aacrjournals.org/lookup/doi/10.1158/1541-7786.MCR-18-0575.

Udvardy A. 1999. Dividing the empire: Boundary chromatin elements delimit the territory of enhancers. *EMBO J* **18**: 1–8.

Uehara R, Cerritelli SM, Hasin N, Sakhuja K, London M, Iranzo J, Chon H, Grinberg A, Crouch RJ. 2018. Two RNase H2 Mutants with Differential rNMP Processing Activity Reveal a Threshold of Ribonucleotide Tolerance for Embryonic Development. *Cell Rep* **25**: 1135-1145.e5. http://dx.doi.org/10.1016/j.celrep.2018.10.019.

Vietri Rudan M, Barrington C, Henderson S, Ernst C, Odom DT, Tanay A, Hadjur S. 2015. Comparative Hi-C Reveals that CTCF Underlies Evolution of Chromosomal Domain Architecture. *Cell Rep* **10**: 1297–1309. http://dx.doi.org/10.1016/j.celrep.2015.02.004.

Vitaliano-Prunier A, Menant A, Hobeika M, Géli V, Gwizdek C, Dargemont C. 2008. Ubiquitylation of the COMPASS component Swd2 links H2B ubiquitylation to H3K4 trimethylation. *Nat Cell Biol* **10**: 1365–1371. http://www.nature.com/doifinder/10.1038/ncb1796.

Voichek Y, Mittelman K, Gordon Y, Bar-Ziv R, Lifshitz Smit D, Shenhav R, Barkai N. 2018. Epigenetic Control of Expression Homeostasis during Replication Is Stabilized by the Replication Checkpoint. *Mol Cell* **70**: 1121-1133.e9. https://doi.org/10.1016/j.molcel.2018.05.015.

Wahba L, Costantino L, Tan FJ, Zimmer A, Koshland D. 2016. S1-DRIP-seq identifies high expression and polyA tracts as major contributors to R-loop formation. *Genes Dev* **30**: 1327–38. http://www.ncbi.nlm.nih.gov/pubmed/27298336.

Wahba L, Gore SK, Koshland D. 2013. The homologous recombination machinery modulates the formation of RNA – DNA hybrids and associated chromosome instability. *Elife* 1–20.

Walker C, Herranz-Martin S, Karyka E, Liao C, Lewis K, Elsayed W, Lukashchuk V, Chiang SC, Ray S, Mulcahy PJ, et al. 2017. C9orf72 expansion disrupts ATM-mediated chromosomal break repair. *Nat Neurosci* **20**: 1225–1235.

Wan Y, Zheng X, Chen H, Guo Y, Jiang H, He X, Zhu X, Zheng Y. 2015. Splicing function of mitotic regulators links R-loop-mediated DNA damage to tumor cell killing. *J Cell Biol* **209**: 235–246.

Wang IX, Grunseich C, Fox J, Burdick J, Zhu Z, Ravazian N, Hafner M, Cheung VG. 2018. Human proteins that interact with RNA / DNA hybrids. *Genome Res* 1–10.

Watanabe S, Radman-Livaja M, Rando OJ, Peterson CL. 2013. A histone acetylation switch regulates H2A.Z deposition by the SWR-C remodeling enzyme. *Science (80-)* **340**: 195–199.

Watanabe S, Resch M, Lilyestrom W, Clark N, Hansen JC, Peterson C, Luger K. 2010. Structural characterization of H3K56Q nucleosomes and nucleosomal arrays. *Biochim Biophys Acta* **1799**: 480–486. http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2885283&tool=pmcentrez&rendertype=abstract (Accessed March 7, 2013).

Watanabe Y. 2012. Geometry and force behind kinetochore orientation: lessons from meiosis. *Nat Rev Mol Cell Biol* **13**: 370–82. http://www.ncbi.nlm.nih.gov/pubmed/22588367 (Accessed July 21, 2014).

Weinstein JN, Collisson EA, Mills GB, Shaw KRM, Ozenberger BA, Ellrott K, Sander C, Stuart JM, Chang K, Creighton CJ, et al. 2013. The cancer genome atlas pan-cancer analysis project. *Nat Genet* **45**: 1113–1120. http://dx.doi.org/10.1038/ng.2764.

Weintraub AS, Li CH, Zamudio A V., Sigova AA, Hannett NM, Day DS, Abraham BJ, Cohen MA, Nabet B, Buckley DL, et al. 2017. YY1 Is a Structural Regulator of Enhancer-Promoter Loops. *Cell* **171**: 1573-1588.e28. https://doi.org/10.1016/j.cell.2017.11.008.

Wells JP, White J, Stirling PC. 2019. R Loops and Their Composite Cancer Connections. *Trends in Cancer* 1–13. http://dx.doi.org/10.1016/j.trecan.2019.08.006.

White CL, Suto RK, Luger K. 2001. Structure of the yeast nucleosome core particle reveals fundamental changes in internucleosome interactions. *EMBO J* **20**: 5207–18. http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=125637&tool=pmcentrez&rendertype=abstract.

Whyte WA, Orlando DA, Hnisz D, Abraham BJ, Lin CY, Kagey MH, Rahl PB, Lee TI, Young RA. 2013. Master transcription factors and mediator establish super-enhancers at key cell identity genes. *Cell* **153**: 307–319. http://dx.doi.org/10.1016/j.cell.2013.03.035.

Wit E De, Vos ESM, Holwerda SJB, Valdes-quezada C, Verstegen MJAM, Teunissen H, Splinter E, Wijchers PJ, Krijger PHL, Laat W De. 2015. CTCF Binding Polarity Determines Chromatin Looping. *Mol Cell* **60**: 1–9. http://dx.doi.org/10.1016/j.molcel.2015.09.023.

Witz G, Stasiak A. 2009. DNA supercoiling and its role in DNA decatenation and unknotting. *Nucleic Acids Res* **38**: 2119–2133.

Wu X, Kim TK, Baxter D, Scherler K, Gordon A, Fong O, Etheridge A, Galas DJ, Wang K. 2017. SRNAnalyzer-A flexible and customizable small RNA sequencing data analysis pipeline. *Nucleic Acids Res* **45**: 12140–12151.

Xu L, Kleckner N. 1995. Sequence non-specific double-strand breaks and interhomolog interactions prior to double-strand break formation at a meiotic recombination hot spot in yeast. *EMBO J* **14**: 5115–5128.

Xu W, Xu H, Li K, Fan Y, Liu Y, Yang X, Sun Q. 2017. The R-loop is a common chromatin feature of the Arabidopsis genome. *Nat Plants* **3**: 704–714.

Yang B, Miller A, Kirchmaier AL. 2008. HST3/HST4-dependent Deacetylation of Lysine 56 of Histone H3 in Silent Chromatin ed. K.S. Bloom. *Mol Biol Cell* **19**: 4993–5005. http://www.ncbi.nlm.nih.gov/pubmed/18799617 (Accessed August 6, 2018).

Yang W, Soares J, Greninger P, Edelman EJ, Lightfoot H, Forbes S, Bindal N, Beare D, Smith JA, Thompson IR, et al. 2013. Genomics of Drug Sensitivity in Cancer (GDSC): A resource for

therapeutic biomarker discovery in cancer cells. *Nucleic Acids Res* **41**: 955–961.

Yasuhara T, Kato R, Hagiwara Y, Shiotani B, Yamauchi M, Nakada S, Shibata A, Miyagawa K. 2018. Human Rad52 Promotes XPG-Mediated R-loop Processing to Initiate Transcription-Associated Homologous Recombination Repair. *Cell* **175**: 558-570.e11. https://doi.org/10.1016/j.cell.2018.08.056.

Yu K, Chedin F, Hsieh C-L, Wilson TE, Lieber MR. 2003. R-loops at immunoglobulin class switch regions in the chromosomes of stimulated B cells. *Nat Immunol* **4**: 442–51. http://www.ncbi.nlm.nih.gov/pubmed/12679812 (Accessed November 21, 2013).

Yun E, Chang C, Novoa CA, Aristizabal MJ, Coulombe Y, Segovia R, Chaturvedi R, Shen Y, Keong C, Tam AS, et al. 2017. RECQ-like helicases Sgs1 and BLM regulate R-loop– associated genome instability. *J Cell Biol* **216**: 3991–4005. https://doi.org/10.1083/jcb.201703168.

Zhang H, Zhu JK. 2012. Active DNA demethylation in plants and animals. *Cold Spring Harb Symp Quant Biol* **77**: 161–173.

Zhang L, Wang S, Yin S, Hong S, Kim KP, Kleckner N. 2014. Topoisomerase II mediates meiotic crossover interference. *Nature* **511**: 551–556. http://www.nature.com/doifinder/10.1038/nature13442 (Accessed July 14, 2014).

Zhang X, Chiang H-C, Wang Y, Zhang C, Smith S, Zhao X, Nair SJ, Michalek J, Jatoi I, Lautner M, et al. 2017. Attenuation of RNA polymerase II pausing mitigates BRCA1-associated R-loop accumulation and tumorigenesis. *Nat Commun* **8**: 1–11. https://www.nature.com/articles/ncomms15908.pdf.

Zhang Y, Liu T, Yuan F, Garcia-Martinez L, Lee KD, Stransky S, Sidoli S, Verdun RE, Zhang Y, Wang Z, et al. 2021. The Polycomb protein RING1B enables estrogen-mediated gene expression by promoting enhancer–promoter interaction and R-loop formation. *Nucleic Acids Res* gkab723. https://academic.oup.com/nar/advance-article/doi/10.1093/nar/gkab723/6357087 (Accessed August 25, 2021).

Zhang Y, Suzuki T, Li K, Gothwal SK, Shinohara M, Shinohara A. 2020. Genetic interactions of histone modification machinery set1 and PaF1C with the recombination complex Rec114-Mer2-Mei4 in the formation of meiotic DNA double-strand breaks. *Int J Mol Sci* **21**: 1–15.

Zheng W, Zhao H, Mancera E, Steinmetz LM, Snyder M. 2010. Genetic analysis of variation in transcription factor binding in yeast. *Nature* **464**: 1187–91. http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2941147&tool=pmcentrez&rendertype=abstract (Accessed December 17, 2013).

Zhong X, Du J, Hale CJ, Gallego-Bartolome J, Feng S, Vashisht AA, Chory J, Wohlschlegel JA, Patel DJ, Jacobsen SE. 2014. Molecular mechanism of action of plant DRM de novo DNA methyltransferases. *Cell* **157**: 1050–1060. http://dx.doi.org/10.1016/j.cell.2014.03.056.

Zhu D, Zhang Y, Wang S. 2021. Histone citrullination: a new target for tumors. *Mol Cancer* **20**: 1–17.

Zuo W, Chen G, Gao Z, Li S, Chen Y, Huang C, Chen J, Chen Z, Lei M, Bian Q. 2021. Stage-resolved Hi-C analyses reveal meiotic chromosome organizational features influencing homolog alignment. *Nat Commun* **12**: 5827. http://www.ncbi.nlm.nih.gov/pubmed/34625553.

A DISSZERTÁCIÓ ALAPJÁT KÉPEZŐ SAJÁT KÖZLEMÉNYEK

1. Karányi Zs. Mosolygó-L Á, Feró O, Horváth A, Boros-Oláh B, Nagy É, Hetey Sz, Holb I, Szaker H, Miskei M, Csorba T, **Székvölgyi L**: NODULIN HOMEOBOX is required for heterochromatin homeostasis in Arabidopsis. *Nature Communications.* 13 5058, 2022. (D1, IF: 17.694)
2. Karányi, Z., Hornyák, L., **Székvölgyi, L.**: Histone H3 lysine 56 acetylation is required for formation of normal levels of meiotic DNA breaks in S. cerevisiae. *Front. Cell. Dev. Biol.* 7 1-18, 2020. (Q1, IF: 6.684)
3. Boros-Oláh, B., Dobos, N., Hornyák, L., Szabó, Z., Karányi, Z., Halmos, G., Roszik, J., **Székvölgyi, L.**: Drugging the R-loop interactome: RNA-DNA hybrid binding proteins as targets for cancer therapy. *DNA Repair.* 84 1-10, 2019. (Q1, IF: 3.339)
4. Hegedűs, É., Kókai, E., Nánási, P., Imre, L., Halász, L., Jossé, R., Antunovics, Z., Webb, M., El Hage, A., Pommier, Y., **Székvölgyi, L.**, Dombrádi, V., Szabó, G.: Endogenous single-strand DNA breaks at RNA polymerase II promoters in Saccharomyces cerevisiae. *Nucleic Acids Res.* 46 (20), 10649-10668, 2018. (D1, IF: 11.147)
5. Karányi, Z., Halász, L., Acquaviva, L., Jonás, D., Hetey, S., Boros-Oláh, B., Peng, F., Chen, D., Klein, F., Géli, V., **Székvölgyi, L.**: Nuclear dynamics of the Set1C subunit Spp1 prepares meiotic recombination sites for break formation. *J. Cell Biol.* 217 (10), 3398-3415, 2018. (D1, IF: 8.891)
6. Hornyák, L., Dobos, N., Koncz, G., Karányi, Z., Páll, D., Szabó, Z., Halmos, G., **Székvölgyi, L.**: The role of indoleamine-2,3-dioxygenase (IDO) in cancer development, diagnostics, and therapy. *Front. Immunol.* 9 (151), 1-8, 2018. (Q1, IF: 4.716)
7. Roszik, J., Fenyőfalvi, G., Halász, L., Karányi, Z., **Székvölgyi, L.**: In Silico Restriction Enzyme Digests To Minimize Mapping Bias In Genomic Sequencing. *Mol. Ther. Methods. Clin. Dev.* 6 66-67, 2017. (Q2, IF: 3.681)
8. Halász, L., Karányi, Z., Boros-Oláh, B., Kuik-Rózsa, T., Sipos, É., Nagy, É., Mosolygó, Á., Türk-Mázló, A., Rajnavölgyi, É., Halmos, G., **Székvölgyi, L.**: RNA-DNA hybrid (R-loop) immunoprecipitation mapping: an analytical workflow to evaluate inherent biases. *Genome Res.* 27 1063-1073, 2017. (D1, IF: 10.101)
9. **Székvölgyi, L** *., Ohta, K., Nicolas, A*.: Initiation of meiotic homologous recombination: flexibility, impact of histone modifications, and chromatin remodeling. * levelező szerzők *Cold Spring Harbor Perspect. Biol.* 7 (5), 1-16, 2015. (D1, IF: 9.173) * megosztott levelező szerzők
10. Acquaviva, L*, **Székvölgyi, L*,** Dichtl, B., Dichtl, B., Saint André, C., Nicolas, A., Géli, V.: The COMPASS subunit Spp1 links histone methylation to initiation of meiotic recombination. *Science.* 339 (6116), 215-218, 2013. (D1, IF: 31.477) * megosztott elsőszerzők

11. **Székvölgyi, L.**, Nicolas, A.: From meiosis to postmeiotic events: Homologous recombination is obligatory but flexible. *FEBS J.* 277 (3), 571-589, 2010. (Q1, IF: 3.129)
12. **Székvölgyi, L.**, Rákosy, Z., Bálint, B., Kókai, E., Imre, L., Vereb, G., Bacsó, Z., Goda, K., Varga, S., Balázs, M., Dombrádi, V., Nagy, L., Szabó, G.: Ribonucleoprotein-masked nicks at 50-kbp intervals in the eukaryotic genomic DNA. *Proc. Natl. Acad. Sci. U.S.A.* 104 (38), 14964-14969, 2007. (D1, IF: 9.598)

EGYÉB ELSŐ ÉS UTOLSÓSZERZŐS KÖZLEMÉNYEK

13. Miskei, M., Horváth, A., Viola, L., Varga, L., Nagy, É., Feró, O., Karányi, Z., Roszik, J., Miskey, C., Ivics, Z., **Székvölgyi, L.**: Genome-wide mapping of binding sites of the transposase-derived SETMAR protein in the human genome. *Computational and Structural Biotechnology Journal.* 19 4032-4041, 2021. (D1, IF: 6.155)
14. Szabó, Z., Hornyák, L., Miskei, M., **Székvölgyi, L.**: Two targets, one hit: new anticancer therapeutics to prevent tumorigenesis without cardiotoxicity. *Front. Pharmacol.* 11 1-5, 2021. (Q1, IF: 5.988)
15. Hetey, S., Boros-Oláh, B., Kuik-Rózsa, T., Li, Q., Karányi, Z., Szabó, Z., Roszik, J., Szalóki, N., Vámosi, G., Tóth, K., **Székvölgyi, L.**: Biophysical characterization of histone H3.3 K27M point mutation. *Biochem. Biophys. Res. Commun.* 490 (3), 868-875, 2017. (Q1, IF: 2.559)
16. **Székvölgyi, L.**, Imre, L., Doan-Xuan, Q., Hegedűs, É., Bacsó, Z., Szabó, G.: Flow Cytometric and Laser Scanning Microscopic Approaches in Epigenetics Research. *Methods Mol. Biol.* 567 99-111, 2009. (Q3)
17. **Székvölgyi, L.**, Bálint, B., Imre, L., Goda, K., Szabó, M., Nagy, L., Szabó, G.: Chip-on-beads: flow-cytometric evaluation of chromatin immunoprecipitation. *Cytometry A.* 69A (10), 1086-1091, 2006. (Q1, IF: 3.293)
18. **Székvölgyi, L.**, Hegedűs, É., Molnár, M., Bacsó, Z., Szarka, K., Beck, Z., Dombrádi, V., Austin, C., Szabó, G.: Nick-forming sequences may be involved in the organization of eukaryotic chromatin into ~50 kbp loops. *Histochem. Cell Biol.* 125 (1-2), 63-73, 2006. (D1, IF: 3.22)

FÜGGELÉK

dc_1977_21

EXTENDED PDF FORMAT
SPONSORED BY
IMMUNOLOGY2013™ Travel Grants
from R&D Systems
$1000 travel grants will be awarded to 10 immunology researchers
Apply Now!
www.rndsystems.com

The COMPASS Subunit Spp1 Links Histone Methylation to Initiation of Meiotic Recombination
Laurent Acquaviva *et al.*
Science **339**, 215 (2013);
DOI: 10.1126/science.1225739

This copy is for your personal, non-commercial use only.

If you wish to distribute this article to others, you can order high-quality copies for your colleagues, clients, or customers by clicking here.

Permission to republish or repurpose articles or portions of articles can be obtained by following the guidelines here.

The following resources related to this article are available online at www.sciencemag.org (this information is current as of January 16, 2013):

Updated information and services, including high-resolution figures, can be found in the online version of this article at:
http://www.sciencemag.org/content/339/6116/215.full.html

Supporting Online Material can be found at:
http://www.sciencemag.org/content/suppl/2012/11/14/science.1225739.DC1.html

This article **cites 30 articles**, 10 of which can be accessed free:
http://www.sciencemag.org/content/339/6116/215.full.html#ref-list-1

This article appears in the following **subject collections:**
Molecular Biology
http://www.sciencemag.org/cgi/collection/molec_biol

Downloaded from www.sciencemag.org on January 16, 2013

Science (print ISSN 0036-8075; online ISSN 1095-9203) is published weekly, except the last week in December, by the American Association for the Advancement of Science, 1200 New York Avenue NW, Washington, DC 20005. Copyright 2013 by the American Association for the Advancement of Science; all rights reserved. The title *Science* is a registered trademark of AAAS.

dc. 1977 21

The COMPASS Subunit Spp1 Links Histone Methylation to Initiation of Meiotic Recombination

Laurent Acquaviva,[1]* Lóránt Székvölgyi,[2,4]* Bernhard Dichtl,[3]† Beatriz Solange Dichtl,[3] Christophe de La Roche Saint André,[1]† Alain Nicolas,[2]† Vincent Géli[1]†

During meiosis, combinatorial associations of genetic traits arise from homologous recombination between parental chromosomes. Histone H3 lysine 4 trimethylation marks meiotic recombination hotspots in yeast and mammals, but how this ubiquitous chromatin modification relates to the initiation of double-strand breaks (DSBs) dependent on Spo11 remains unknown. Here, we show that the tethering of a PHD-containing protein, Spp1 (a component of the COMPASS complex), to recombinationally cold regions is sufficient to induce DSB formation. Furthermore, we found that Spp1 physically interacts with Mer2, a key protein of the differentiated chromosomal axis required for DSB formation. Thus, by interacting with H3K4me3 and Mer2, Spp1 promotes recruitment of potential meiotic DSB sites to the chromosomal axis, allowing Spo11 cleavage at nearby nucleosome-depleted regions.

Meiotic recombination is initiated by the introduction of DNA double-strand breaks (DSBs) by Spo11, a meiosis-specific transesterase, which is highly conserved throughout evolution (*1*). To ensure at least one crossover per chromosome pair, a large number of DSBs are formed in each meiotic cell, and their occurrence is controlled at multiple levels, comprising both higher-order chromosome structure and local determinants (*2*). In *Saccharomyces cerevisiae*, ~150 DSBs are formed per meiosis, preferentially in promoter regions and with a highly variable frequency. DSB sites are largely independent of DNA sequence composition (*3*). The control of DSB formation also depends on a number of Spo11-accessory proteins that have been shown to form subcomplexes, but whose functions are only partially understood (*4*). In particular, the Mer2/Mei4/Rec114 complex plays a role in linking replication to subsequent DSB formation through Mer2 phosphorylation (*5*–*7*). This complex is also thought to tether potential DSB sites located on chromatin loops to the highly differentiated meiotic chromosome axis, leading to stepwise activation of Spo11 cleavage (*8*–*10*).

In *S. cerevisiae*, all H3K4 methylation is carried out by the COMPASS protein complex that includes Set1, the catalytic subunit, which acts as a scaffold for the other structural and regulatory components (*11*). In particular, the PHD-finger subunit Spp1 specifically regulates the H3K4me3 state (*11*, *12*). Our previous studies revealed that the absence of Set1 severely reduces meiotic DSB levels (*13*, *14*) and that the level of H3K4me3 is constitutively higher near DSB sites (*14*). However, beyond these correlations, the mechanistic link between H3K4 methylation and DSB formation has remained elusive (*15*).

To uncover functional connections, we first asked whether the mutation of each COMPASS subunit affected DSB frequencies at natural recombination hotspots (*16*). Similar to the deletion of *SET1*, the absence of each COMPASS subunit, except Shg1, reduced DSB frequencies at the *BUD23* (Fig. 1A) and *CYS3* hotspots (fig. S1A; strain genotypes in table S1). We observed the frequency reduction for Spp1, which is specifically required for H3K4me3 formation. The similar effect of *set1*Δ and *spp1*Δ mutants on DSB formation at hotspots also extends to DSB formation near the naturally cold *PES4* locus (fig. S1B) (*14*). To check if the decrease in DSB formation at hotspots reflected the role of only the COMPASS complex on H3K4 methylation, we asked whether this decrease was recapitulated by mutation of the H3K4 residue. At *BUD23*, we observed equivalent DSB reduction in the *set1*Δ and *H3K4R* strains (fig. S1C). Parallel analysis of the *set1*Δ *H3K4R* double mutant revealed an identical effect,

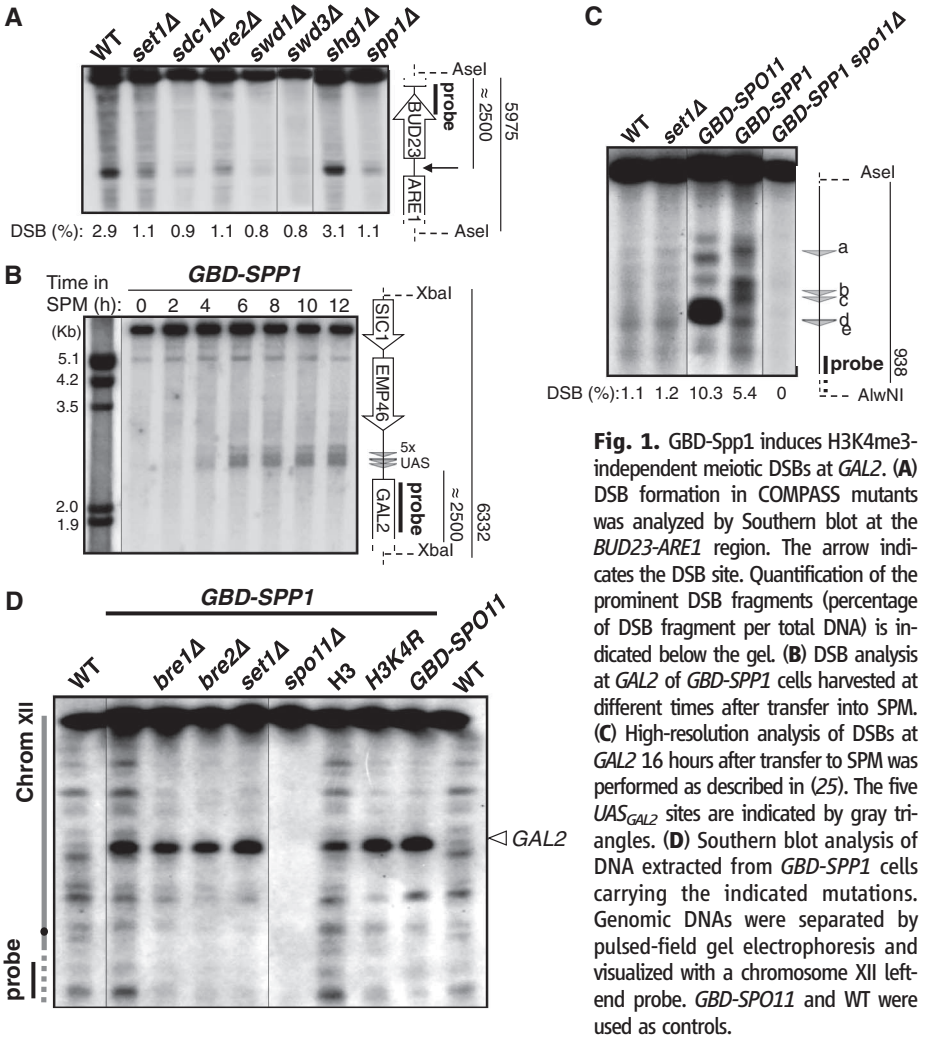


Fig. 1. GBD-Spp1 induces H3K4me3-independent meiotic DSBs at *GAL2*. (**A**) DSB formation in COMPASS mutants was analyzed by Southern blot at the *BUD23-ARE1* region. The arrow indicates the DSB site. Quantification of the prominent DSB fragments (percentage of DSB fragment per total DNA) is indicated below the gel. (**B**) DSB analysis at *GAL2* of *GBD-SPP1* cells harvested at different times after transfer into SPM. (**C**) High-resolution analysis of DSBs at *GAL2* 16 hours after transfer to SPM was performed as described in (*25*). The five UAS_{GAL2} sites are indicated by gray triangles. (**D**) Southern blot analysis of DNA extracted from *GBD-SPP1* cells carrying the indicated mutations. Genomic DNAs were separated by pulsed-field gel electrophoresis and visualized with a chromosome XII left-end probe. *GBD-SPO11* and WT were used as controls.

[1]Marseille Cancer Research Center (CRCM), U1068 Inserm, UMR7258 CNRS, Aix-Marseille University, Institut Paoli-Calmettes, Marseille 13009, France. [2]Institut Curie Centre de Recherche, UMR3244 CNRS, Université Pierre et Marie Curie, 75248 Paris, France. [3]Centre for Cellular and Molecular Biology, School of Life and Environmental Sciences, Deakin University, Victoria 3125, Australia. [4]Department of Biophysics and Cell Biology, Medical and Health Science Center, University of Debrecen, Debrecen, Hungary.

*These authors contributed equally to this work.
†To whom correspondence should be addressed. E-mail: b.dichtl@deakin.edu.au (B.D.); laroche@ifr88.cnrs-mrs.fr (C.d.L.R.S.A.); alain.nicolas@curie.fr (A.N.); geli@ifr88.cnrs-mrs.fr (V.G.)

Downloaded from www.sciencemag.org on January 16, 2013

dc_1977_21

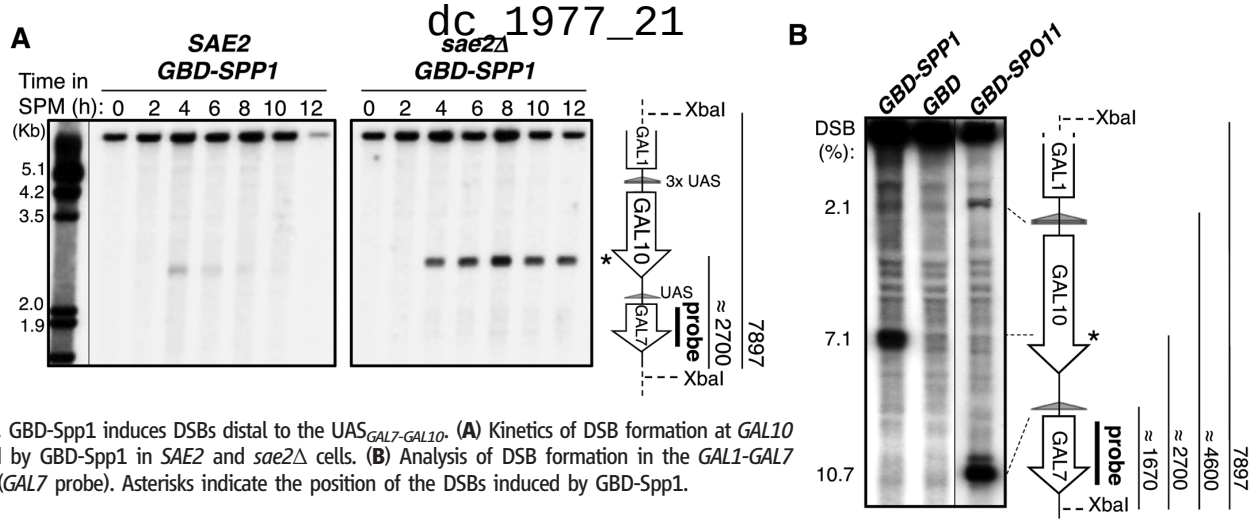


Fig. 2. GBD-Spp1 induces DSBs distal to the $UAS_{GAL7\text{-}GAL10}$. (**A**) Kinetics of DSB formation at *GAL10* induced by GBD-Spp1 in *SAE2* and *sae2*Δ cells. (**B**) Analysis of DSB formation in the *GAL1-GAL7* region (*GAL7* probe). Asterisks indicate the position of the DSBs induced by GBD-Spp1.

underscoring their epistatic relationship (fig. S1C). Together, these observations strongly support a role for the COMPASS complex and H3K4 trimethylation in the control of DSB formation and tie them to the same pathway.

To elucidate the role of H3K4me3, Set1, and Spp1 in DSB formation, we next asked whether the tethering of Set1 and Spp1 proteins to UAS_{GAL} (UAS, upstream activating sequence) binding sites located in cold DSB regions was sufficient to stimulate DSB formation, as previously observed for the fusion of the Gal4 binding domain (GBD) with Spo11 (*17*). We generated and expressed in-frame fusions of the coding region with the GBD. With respect to H3K4 methylation and DSB formation, both GBD-Set1 and GBD-Spp1 fusions behaved the same as the wild type, indicating no adverse interference with COMPASS and the Spo11 machinery (fig. S2 and table S2). We examined DSB formation in the naturally cold *GAL2* locus, which contains five UAS_{GAL} sequences in its promoter. Notably, the GBD-Set1 construct stimulated DSBs near the UAS_{GAL2} sites (fig. S3, A and B). In terms of genetic requirements, the data reported in fig. S3, C to E, establish that the GBD-Set1–induced DSBs were not associated with a coincident increase in H3K4me3 at *GAL2* but did depend on (i) the presence of Spo11, (ii) the integrity of the COMPASS, (iii) the histone methyl transferase activity of Set1, and (iv) the presence of the H3 lysine 4 residue.

Similar analyses revealed that GBD-Spp1 strongly stimulated DSB formation at *GAL2*. These DSBs appeared and disappeared upon repair with similar kinetics as natural DSBs (Fig. 1, B and C, see also Fig. 2A). Notably, DSBs targeted by GBD-Spp1 required neither the presence of Set1 or Bre2 nor the E3 ligase Bre1 that controls H3K4 trimethylation through the ubiquitinylation of H2B (*18*). In agreement with this observation, DSBs resulting from the tethering of Spp1 also appeared in the *H3K4R* mutant (Fig. 1D). Tethering Spp1 to the *GAL1*/*GAL10* and *GAL7* loci also efficiently induced DSBs (Fig. 2A), but, surprisingly, these breaks did not occur

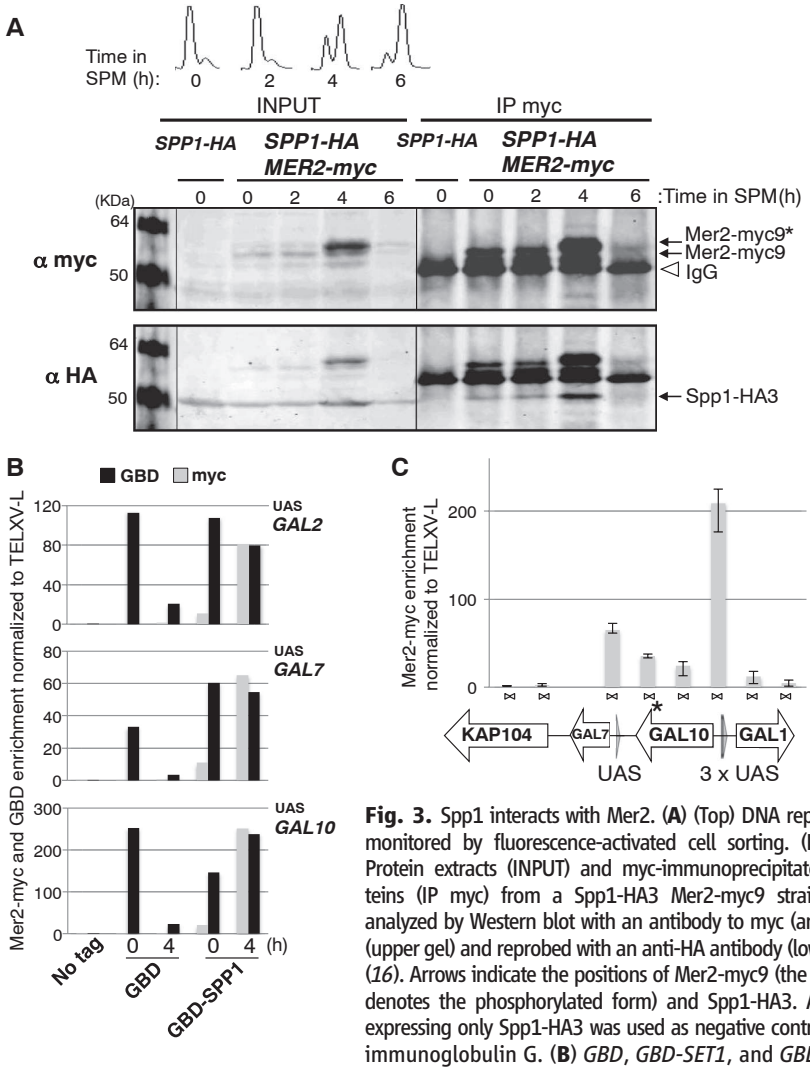


Fig. 3. Spp1 interacts with Mer2. (**A**) (Top) DNA replication monitored by fluorescence-activated cell sorting. (Bottom) Protein extracts (INPUT) and myc-immunoprecipitated proteins (IP myc) from a Spp1-HA3 Mer2-myc9 strain were analyzed by Western blot with an antibody to myc (anti-myc) (upper gel) and reprobed with an anti-HA antibody (lower gel) (*16*). Arrows indicate the positions of Mer2-myc9 (the asterisk denotes the phosphorylated form) and Spp1-HA3. A strain expressing only Spp1-HA3 was used as negative control. IgG, immunoglobulin G. (**B**) *GBD*, *GBD-SET1*, and *GBD-SPP1* strains expressing Mer2-myc9 were harvested after 0 and 4 hours in SPM. Binding of the GBD-fusion proteins and Mer2-myc9 at the *GAL2*, *GAL7*, and *GAL10* UASs was analyzed by chromatin immunoprecipitation (ChIP) with anti-GBD and anti-myc antibodies. Polymerase chain reaction (PCR) primers specific for each GAL_{UAS} and TELXV-L (control) were used to amplify immunoprecipitated DNAs (*16*). (**C**) Chromatin binding of Mer2-myc9 in the *GBD-SPP1* strain was analyzed by ChIP–quantitative PCR after 4 hours in SPM along the genomic region comprising *GAL7*, *GAL10*, and *GAL1* at positions indicated by two convergent arrowheads. The asterisk indicates the Gal1 ucut promoter. Error bars represent SDs from three independent experiments.

Downloaded from www.sciencemag.org on January 16, 2013

dc_1977_21

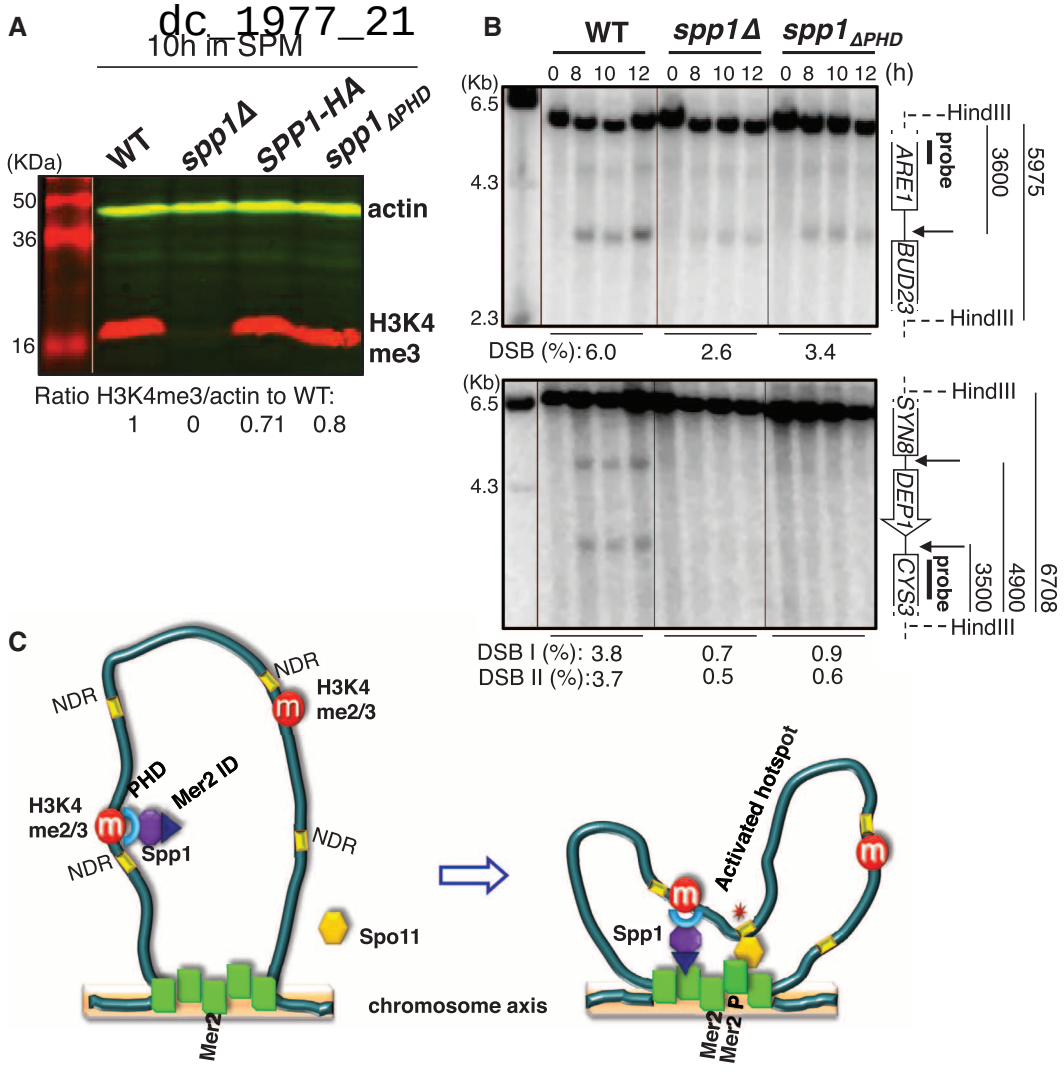


Fig. 4. The PHD-finger domain of Spp1 regulates meiotic DSB formation. (**A**) Analysis of H3K4me3 levels (*16*). (**B**) DSB formation at the *BUD23* and *CYS3* hotspots. (**C**) Extension of the chromatin loop-axis model. The tethering of a H3K4me-rich region to the chromosomal axis via the Spp1-Mer2 interaction allows Spo11 cleavage at a proximal nucleosome-depleted region (NDR). The red circles labeled with "m" indicate di- and trimethylated H3K4.

at the same location as those observed in the *GBD-SPO11* (Fig. 2B). Whereas tethering of Spo11 introduced DSBs in the vicinity the *UAS* sites, tethering of Spp1 to the same sites produced cleavage in the 3′ end of the *GAL10* coding sequence (Fig. 2B and fig. S4A). As for the majority of meiotic DSBs in *S. cerevisiae*, occurring in promoter regions, the unexpected location of these GBD-Spp1–induced DSBs coincides with the promoter of the well-characterized antisense *GAL1* ucut (SUT013) noncoding RNA. This RNA is expressed in meiosis (*19*) under the control of the transcription factor Reb1 (*20*). This situation is also consistent with the high-resolution, genome-wide map of DSBs, which revealed a high enrichment of Spo11 binding in nucleosome-depleted regions adjacent to the Reb1 binding sites (*3*). As in *GAL2*, the DSBs generated at *GAL10* required Spo11 and occurred in the absence of H3K4 methylation (fig. S4B). Together, these results demonstrated that the tethering of Spp1 efficiently stimulated Spo11-dependent DSBs independently of H3K4 methylation.

A simple interpretation of the above results is that Spp1 recruits a component of the DSB machinery. To search for interaction partners, we used the full-length Spp1 protein to perform yeast two-hybrid screening (*16*). We readily identified Set1 and Mer2 as prominent interactors of Spp1 (table S3). We mapped the Spp1-interacting region of Mer2 to the central part of the protein (residues 105 to 172) (fig. S5A). We validated the Mer2-Spp1 interaction by glutathione *S*-transferase–Mer2 pull-down experiments using various in vitro–translated Spp1 polypeptides. We found that the 131–amino acid C-terminal region of Spp1 was required for interaction with Mer2 (fig. S5B). Finally, we used tagged versions of both proteins to examine the in vivo interaction of Spp1 and Mer2 during meiosis. The C-terminally tagged *SPP1-HA3* strain showed a level of H3K4me3 near that of the wild type and a slight reduction (35%) of DSB formation at the *BUD23* hotspot, but, nevertheless, wild-type (WT) spore viability, indicating efficient DSB formation throughout the genome (fig. S6 and table S2). Spp1-HA3 was efficiently coimmunoprecipitated with Mer2-myc9 during meiosis, with a peak at 4 hours (Fig. 3A). Deoxyribonuclease and phosphatase treatments of the Myc immunoprecipitated proteins from the double-tag strain did not reduce the recovery of Spp1-HA (fig. S5, B and C), suggesting an interaction of Spp1 with the nonphosphorylated form of Mer2 (*5*, *7*, *9*). A schematic representation of the Spp1-Set1 and Spp1-Mer2 interaction domains is illustrated in fig. S5D.

The essential role of Mer2 in DSB formation and its interaction with Spp1 prompted us to determine whether Mer2 was present at the Spp1-tethered DSB sites. At the control *CTR86* and *SED4* regions known to associate with Mer2 in WT cells (*9*), we found that Mer2 was enriched in the strains expressing GBD and GBD-Spp1 (fig. S7A), suggesting that the expression of these GBD fusion proteins did not alter chromatin occupancy of Mer2. Four hours after transfer to the sporulation media (SPM), Mer2 was strongly enriched at the UAS_{GAL} sites in the strain expressing GBD-Spp1, but not in the strains expressing *GBD* alone (Fig. 3B). Importantly, Mer2-myc was also enriched in the vicinity of the Gal1 ucut promoter, where the major DSB site is detected (Fig. 3C). This enrichment appears to be specific, as its level is greater than the one expected to spread for the UAS sites.

Downloaded from www.sciencemag.org on January 16, 2013

doi: 10.1126/science.1225739

Examination of the genetic requirements for DSB formation indicated that the cleavage site induced by the tethering of Spp1 (i) was independent of Set1, (ii) required Mer2, and (iii) did not result from Spp1 overexpression (fig. S7B). The previously described Spp1 zinc finger–like domain (SZF) was included in the Mer2-interacting region (*21*). We deleted the SZF canonical CXXC motif in GBD-Spp1 (GBD-Spp1$_{\Delta CXXC}$) and tested whether deleting this evolutionarily conserved motif impairs DSB formation and Mer2 binding at *GAL10* (*16*). DSB formation and Mer2 binding were both strongly reduced at *GAL10*, providing a strong correlation between the ability of GBD-Spp1 to recruit Mer2 at *GAL10* and DSB formation (fig. S8). This finding outlines the importance of this interaction for DSB formation.

What is the role of H3K4 methylation in the control of DSB formation? We found that the amount of Spp1 was strongly reduced both in the absence of Set1 and in COMPASS mutants affecting the stability of Set1, but either remained normal in the catalytically inactive *set1G951S* mutant or was slightly reduced in the *H3K4R*, *bre2*Δ, and *sdc1*Δ mutants (fig. S9, A and B). Therefore, decreased levels of Spp1 might only partially explain the DSB formation defect of mutants that are compromised in H3K4 methylation. Our interpretation is that the function of Spp1 in DSB formation depends on H3K4 methylation through an interaction of its PHD-finger with H3K4me2/3 (*21*). We therefore deleted the PHD domain of Spp1 (*16*) and tested whether the presence of the PHD domain was important for H3K4me3 and DSB formation. We detected a level of H3K4me3 near that of the wild type in the *spp1*$_{\Delta PHD}$ mutant (Fig. 4A), whereas DSB formation was clearly reduced at the *BUD23* and *CYS3* hotspots (Fig. 4B). We conclude that the Spp1 PHD domain itself plays a specific role in DSB formation.

Our work led us to propose an enriched chromatin loop-axis model (Fig. 4C) for the regulation of DSB formation that addresses how the meiotic DSB sites are mechanistically selected. We propose that the interaction between the COMPASS subunit, Spp1, and Mer2 brings potential meiotic DSB sites to the chromosome axis for further downstream events that will ultimately lead to Spo11-dependent DSB formation at axis-proximal regions that are depleted of nucleosomes. In mammals, H3K4 methylation has been reported to be enriched at recombination hotspots where the meiosis-specific PRDM9 H3K4 methyltransferase is known to act (*22*–*24*). How PRDM9 connects to the mammalian DSB machinery is not known, but, as in yeast, it may be the consequence of a direct interaction with a protein of the DSB machinery. Our results indicate that H3K4me3 is required for the function of Spp1, probably through its recognition by the PHD domain within Spp1, and this requirement can be bypassed by tethering Spp1 to the DNA locus. The broadly localized H3K4me3 modification has the virtue of permitting the initiation of recombination at numerous places of the genome, a molecular strategy that ensures a large diversity of recombinant haplotypes to be transmitted by the gametes. In conclusion, this model offers a clue of how chromosome structure and DSB regulation are interrelated, and it attributes a pivotal role to Spp1 in the recruitment of components acting at meiotic DSB sites to the chromosomal axis.

References and Notes

1. D. Zickler, N. Kleckner, *Annu. Rev. Genet.* **33**, 603 (1999).
2. M. Lichten, B. de Massy, *Cell* **147**, 267 (2011).
3. J. Pan *et al.*, *Cell* **144**, 719 (2011).
4. S. Keeney, *Genome Dyn Stab.* **2**, 81 (2008).
5. K. A. Henderson, K. Kee, S. Maleki, P. A. Santini, S. Keeney, *Cell* **125**, 1321 (2006).
6. J. Li, G. W. Hooker, G. S. Roeder, *Genetics* **173**, 1969 (2006).
7. H. Sasanuma *et al.*, *Genes Dev.* **22**, 398 (2008).
8. Y. Blat, R. U. Protacio, N. Hunter, N. Kleckner, *Cell* **111**, 791 (2002).
9. S. Panizza *et al.*, *Cell* **146**, 372 (2011).
10. T. Miyoshi *et al.*, *Mol. Cell* **47**, 722 (2012).
11. P. M. Dehé *et al.*, *J. Biol. Chem.* **281**, 35404 (2006).
12. J. Schneider *et al.*, *Mol. Cell* **19**, 849 (2005).
13. J. Sollier *et al.*, *EMBO J.* **23**, 1957 (2004).
14. V. Borde *et al.*, *EMBO J.* **28**, 99 (2009).
15. S. E. Tischfield, S. Keeney, *Cell Cycle* **11**, 1496 (2012).
16. Materials and methods are available as supplementary materials on *Science* Online.
17. A. Peciña *et al.*, *Cell* **111**, 173 (2002).
18. A. Wood *et al.*, *Mol. Cell* **11**, 267 (2003).
19. A. Lardenois *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **108**, 1058 (2011).
20. M. Pinskaya, S. Gourvennec, A. Morillon, *EMBO J.* **28**, 1697 (2009).
21. B. L. Murton, W. L. Chin, C. P. Ponting, L. S. Itzhaki, *J. Mol. Biol.* **398**, 481 (2010).
22. F. Baudat *et al.*, *Science* **327**, 836 (2010).
23. S. Myers *et al.*, *Science* **327**, 876 (2010).
24. F. Smagulova *et al.*, *Nature* **472**, 375 (2011).
25. H. Murakami, A. Nicolas, *Mol. Cell. Biol.* **29**, 3500 (2009).

Acknowledgments: We thank F. Klein and V. Borde for materials, J. B. Boulé for his help, and S. Kowalczykowski for reading of the manuscript. Work in the V.G. and A.N. laboratories is supported by the "Ligue contre le Cancer" (Equipe Labelisées). Work in the laboratory of B.D. was supported by the Swiss National Science Foundation. L.A. was supported by grants from the Fondation ARC and L.Sz. by the European Union and the Hungarian Research Fund.

Supplementary Materials

www.sciencemag.org/cgi/content/full/science.1225739/DC1
Materials and Methods
Figs. S1 to S9
Tables S1 to S5
References (*26*–*31*)

6 June 2012; accepted 5 November 2012
Published online 15 November 2012;
10.1126/science.1225739

JNK Expression by Macrophages Promotes Obesity-Induced Insulin Resistance and Inflammation

Myoung Sook Han,[1,2] Dae Young Jung,[2] Caroline Morel,[1,2] Saquib A. Lakhani,[3]* Jason K. Kim,[2,4] Richard A. Flavell,[3] Roger J. Davis[1,2]†

The cJun NH_2-terminal kinase (JNK) signaling pathway contributes to inflammation and plays a key role in the metabolic response to obesity, including insulin resistance. Macrophages are implicated in this process. To test the role of JNK, we established mice with selective JNK deficiency in macrophages. We report that feeding a high-fat diet to control and JNK-deficient mice caused similar obesity, but only mice with JNK-deficient macrophages remained insulin-sensitive. The protection of mice with macrophage-specific JNK deficiency against insulin resistance was associated with reduced tissue infiltration by macrophages. Immunophenotyping demonstrated that JNK was required for pro-inflammatory macrophage polarization. These studies demonstrate that JNK in macrophages is required for the establishment of obesity-induced insulin resistance and inflammation.

Obesity is an important public health problem that is associated with inflammation, cardiovascular disease, metabolic syndrome, and type 2 diabetes (*1*). Tissue infiltration by macrophages is a major contributor to inflammation and insulin resistance (*2*). Tissue macrophages comprise multiple populations (*3*); however, there are two well-known subtypes that are capable of dynamic interconversion (*4*). Classically activated macrophages (M1) induced by interferon-γ (IFN-γ) or endotoxin promote interleukin-12 (IL12)–mediated T helper 1 (T_H1) immune responses. Alternatively, activated macrophages induced by IL4 or IL13 (M2a), immune complexes (M2b), and the anti-inflammatory cytokines IL10 or transforming growth factor-β (M2c) can promote T_H2 immune responses and mediate wound healing, tissue repair, and the resolution of inflammation. Obesity increases tissue infiltration by macrophages (*5*) and polarization to the pro-inflammatory M1 state (fig. S1) (*6*, *7*). Indeed, the inflammation associated with M1-polarized tissue macrophages is implicated in the development of obesity-related insulin resistance (*8*–*11*).

Downloaded from www.sciencemag.org on January 16, 2013

Published Online: 23 July, 2018 | Supp Info: http://doi.org/10.1083/jcb.201712122
Downloaded from jcb.rupress.org on October 1, 2018

dc_1977_21

JCB Journal of Cell Biology

ARTICLE

Nuclear dynamics of the Set1C subunit Spp1 prepares meiotic recombination sites for break formation

Zsolt Karányi[1,2], László Halász[1], Laurent Acquaviva[3], Dávid Jónás[1], Szabolcs Hetey[1], Beáta Boros-Oláh[1], Feng Peng[4], Doris Chen[4], Franz Klein[4], Vincent Géli[3]*, and Lóránt Székvölgyi[1]*

Spp1 is the H3K4me3 reader subunit of the Set1 complex (COMPASS/Set1C) that contributes to the mechanism by which meiotic DNA break sites are mechanistically selected. We previously proposed a model in which Spp1 interacts with H3K4me3 and the chromosome axis protein Mer2 that leads to DSB formation. Here we show that spatial interactions of Spp1 and Mer2 occur independently of Set1C. Spp1 exhibits dynamic chromatin binding features during meiosis, with many de novo appearing and disappearing binding sites. Spp1 chromatin binding dynamics depends on its PHD finger and Mer2-interacting domain and on modifiable histone residues (H3R2/K4). Remarkably, association of Spp1 with Mer2 axial sites reduces the effective turnover rate and diffusion coefficient of Spp1 upon chromatin binding, compared with other Set1C subunits. Our results indicate that "chromosomal turnover rate" is a major molecular determinant of Spp1 function in the framework of meiotic chromatin structure that prepares recombination initiation sites for break formation.

Introduction

Regulation of chromatin structure through covalent histone modifications is a central mechanism for modulating DNA-directed biological processes, including gene transcription, mRNA processing, and DNA replication, recombination, and repair. Histone modifications act by directly altering the chromatin structure or by creating docking sites that facilitate the binding of chromatin readers (Rothbart and Strahl, 2014). These readers in turn recruit remodeling enzymes or additional chromatin modifiers (Tessarz and Kouzarides, 2014). In the past years, methylation of lysine 4 on histone H3 (H3K4) has received considerable attention and was linked to several aspects of transcriptional regulation (Ruthenburg et al., 2007) and class-switch recombination (Daniel et al., 2010), S-phase DNA damage checkpoint (Liu et al., 2010), and meiotic recombination (Borde and de Massy, 2013). The family of H3K4 methylases is highly conserved from yeast to human (Shilatifard, 2012). They share a canonical organization in which the catalytic subunit acts as a docking platform for multiple subunits that regulates the activity of the enzyme (Ernst and Vakoc, 2012). The budding yeast Set1 complex has proved to be an excellent model to study the *SET1/MLL* family complexes. In *Saccharomyces cerevisiae*, all H3K4 methylation is performed by a complex called Complex of Proteins Associated with Set1 (COMPASS; Miller et al., 2001) or Set1C (Roguev et al., 2001) composed of Set1, the catalytic subunit, acting as a scaffold for seven other components (Swd1 [RbBP5], Swd2 [Wdr82], Swd3 [Wdr5], Bre2 [Ashl2], Sdc1 [Dpy30], Spp1 [Cfp1], and Shg1 [Bod1]; Miller et al., 2001; Nagy et al., 2002). In the past years, several studies contributed to define how each subunit of Set1C was bound to the docking platform established by the catalytic Set1 subunit. Swd1, Swd3, Bre2, and Sdc1 were shown to interact with the isolated SET domain to form the SET-c (Dehé et al., 2006; Trésaugues et al., 2006; Kim et al., 2013). On the other hand, Spp1, Swd2, and Shg1 directly interact with the n-SET domain, the N-terminal domain, and the second RNA-recognition motif (RRM) motif of Set1, respectively (Dehé et al., 2006; Halbach et al., 2009; Kim et al., 2013). Loss of individual Set1C subunits differentially affects Set1 stability, complex integrity, global H3K4 methylation patterns, and H3K4 methylation along active genes (Soares et al., 2014).

The recruitment of Set1C to chromatin is not fully understood. Set1C has been shown to be targeted to the 5′ regions of transcription units via the Paf1C elongation factor and the CTD of Pol II (Krogan et al., 2003; Ng et al., 2003). These interactions are thought to contribute to the prevalence of H3K4me3 at chromatin domains at the 5′-end of active genes. However, directly interacting protein(s) that would recruit Set1C to actively tran-

[1]MTA-DE Momentum Genome Architecture and Recombination Research Group, Department of Biochemistry and Molecular Biology, Faculty of Medicine, University of Debrecen, Debrecen, Hungary; [2]Department of Internal Medicine, University of Debrecen, Debrecen, Hungary; [3]Marseille Cancer Research Center (CRCM), U1068 Institut National de la Santé et de la Recherche Médicale, UMR7258 Centre National de la Recherche Scientifique, Aix Marseille University, Institut Paoli-Calmettes, Marseille, France. Equipe labellisée Ligue; [4]Max F. Perutz Laboratories, University of Vienna, Vienna, Austria.

*V. Géli and L. Székvölgyi contributed equally to this paper; Correspondence to Lóránt Székvölgyi: lorantsz@med.unideb.hu; Vincent Géli: vincent.geli@inserm.fr.

© 2018 Karányi et al. This article is distributed under the terms of an Attribution–Noncommercial–Share Alike–No Mirror Sites license for the first six months after the publication date (see http://www.rupress.org/terms/). After six months it is available under a Creative Commons License (Attribution–Noncommercial–Share Alike 4.0 International license, as described at https://creativecommons.org/licenses/by-nc-sa/4.0/).

scribed genes have not yet been identified. Recently, Set1C was shown to bind mRNAs in vitro and in vivo (Trésaugues et al., 2006; Battaglia et al., 2017; Luciano et al., 2017; Sayou et al., 2017). Unexpectedly, multiple protein surfaces in Set1C, as well as the dRRM, N-SET domain, and Spp1, were shown to be important to bind RNA in vitro. RNA binding of Set1 was found to be important for the proper topology of Set1C distribution along transcription units (Luciano et al., 2017; Sayou et al., 2017).

Meiosis is a differentiation process involving two successive cell divisions required for the formation of haploid nuclei and gametes from germ cells (Székvölgyi and Nicolas, 2010; Székvölgyi et al., 2015). During the first meiotic division, meiotic recombination is initiated by the formation of DNA double-strand breaks (DSBs) catalyzed by the meiosis-specific type II topoisomerase-like DNA transesterase Spo11 (Keeney et al., 1997). In *S. cerevisiae*, DSBs localize to specific regions called hot spots that mainly overlap with nucleosome depleted intergenic regions, near promoters (Pan et al., 2011). DSB formation also requires several Spo11-associated proteins that form sub-complexes (Lam and Keeney, 2015). The Mer2/Mei4/Rec114 (RMM) sub-complex has been proposed to link DSB sites located within chromatin loops to chromosome axial structures to undergo Spo11-mediated cleavage (Panizza et al., 2011). It was initially observed that Set1 inactivation severely reduced the level and distribution of meiotic DNA breaks (Sollier et al., 2004). Moreover, nucleosomes flanking DSB sites were shown to be enriched in histone H3K4 trimethylation that was independent of the mRNA expression level of nearby genes (Borde et al., 2009). Consistent with these observations, inactivation of *RAD6* (Yamashita et al., 2004) or the PAF1 complex (Gothwal et al., 2016) that both reduce H3K4 methylation significantly also reduced meiotic DSB frequencies; however, this by itself did not prove that H3K4me directly promoted DSB formation. The link between H3K4me3 and meiotic DNA breaks is conserved in many organisms including mammals, where the H3K4me3 mark is deposited by a sequence-specific histone methylase, Prdm9, which directs DSBs to certain DNA motifs recognized by its zinc finger domain (Baudat et al., 2010; Parvanov et al., 2010).

The mechanism by which H3K4 methylation is linked to DSB formation was further highlighted by the discovery that Spp1 (the PHD-finger subunit of Set1C) can physically interact with both H3K4me2/3 and Mer2, located at the meiotic chromosomal axis (Acquaviva et al., 2013b; Sommermeyer et al., 2013). It was proposed that the interaction between Spp1 and Mer2 anchors meiotic DSB hot spots to the chromosome axis for downstream events that will ultimately lead to Spo11-dependent DSB formation at axis-proximal regions (Acquaviva et al., 2013a). This model proposed an explanation of how chromosome architecture and DSB regulation are interrelated at the molecular level and revealed the key role of Spp1 in recruiting meiotic DSB sites to the chromosome axis.

The above data demonstrate that Spp1 not only regulates the catalytic activity of Set1C, but also interacts with the deposited H3K4me3 mark and mediates its downstream effects. Remarkably, dissociation of Spp1 from Set1C repurposed its biological function to promote epigenetic transcriptional memory at the *INO1* gene (D'Urso et al., 2016). These results refer to the highly dynamic behavior of Spp1; however, the nuclear dynamics of Set1C subunits have not been directly and systematically studied to date. The temporal and spatial sequence of events controlling the chromatin binding of Set1C and the exact molecular mechanism of Spp1 chromosomal redistribution during meiosis have also remained important unanswered questions.

Here, we used dynamic chromatin mapping and quantitative imaging to unravel the chromatin binding characteristics and turnover rate of Set1C and Spp1 in live meiotic nuclei. We present a detailed spatio-temporal picture showing how Spp1 becomes redistributed from actively transcribed genes to chromosome axial sites, independently of Set1C.

Results

Spp1 and Mer2 partially colocalize in meiotic chromosome spreads

Since Spp1 was shown to physically interact with Mer2 and to bind to Mer2 chromosomal sites at the time of meiotic DSB formation (Acquaviva et al., 2013b; Sommermeyer et al., 2013; Adam et al., 2018), we tagged Spp1/Mer2 (and Bre2/Mer2) to perform double-immunofluorescence labeling on meiotic chromosome spreads to ascertain whether Spp1/Mer2 would colocalize in one and the same nucleus by the time of DSB formation. Cells coexpressing Mer2 internally tagged with 3xHA, Mer2-HA.int, and Spp1-myc (or Bre2-myc) as their only source for these proteins were synchronized, and samples were collected at different time points during the meiotic time course. As exemplified in Fig. 1 A, Spp1 and Mer2 foci colocalized 4 h after transfer to sporulation medium (SPM), whereas Bre2 and Mer2 foci (Fig. 1 B) did not. Quantification of Spp1 foci overlapping with Mer2 showed colocalization for all time points analyzed, reaching a plateau at 3–6 h after transfer to SPM, with ∼50% of Spp1 foci overlapping with Mer2 foci (Fig. 1 C). To assign statistical significance for colocalization frequencies, we simulated random colocalization using FociSim (Kurzbauer et al., 2012). For each nucleus, Monte Carlo simulations with 200,000 random seeds were performed, with nuclear area, foci numbers, and foci areas from the experimental data, yielding a distribution of random overlaps (dashed lines in Fig. 1, C–F). Colocalization was significant for each of the analyzed wild-type nuclei (n = 25) based on random simulations at all time points ($P < 0.02$; Table S1). Moreover, colocalizations per time point were also significantly higher than random overlaps (Welch's t test, $P < 0.001$ for 3, 4, and 5 h in SPM; Fig. 1 C and Table S1). In particular, Mer2 foci numbers showed a dynamic behavior, with fewer foci at time points 2 and 6 h in SPM (Fig. 1 H), contributing to the slightly lower colocalization at these time points. We next analyzed the colocalization of Mer2 with two Spp1 mutants whose function in bridging H3K4 and Mer2 is impaired: Spp1CxxCΔ (Fig. 1 D) lacks the CxxC zinc finger motif involved in Mer2-Spp1 interaction at chromosomal axis sites, while Spp1PHDΔ (Fig. 1 E) contains a truncation of the PHD finger domain that affects the binding of Spp1 to H3K4 trimethylated nucleosomes (Acquaviva et al., 2013b). Both mutants showed a decreased Spp1 focus number, but did not change Mer2 focus counts (Fig. 1, G and H). Although the number of potential Mer2 partner foci was unchanged, the proportion of colocalizing

Spp1 foci decreased significantly compared with wild-type cells (Welch's *t* test, $P \leq 0.02$ for 3, 4, 5, and 6 h in SPM; Fig. 1 D and E and Table S1). Colocalization reached a plateau only in 30% of both mutants, which was still higher than randomized colocalization (Welch's *t* test, $P < 0.01$ for 2, 3, 5 h for Spp1CxxCΔ; and $P < 0.05$ for 2, 3, 5 h in SPM for Spp1PHDΔ; Table S1). Colocalization of Mer2 with the Set1 subunit Bre2 that stably interacts with Set1C (see below) was lower or equal than assumed from a random distribution (Welch's *t* test, $P > 0.1$; not significant for all time points; Table S1); however, it was significantly lower than that of Spp1-Mer2 (Welch's *t* test, $P < 0.01$; Fig. 1 F and Table S1). Together, these data highlight spatial associations between Mer2 and Spp1 that are snot seen between Mer2 and the Set1C/Bre2 holocomplex. Wild-type level of association requires Spp1's Mer2 and H3K4me3 interaction motifs, suggesting that these domains critically enhance the lifespan of the observed interactions.

Spp1 exhibits static and dynamic chromosome binding kinetics during meiosis

To assess the chromatin dynamics of Spp1 during the progression of meiotic prophase, we mapped the chromosomal locations of epitope-tagged Spp1 and Bre2 by chromatin immunoprecipitation (ChIP) sequencing in synchronously sporulating yeast cultures (Fig. S1). The distribution of Bre2 was used as a proxy to mark the chromosomal position of Set1C. Peak sets identified at individual meiotic time points (pre-SPM [SPS]; 0, 2, 4, and 6 h in SPM) were concatenated and sorted by chromosomal position and then merged (all peaks and ChIP-seq profiles established in this study can be accessed via JBrowse, see Materials and methods for details). Venn diagram analysis of chromatin binding sites shows that ∼46% of the Spp1 peaks coincide with Bre2 (Fig. 2 A), which may represent a group of Spp1 molecules associated with Set1C during meiosis.

Overall, Spp1 and Bre2 (common) peaks and Bre2-only peaks show strong enrichment on ribosomal protein genes (RPGs), snoRNA/ncRNA genes and transcription start sites (TSSs), but they are absent from Mer2/Red1 axial sites (Fig. 2 B), as defined by genome-wide ChIP analysis (Panizza et al., 2011; Sun et al., 2015). In contrast, Spp1-only peaks are significantly overrepresented at Mer2/Red1 sites. Strikingly, Bre2-only peaks are highly enriched at RPG and tRNA genes compared with common peaks of Spp1 and Bre2, indicating the presence of Spp1-free Set1C on these genes during meiosis.

Importantly, Spp1 showed a progressive loading onto Mer2 binding sites during meiotic prophase, while Bre2 remained depleted throughout the sporulation process (Fig. 2 C). Although Spp1 binding sites appear to be more dynamic than common (Spp1 and Bre2) sites (representative JBrowse example for dynamic Spp1 peaks is shown in Fig. 2 D), the latter peaks show much higher ChIP signal compared with Spp1-only or Bre2-only sites (ANOVA with Tukey HSD, $P < 0.0001$; Fig. 2 E). We explain these differences with the differential turnover rate characteristics of the sites (see below).

To gain more mechanistic insights into the dynamics of Spp1, we performed k-means clustering analysis on the time-resolved Spp1 ChIP signals, classifying the identified binding sites based on their similarity (see Materials and methods). Two kinetic fractions were readily revealed based on the relative change of Spp1 peak signals over time (Fig. 2 F): dynamic sites, which gradually appeared (red) or disappeared (blue) as meiosis progressed, and static sites (green) showing permanent association with Spp1. These separate classes were reproduced by a clustering-independent approach that relied on the absolute change of Spp1 signal intensities in terms of time (Fig. 2 G; see legend for explanation).

Functional annotation revealed that (1) appearing Spp1 peaks are strongly enriched at chromosome axial sites (Red1 and Mer2), (2) disappearing Spp1 sites are enriched at RPG and snoRNA genes, and (3) constant Spp1 peaks show strong association with ncRNAs (Fig. 2 H). We conclude that the dynamic properties of Spp1 correlate with its noncanonical (Set1C independent) functions and the remodeling of Set1C at RPG and snoRNA genes during the meiotic process.

Functional analysis of Spp1 chromatin binding in meiosis

To further shed light on the molecular determinants of Spp1 chromatin binding, we also examined the binding sites of Spp1PHDΔ and Spp1CxxCΔ mutants and that of H3R2A and H3K4R mutants. Mutation of lysine 4 prevents H3K4 methylation, while substitution of arginine 2 by alanine inhibits the deposition of H3K4me3 (Kirmizis et al., 2007; Yuan et al., 2012). Both modifications are expected to phenocopy the meiotic phenotype of the Spp1PHDΔ mutation (Fig. 3 A). We performed time-resolved meiotic ChIP-seq and mapped the binding of Spp1PHDΔ, Spp1CxxCΔ, and Spp1 in H3R2A/H3K4R mutants (the data can be accessed through JBrowse; see Materials and methods). As shown in Venn diagrams (Fig. 3 B), all four mutations eliminate ∼50% of Spp1 binding sites during the meiotic time-course identified in the wild-type strain. Interestingly, some new Spp1 sites (∼10%) are also generated in each mutant (Fig. 3 B).

We next performed multidimensional scaling (MDS) analysis on the identified binding sites to highlight temporal and cell type–specific differences in Spp1 chromosomal localization (Fig. 3 C). For all cell types and meiotic time points, exact chromosomal position and enrichment of all the identified Spp1 ChIP peaks were assigned to N-dimensional coordinates, defining Spp1 "states" by cell type and meiotic stage. All Spp1 states were then projected to a 2D plane (highlighted as dots in the MDS maps, Fig. 3 C) such that the closer is the difference between any two datapoints the more similar the Spp1 states are. As shown in the upper panel of Fig. 3 C, wild-type cells and Spp1 PHD- and CxxC-domain mutants behave very differently at the beginning of sporulation. Then, in the first 2 h, there will be a large, rapid, and identical change in both wild-type and mutant cells. By the end of the process, each cell type converges to a similar Spp1 state, which is shown by the small distance of dots at the 6-h time point. In the histone mutant backgrounds (lower panel in Fig. 3 C), Spp1 binding sites are more similar to the wild type at the beginning of sporulation (0 h in SPM). Subsequently, fast and dynamic changes occur in the first few hours such that both mutants quickly move away from the wild type. By the end of the process all three cell types are characterized by a different Spp1 state.

We next analyzed the overlap of Spp1 binding sites with annotated functional genomic elements in each mutant (Spp1PHDΔ, Spp1CxxCΔ, H3R2A, and H3K4R). As shown in Fig. 3 D, the result-

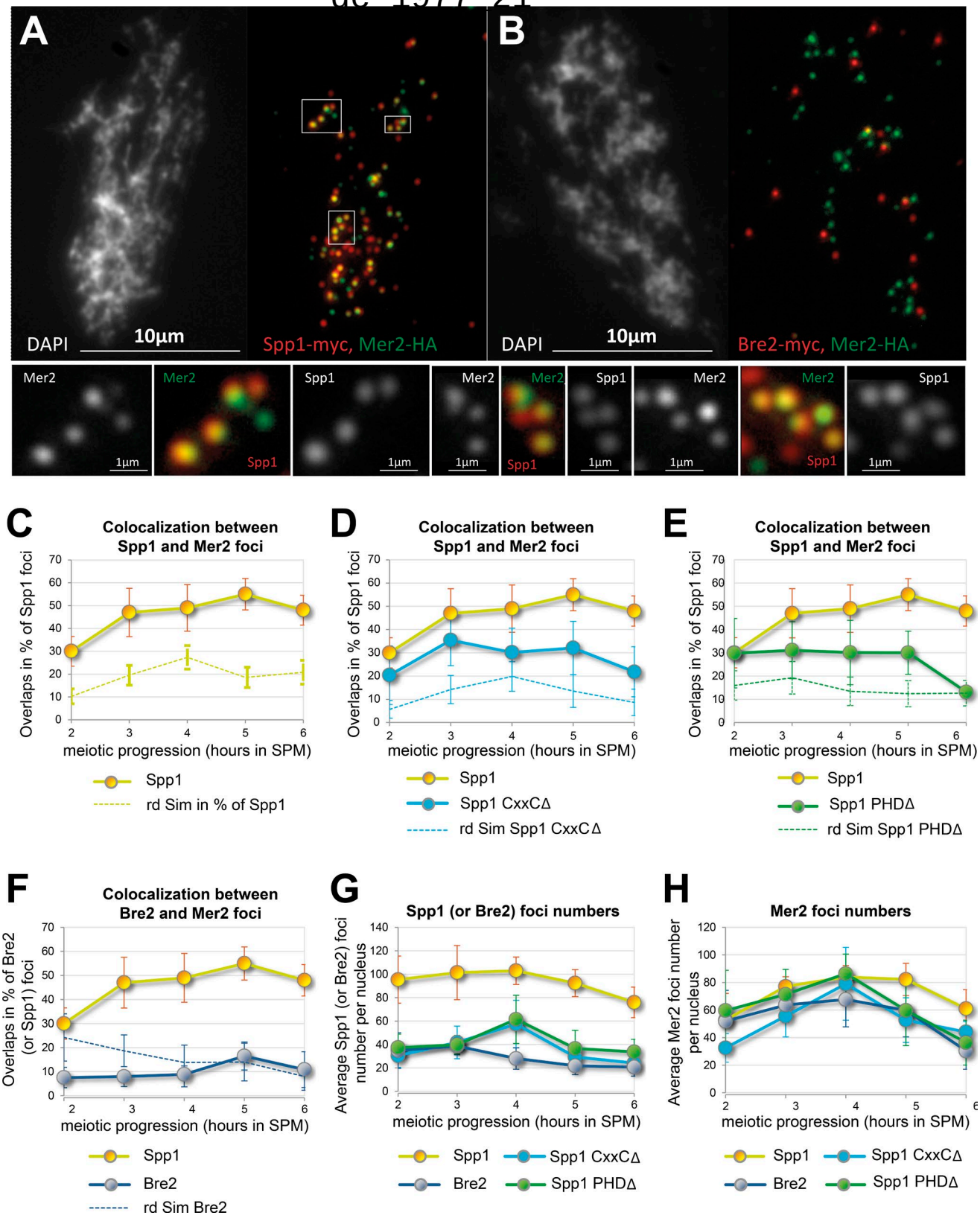


Figure 1. **Localization of Spp1-myc, Bre2-myc, and Mer2-HA in meiotic chromosome spreads and quantification of overlaps. (A)** Representative chromatin spread showing Spp1-myc foci (red) and Mer2-HA.int foci (green). Insets are also displayed at higher magnification in lower panels. **(B)** The same as A, but Bre2-myc (red) and Mer2-HA.int (green) foci are shown. Cells were taken from a meiotic time course at 4 h after transfer to SPM. DNA was stained by DAPI (white). **(C–E)** The graphs show the percentage of Spp1-myc foci overlapping with Mer2-HA.int in wild-type cells (yellow; C), and in Spp1CxxCΔ-myc (light blue; D), and Spp1PHDΔ-myc mutants (green; E). **(G)** Expressing overlap in percentage of Spp1 foci demonstrates changes in the overlap ratio between Spp1 and Mer2 independent of the reduction in Spp1 foci numbers seen in the mutants. **(F)** Percentage of Bre2-myc foci overlapping with Mer2-HA.int (gray). **(C–F)**

ing peaks are differentially enriched over several genomic elements and show variable overlap with each other (from 7 to 58%; Fig. S2 A). Importantly, all mutations reduce the binding of Spp1 to axis sites (Fig. 3 D) and abrogate the association of dynamic clusters of Spp1 peaks with Mer2 sites (Fig. 3 E). The PHDΔ mutant shows a very high enrichment of Spp1 at RPG genes, which highlights the role of the PHD domain in the removal of Spp1 from RPG genes (Fig. 3 D). Similarly, H3R2A and H3K4R mutants exhibit specific Spp1 enrichment at snoRNA genes, indicating that H3R2 and H3K4 methylation promotes the disappearance of Spp1 from snoRNAs.

The heat maps shown in Fig. 3 E reveal that enrichment of appearing Spp1 peaks at Mer2 sites is abolished in the Spp1CxxCΔ, H3R2A, and H3K4R mutants. Deleting the PHD finger domain of Spp1 eliminates ∼75% of appearing Spp1 peaks (264/1,021) detected in wild-type cells; however, about half of the remaining Spp1PHDΔ sites (130 peaks) still exhibit significant enrichment at Mer2 sites. This is in contrast to the Spp1CxxCΔ binding sites and the effects of H3R2A/K4R mutations that apparently prevent enrichment of Spp1 at Mer2 sites. For comparison, we also analyzed the association of the dynamic clusters of Bre2 binding sites defined by cluster analysis (similarly to Spp1 sites) with Mer2 sites. Clearly the appearing Bre2 binding sites are particularly low in Mer2 signal (Fig. 3 E, right panel).

Together, these results further strengthen the tethered loop axis model of meiotic DSB formation proposing that proper localization of Spp1 to chromosome axial sites requires (1) the Mer2-binding (CxxC) motif of Spp1; (2) to a lesser extent, the PHD finger domain; and (3) the presence of histone modifications and modifiable residues (H3K4me3 and H3R2me2s). Consistent with our ChIP-seq data in H3R2/K4 mutants (Fig. 3 E), Spp1 chromatin binding show a global decrease over Mer2 axial sites in *set1Δ* cells (Fig. S2 C; microarray dataset is from Adam et al., 2018), suggesting that the activity of Set1 is a prerequisite for the full binding of Spp1 to chromatin.

We next examined meiotic chromosome spreads on the colocalization of Spp1 and Mer2 in H3R2 and H3K4 mutants. Unexpectedly, microscopic inspection of Spp1-myc and Mer2-HA.int foci did not show a significant decrease in the number of colocalized spots for H3R2 and H3K4 mutants (Fig. S2 D), in contrast to the reduced ChIP enrichment of Spp1 at axial sites in the same mutants and in *set1Δ* cells (Fig. 3 and Fig. S2 C). This result indicates that the colocalization, reflecting an apparent interaction of Spp1 with Mer2, is apparently insufficient to promote stable chromatin binding of Spp1 to chromosome axial sites. In the absence of H3R2/K4 methylation, the affinity of Spp1 to chromatin (measured by ChIP enrichment) may decrease significantly and may thus reduce the association of Spp1 with chromatin axial sites. Therefore, we conclude that both Set1-mediated histone methylation and modifiable histone residues are important for the effective binding of Spp1 to the chromosome axis.

Dynamics of Spp1 is influenced by the kinetics of meiotic gene expression

Meiosis involves ∼70% of the genes showing a constant transcriptional level, but 30% are regulated up or down (Primig et al., 2000; Borde et al., 2009; Lardenois et al., 2011; Yamaguchi et al., 2012). H3K4me3 and Set1 occupancy have been shown to be increased at the 5′-end of coding regions and correlate with the level of transcription (Ng et al., 2003; Luciano et al., 2017). We therefore asked whether the differential dynamics of Spp1 and Bre2 correlate with the transcriptional level of flanking genes (data are from Brar et al., 2012). We classified protein coding genes into three categories that are associated with Spp1-only, Bre2-only, and common (Bre2 and Spp1) peaks (Fig. 4 A and Fig. S2 B). Based on the median mRNA expression levels, ORFs linked to common (Spp1 and Bre2) sites showed significantly higher transcription rate than Spp1-only and Bre2-only genes (ANOVA with Tukey HSD; $P < 0.0001$). Intensive transcription at common (Spp1 and Bre2) sites suggests a spatial correlation between the presence of the full Set1 complex over protein coding ORFs and increased mRNA expression levels. This conclusion is also supported by the fact that when we repeated the previous measurement with genes related to dynamic Spp1 binding sites (Fig. 4 B), the smallest and highest mRNA levels have been measured in the appearing and disappearing categories, respectively. It should be also noted that expression of Spp1-associated genes tightly follows the dynamics of Spp1 chromatin binding, since genes in the "appearing" class become rapidly up-regulated, while the "disappearing" class becomes down-regulated in the first few hours of meiosis.

Regarding the chromatin factors that have been implicated in meiotic DSB formation, appearing Spp1 sites show strong association with Mer2 binding sites (Fig. 4 C) with reduced H3K4me3 and Spo11-oligo levels (Fig. 4, D and E), whereas disappearing Spp1 sites are highly enriched in histone H3K4me3 (Fig. 4 D) and Spo11-oligo DSBs (Fig. 4 E) with decreased Mer2 levels (Fig. 4 C). Overall, these results indicate that de novo formed Spp1 binding sites (appearing class) reflect either activated transcription or linkage to axial regions, while loss of Spp1 (disappearing sites) is related to down-regulated/repressed or poised transcription (D'Urso et al., 2016).

Quantitative analysis of Spp1 chromatin binding: estimating effective turnover rates by competition ChIP (c-ChIP)

To quantify the binding characteristics of Spp1 in terms of turnover and residence time, we performed dynamic chromatin mapping using competition c-ChIP, which allowed estimation of turnover rates at Spp1 binding sites (Lickwar et al., 2013; c-ChIP profiles are available through JBrowse, see Materials and methods). We differentially tagged a constitutive and an inducible isoform of *SPP1* with 9xmyc and GFP epitopes, respectively (Fig. 5 A) and turned on the expression of the inducible allele

800–2,000 foci were assessed manually for overlap in a total of 10 spread nuclei per time point. Expected random overlaps are represented by dashed lines determined by Monte Carlo simulation (FociSim, see Materials and methods). FociSim mimics random focus distribution based on the assessed parameters: nucleus area, foci numbers, and foci sizes for each nucleus. **(G)** Average Spp1 focus numbers for the meiotic time courses shown in C–F. **(H)** The same as G, except that average Mer2 focus counts are shown. Error bars represent SD.

dc_1977_21

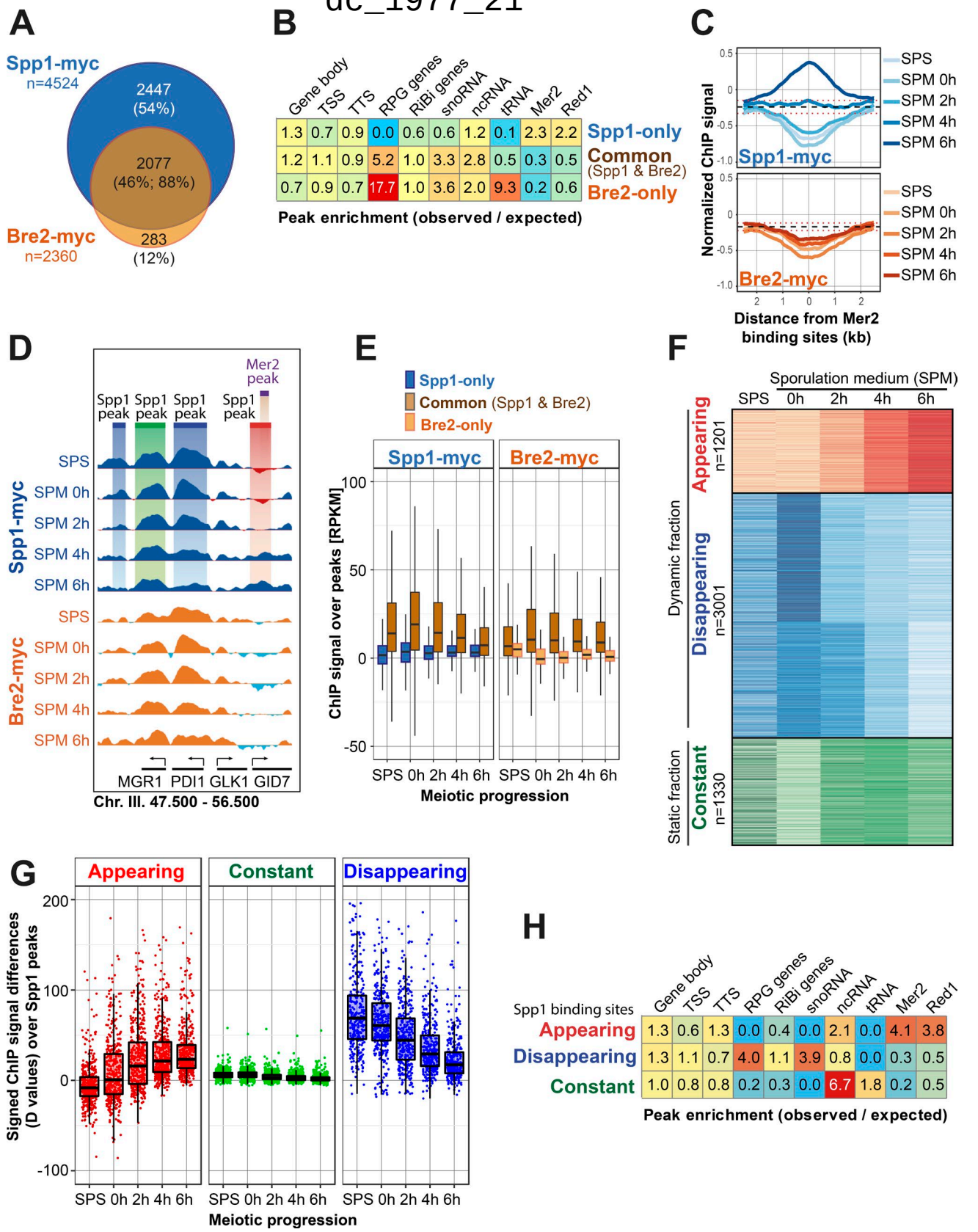


Figure 2. **Chromosomal distribution of Spp1 and Bre2 binding sites during meiotic prophase. (A)** Proportional Venn diagram showing the overlap of Spp1-myc and Bre2-myc binding sites identified in a meiotic time course (SPS; 0, 2, 4, and 6 h in SPM) by ChIP-seq. 54% of Spp1 peaks show no overlap with Bre2, while 88% of Bre2 peaks coincide with Spp1 binding sites. **(B)** Functional annotation of Spp1 and Bre2 sites show differential enrichment over several genomic regions. Spp1-only peaks are highly enriched at chromosome axial sites (Mer2, Red1); common peaks (Spp1 and Bre2) are associated with RPG, snoRNA, and ncRNA genes and depleted over Mer2/Red1 sites; and Bre2-only peaks are enriched at RPG, snoRNA, ncRNA, and tRNA genes and depleted over Mer2/Red1

(driven by a p*CUP1* promoter) with copper addition during the meiotic time course (Fig. S3 A). The level of GFP-Spp1 increased exponentially during the time course and reached its maximum after 6 h in SPM (Fig. 5 B and Fig. S3 B). By sampling dense meiotic time points (4.5, 5.0, 5.5, 6.0, and 6.5 h in SPM), both Spp1 isoforms were immunoprecipitated using anti-myc and anti-GFP antibodies, and binding sites were determined according to conventional ChIP-seq pipelines. The resulting Spp1-enriched sites were validated by polar plot comparison of c-ChIP and ChIP peak sets detected at various time points in meiosis (Fig. S3 C). The polar plot shows the highest similarity between the two datasets between 4 and 6.5 h in SPM (dots at the smallest distance from the center, surrounded by a red dashed line), consistent with the c-ChIP experimental design (Fig. S3 A). Then we calculated Spp1 turnover rates by determining the ratio of the GFP (new Spp1) and myc (old Spp1) signals and by fitting the data with an exponential model (Fig. S4 A). Our kinetic model resulted in 977 high confidence peaks that allowed the expression of Spp1 turnover rates as number of Spp1 replacements per unit time (1/min) per genomic region (Fig. S4 B). The analysis shows that Spp1-only sites exhibit different replacement dynamics compared with common (Spp1 and Bre2) binding sites (Fig. 5 C). Spp1-only sites move much slower than common (Set1C-associated) sites over gene bodies, transcription termination sites (TTSs), RPG/Ribi genes, and Mer2/Red1 sites ($P < 0.001$, ANOVA with Tukey HSD). Interestingly, the turnover rate of TSSs is no different between Spp1-only and common peaks ($P = 0.181$, not significant; ANOVA with Tukey HSD); however, in these genomic regions the highest turnover rates have been measured from all examined regions ($P < 0.0001$; ANOVA with Tukey HSD). Increased Spp1 mobility in promoter-proximal regions is consistent with the fast turnover of the nucleosomal substrate of Set1C (i.e., H3K4 trimethylated histones; Dion et al., 2007; Kraushaar et al., 2013).

Turnover rate/occupancy plots show that distribution of Spp1 turnover rates only slightly correlates with Spp1 ChIP enrichment (occupancy, $R^2 = 0.18$; Fig. 5 D). However, when Spp1 sites are grouped according to their kinetic behavior (disappearing, appearing, and constant fractions; Fig. 5 D–H), the disappearing and emerging Spp1 sites sharply stand apart based on the distribution of turnover rates (disappearing Spp1 sites tend to have higher turnover rates and higher occupancies compared with appearing sites, Fig. 5 D). Constant sites are associated with stochastically distributed values in terms of the above parameters, indicating that a combinatorial action of stably bound and transiently bound Spp1 molecules (present in few cells or large number of cells, respectively) could derive the same apparent ChIP occupancy level. Appearing Spp1 sites exhibit low turnover rates with low Bre2 occupancy (Fig. 5 E), low H3K4me3 (Fig. 5 F), and high Mer2 enrichment levels (Fig. 5 G), whereas disappearing Spp1 sites can be characterized by high turnover rates with high H3K4me3 and Bre2 levels (Fig. 5, E and F) and reduced Mer2 occupancies (Fig. 5 G). The rate of mRNA expression change of the genes associated with Spp1 negatively correlates with Spp1 turnover rate (Fig. 5 H) and sharply separates the kinetic classes of Spp1 binding sites.

From the observed trends, we conclude that (1) turnover rate and occupancy are two measurable properties of the chromatin binding dynamics of Spp1 that effectively discriminates between the different functional types of Spp1 binding sites, (2) differential turnover dynamics of Set1C/Bre2-associated Spp1 sites and Spp1-only sites indicate the presence of two separate Spp1 pools that are characterized by different kinetics and are distributed differently between the Set1 complex and meiotic DSB proteins, and (3) binding of Spp1 to Mer2/Red1 axial sites reduces the rate of Spp1 turnover.

Quantitative microscopic analysis of Spp1 chromatin binding by FRAP and FCS techniques

The c-ChIP approach has superior spatial resolution but its temporal resolving power is relatively low, so we further characterized the nuclear dynamics of Spp1 at an increased temporal resolution. Spp1 and Set1 were tagged with a GFP fluorescent reporter at their N-termini (GFP-Spp1 and GFP-Set1; Fig. 6 A) and we measured their mobility in live meiotic cells using fluorescence recovery after photobleaching (FRAP) and fluorescence correlation spectroscopy (FCS). Expression of the fluorescent proteins was controlled by the p*CUP1* promoter by adding 100 µM $CuSO_4$ into the SPM at 0 h. Cells selected for the measurements had a fluorescence intensity well above the background, but were not saturated, which allowed us to stay within the single molecular sensitivity range of FCS (Stasevich et al., 2010).

In the FRAP setting, whole live-cell nuclei were bleached, and fluorescence recovery was recorded in the first 5 h of meiosis (Fig. 6 B). GFP-Spp1 and GFP-Set1 reached saturation in <50 s

sites. Heat map shows the overlap ratio of observed and computer randomized binding sites (observed/expected) with the indicated annotation category. **(C)** Spp1-myc is progressively loaded to Mer2 binding sites during meiotic prophase, while the Bre2-myc signal remains depleted throughout the sporulation process. Horizontal dashed line and red dotted lines show the genome-wide average ChIP signal ± SD. **(D)** Representative genome browser snapshot showing the chromosomal distribution of Spp1-myc (blue) and Bre-myc (orange) ChIP signal. Tracks represent meiotic time points. Disappearing, constant, and appearing Spp1 peaks are highlighted in blue, green and red, respectively. A Mer2 site is also shown in purple. **(E)** Common (Spp1 & Bre2) binding sites show increased chromatin association compared with Spp1-only and Bre2-only sites (ANOVA with Tukey HSD; $P < 0.0001$). Box plots show the distribution of ChIP signals over the three categories (Spp1-only, common, and Bre2-only). Left, Spp1-myc enrichment. Right, Bre2-myc enrichment. **(F)** Temporal classes of Spp1 binding sites identified by cluster analysis. Appearing (red) and disappearing (blue) sites show dynamically increasing/decreasing ChIP enrichment, while constant sites (green) do not show significant temporal changes. Heat maps show the relative changes of ChIP enrichment over time (normalized by rows). **(G)** Confirming the kinetic classes of Spp1 binding sites by an independent approach, based on the absolute values of ChIP enrichments. Spp1 peaks were rank-ordered by their signed ChIP signal differences (D values) between 0 and 6 h in SPM. Sampling the bottom (<q20), middle (q40–q60) and top (>q80) quantiles of the D values recapitulated the dynamic classes of Spp1 sites visualized by cluster analysis (in panel F). **(H)** Functional annotation of the dynamic classes of Spp1 binding sites. Appearing Spp1 peaks are strongly enriched at chromosome axial sites (Red1 and Mer2). Disappearing Spp1 sites are enriched at RPG and snoRNA genes and depleted at Mer2/Red1 sites. Constant Spp1 peaks show strong association with ncRNAs and depletion over Mer2 binding sites. The data are representative of two independent biological replicate experiments. Sample size (n, number of peaks analyzed in each category) is indicated in panels A and F.

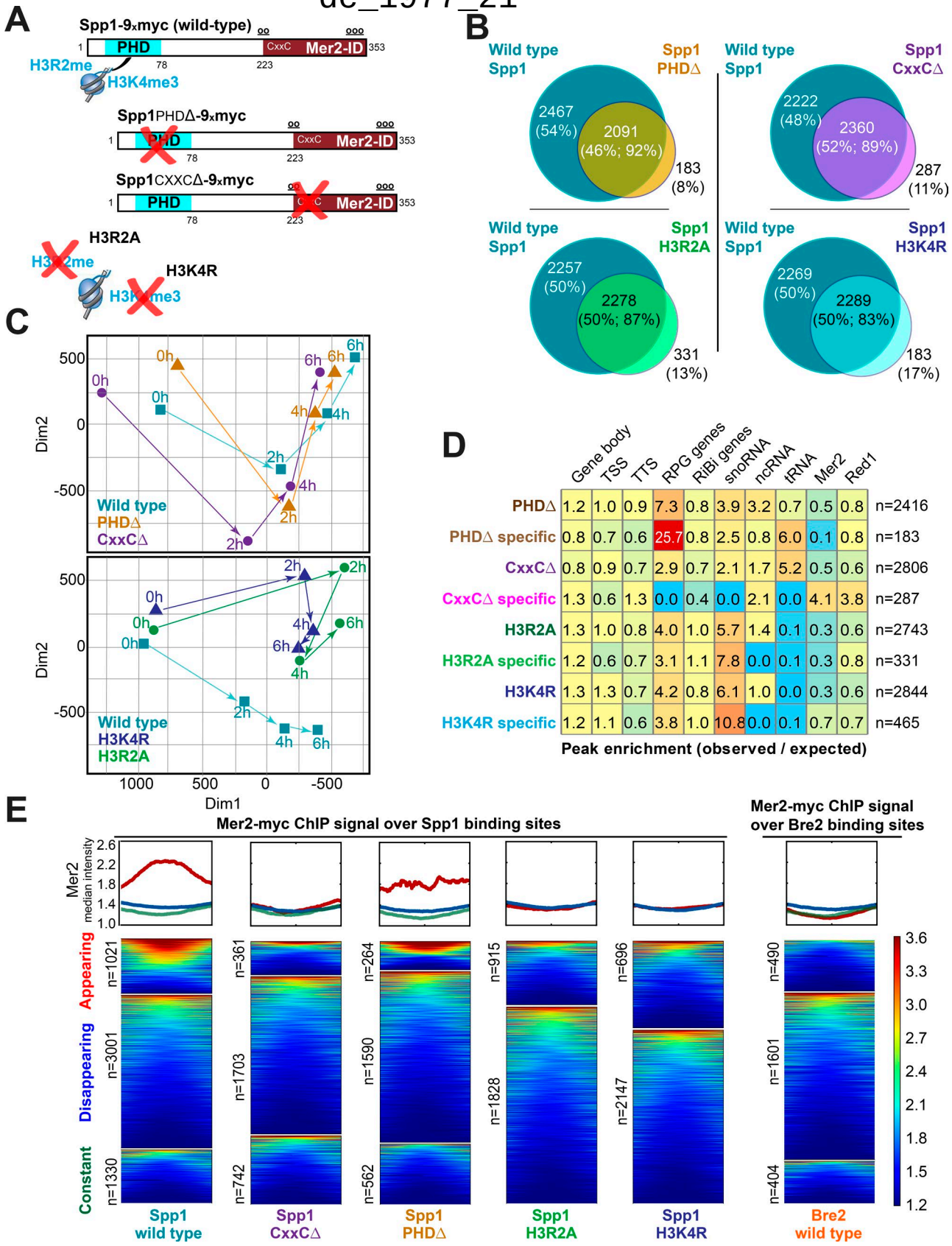


Figure 3. **Functional analysis of Spp1 chromosomal binding in meiosis. (A)** Schematic structure of the Spp1 mutant proteins studied in ChIP-seq experiments. C-terminal tags are not illustrated. Mutations (highlighted by red Xs) were introduced into the PHD finger domain (blue) and Mer2-interacting motif (brown) of Spp1 and into histone H3R2 and H3K4 (H3R2A and H3K4R). **(B)** Proportional Venn diagram showing the overlap of Spp1 binding sites identified in wild-type and mutant cells during the meiotic time course (0–6 h in SPM). Reduction of Spp1 binding sites for each mutation is indicated on the diagrams.

during recovery, which demonstrates the highly dynamic properties of the mobile fraction of the two molecules. About half of the FRAP signal did not return after the initial bleach pulse, indicating that ∼50% of Spp1 and Set1 remain tightly bound to chromatin representing the immobile fraction. The mobile fractions were further characterized by FCS between 0 and 6 h in SPM, allowing us to track the diffusion of one to five molecules at a time (on the scale of milliseconds) within a femtoliter-sized confocal volume. The time-dependent autocorrelation functions were fitted with a 3D normal diffusion model supposing two autonomous diffusing components (ρ_1 and ρ_2; Fig. 6 C), deriving several physicochemical parameters that described the mobility of Set1 and Spp1 (e.g., number of diffusing molecules, diffusion time, diffusion coefficient, and apparent molecular mass). The distribution of fast and slow components did not differ between Spp1 and Set1 (Fig. 6 D, left panel); however, the average diffusion coefficient (D_2) of Spp1 was significantly slower compared with Set1 (Fig. 6 D, middle panel). The changed diffusion coefficient shows the decreased nuclear mobility of Spp1. The average apparent molecular mass of GFP-Set1 (calculated from the Stokes-Einstein equation for spherical objects; Brazda et al., 2014; Hetey et al., 2017) was equal to the expected molecular mass of Set1C (379 kD), while GFP-Spp1 gave ∼43-fold higher molecular mass (1,764 kD) than the real molecular mass of the fusion protein, which suggests that Spp1 is attached to a huge macromolecular complex that is different from Set1C (Fig. 6 D, right panel). The large difference between the expected and observed molecular weights of Spp1 cannot be justified by nucleoplasmic interactions with diffusible protein factors alone; instead, transient chromatin associations (e.g., tethering to axial sites) may better explain the differential diffusion behavior of Spp1.

Discussion

Previously, we proposed a model in which the interaction between Spp1 and Mer2 links potential meiotic DSB sites to the chromosome axis, thereby enabling axis-proximal regions that are depleted in nucleosomes to be cut by Spo11 (Acquaviva et al., 2013a,b). However, it remained unknown whether Spp1 was still related to Set1C during this process and whether a specific subpopulation of Spp1 was relocalized from transcribed genes to chromosome axial sites. In addition, the spatial and temporal dynamics of Spp1 redistribution have not been studied so far.

The results presented in this study clearly show that in meiosis Spp1 behaves differently from Set1C/Bre2. This conclusion is based on the differential chromosomal localization of Spp1 and Bre2, the differential turnover rate dynamics of Spp1 and Set1 binding sites, and the different apparent molecular mass and diffusion coefficient of Spp1 and Set1. Unlike Set1C, the chromatin binding of Spp1 is very dynamic that is characterized by the creation of many new binding sites and the disappearance of many existing binding sites during the first hours of meiotic kinetics. We have shown that dynamic Spp1 sites reflect two populations that bind to chromosome axial sites or highly transcribed genes, respectively. Interestingly, the turnover rate of newly emerging Spp1 binding sites is low, suggesting that prolonged binding of Spp1 to Mer2 might be a prerequisite for DSB selection and formation. Disappearing Spp1 sites were associated with down-regulated genes, indicating that Spp1 might be released from repressed or poised genes similarly to transcriptional memory genes (D'Urso et al., 2016). Disappearing Spp1 binding sites exhibit high turnover rate characteristics and low association with Mer2 binding sites. Interestingly, disappearing Spp1 peaks were predominant at RPG and snoRNA genes that are transiently repressed in the first hours after transfer to SPM (Brar et al., 2012). The mechanism that triggers the dissociation of Spp1 from genomic sites to which Spp1 is tightly bound in rich media needs to be elucidated. Interestingly, the strong association between constant Spp1 peaks and ncRNAs may reflect an unexplored role of Spp1 in regulating noncoding RNA expression. A refined loop axis model (Fig. 7) shows how the dynamic behavior of Spp1 is associated with the different classes of Spp1 binding sites.

Together, our results are consistent with earlier works and further reinforce the tethered loop axis model in the framework of meiotic chromatin structure (Acquaviva et al., 2013b; Sommermeyer et al., 2013; Adam et al., 2018). We demonstrated in our previous work that H3K4me3 is required for Spp1 function, probably by recognizing the PHD-domain of Spp1, and this requirement could be bypassed by artificially tethering Spp1 to a DNA locus. We also reported that the CxxC motif of Spp1 is important for Mer2-Spp1 binding when Spp1 was artificially linked to the $GAL10_{UAS}$ region (Acquaviva et al., 2013b). Our new immunofluorescence experiments with spread meiotic chromosomes clearly show that both Spp1PHDΔ and Spp1CxxCΔ mutations affect the colocalization of Spp1 with Mer2, which usually reaches its maximum by the time of DSBs formation. Therefore, the PHD domain that is located outside the canonical Mer2 domain of Spp1 also contributes to the colocalization of Spp1 and Mer2. This new result is consistent with the genome-wide analysis of Spp1 chromatin binding sites during meiosis, demonstrating

About 90% of Spp1 peaks observed in the mutants overlapped with wild-type Spp1 sites. About 10% of Spp1 peaks formed de novo in the mutants. **(C)** MDS plots visualizing the similarities and differences of Spp1 binding sites identified in wild-type and mutant cells during the meiotic time course (0–6 h in SPM). Each datapoint represents a characteristic Spp1 state specified by cell type and temporal stage in meiosis. Distance of any two datapoints in the MDS map is proportional to the variability of Spp1 states (i.e., Spp1 peak sets). The upper map compares wild-type, Spp1PHDΔ, and Spp1CxxCΔ cells at four meiotic time points (0, 2, 4, and 6 h in SPM). The lower map depicts wild-type, Spp1 H3R2A, and Spp1 H3K4R cells at the same time points. **(D)** Functional annotation of Spp1 binding sites identified in the mutants. Color scale indicates enrichment or depletion within the annotation category. **(E)** Analysis of Mer2 enrichment over the dynamic classes of Spp1 binding sites identified by cluster analysis. Left, Mer2-myc signal enrichment shown on metaplots, centered to Spp1 peak positions identified in wild-type and mutant cells. In wild-type cells, the appearing class of Spp1 binding sites show strong enrichment in Mer2. Dynamic Spp1 clusters are also revealed in the mutants by cluster analysis, however, none of these dynamic sites are associated with Mer2. Right, Mer2-myc signal enrichment over clustered Bre2 chromatin binding sites. The data are representative of two independent biological replicate experiments. Sample size (n, number of peaks analyzed in each category) is indicated in panels B, D, and E.

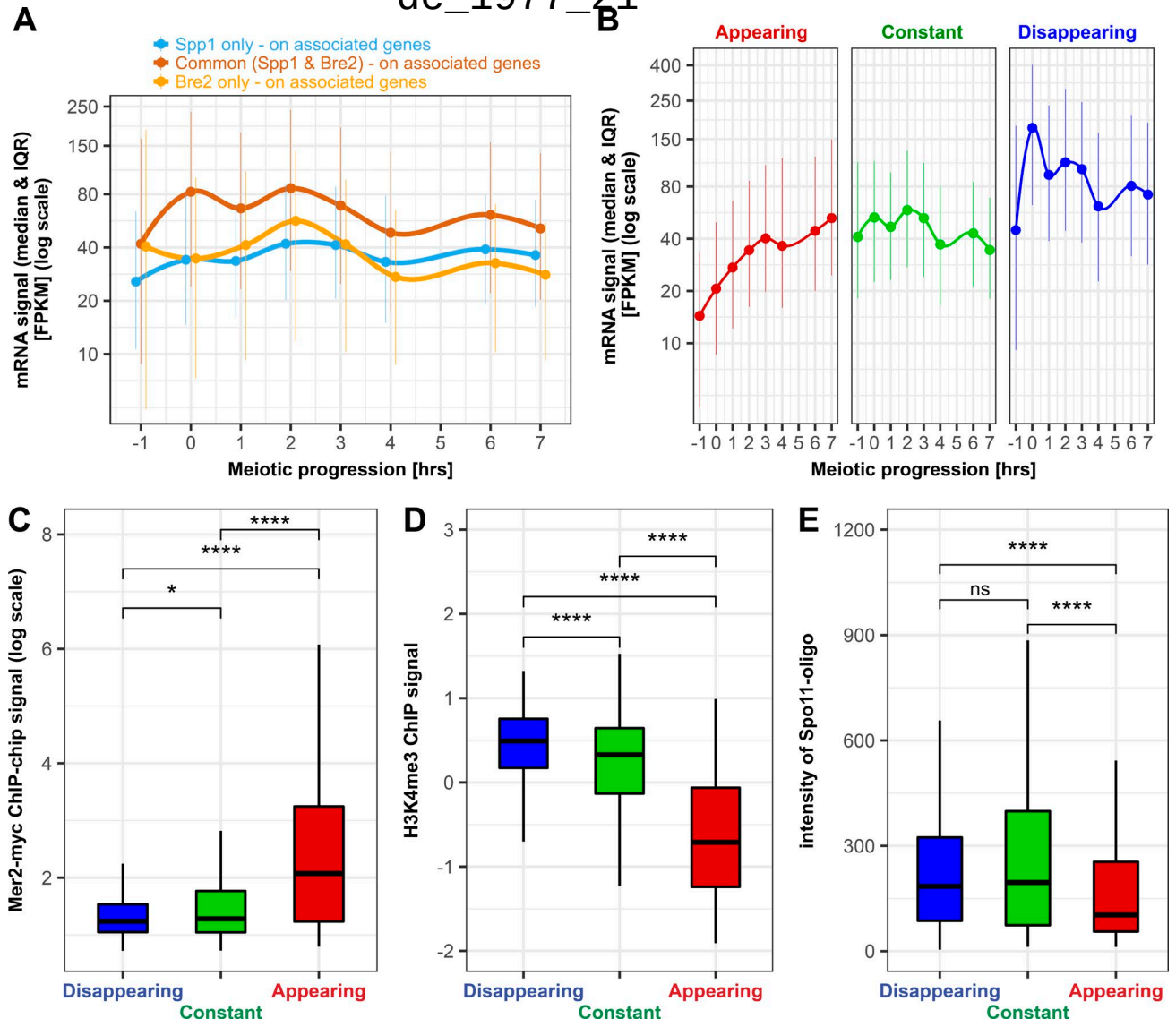


Figure 4. **Spp1 chromatin binding dynamics is influenced by meiotic gene expression. (A)** Spp1-myc and Bre2-myc peaks were assigned to the closest protein coding genes (Fig. S2 B) for which meiotic mRNA levels were determined (data are from Brar et al., 2012) in the three categories of binding sites (Spp1-only, common, and Bre2-only). Median gene expression levels are plotted during the progress of meiotic prophase. Genes associated with both Spp1 and Bre2 show significantly higher expression levels than Spp1-only and Bre2-only genes (ANOVA with Tukey HSD; $P < 0.0001$). The data are representative of two independent biological replicate experiments. Sample sizes (n): Spp1-only genes (3,479), common genes (1,818), and Bre2-only genes (1,300). **(B)** The mRNA expression level of Spp1-associated genes follows the dynamics of Spp1 chromatin binding. Spp1-myc peaks were assigned to the closest protein coding ORFs, and meiotic mRNA levels were determined (similarly to A). Median mRNA signals (with interquartile ranges) are plotted as a function of meiotic time. Dynamic Spp1 clusters (appearing, constant, and disappearing) that are associated with the flanking genes are highlighted in red, green, and blue. The data are representative of two independent biological replicate experiments. Sample sizes (n, number of associated genes): appearing class (1,043), constant class (1,119), and disappearing class (3,135). **(C–E)** Dynamic classes of Spp1 binding sites show differential enrichment in Mer2 binding (left, data from Panizza et al., 2011), H3K4me3 (middle, data from Borde et al., 2009), and Spo11-oligos (right, data from Mohibullah and Keeney, 2017). Box-whiskers plots show the medians (with interquartile ranges) of Mer2/H3K4me3/Spo11-oligo ChIP signals over Spp1 sites. Statistically significant difference is indicated (Mann-Whitney *U* test; *, $P < 0.05$; ****, $P < 0.0001$; ns, not significant). The data are representative of two independent biological replicate experiments. Sample size (n; number of peaks analyzed in each category): appearing class (464), disappearing class (464), constant class (463).

that Mer2 enrichment in the Spp1CxxCΔ mutant is prevented over newly formed Spp1 peaks and is strongly reduced in the Spp1PHDΔ mutant. These functional data point toward the importance of the PHD and CxxC motifs for the relocation of Spp1. Interestingly, when the H3R2 and H3K4 side chains were mutated to H3R2A and H3K4R, binding of Spp1 to axial sites was compromised, while Spp1 was still able to colocalize with Mer2. This result indicates that (microscopic) colocalization of Spp1 and Mer2 is not sufficient to promote stable binding of Spp1 to chromatin axial sites and that the PHD domain of Spp1 contributes to the interaction of Spp1 and Mer2 in addition to mediating association with methylated H3R2 and H3K4. One possible interpretation of these results is that the enrichment of Spp1 at Mer2 chromatin binding sites requires both protein–protein

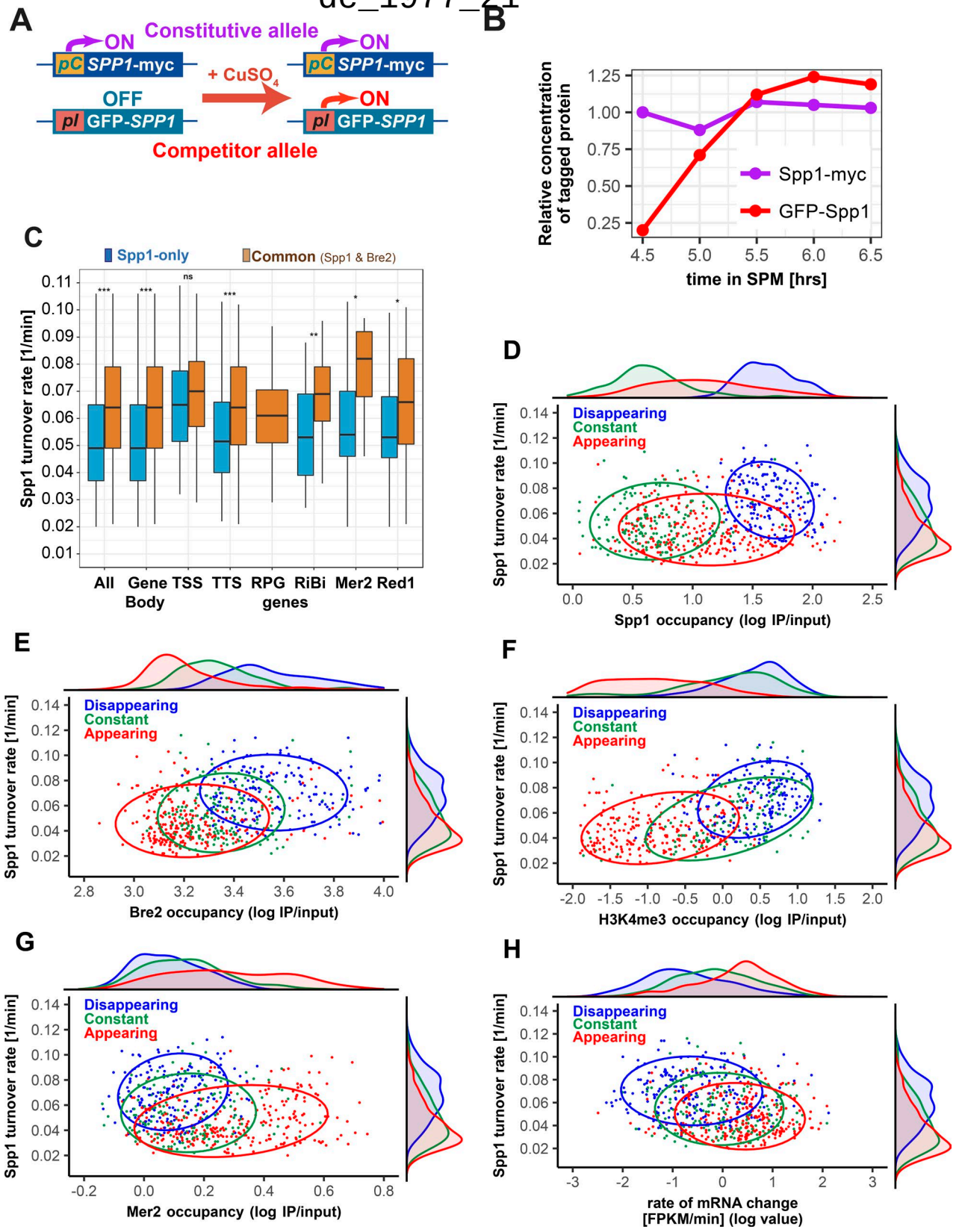


Figure 5. **Competition ChIP unravels turnover rates of Spp1 chromatin binding. (A)** Scheme of the c-ChIP yeast strain. Differentially tagged Spp1 isoforms are expressed from allelic positions in a diploid cell. The constitutive allele (Spp1-myc) is driven by an endogenous *SPP1* promoter while the inducible allele (GFP-Spp1) is controlled by a copper-inducible (p*CUP1*) promoter. Expression of GFP-Spp1 is induced by addition of 200 µM $CuSO_4$. **(B)** Relative protein levels of induced GFP-Spp1 and constitutive Spp1-myc as a function of time. Copper induction was initiated at 4.5 h in SPM and cells were collected in every 30 min until 6.5 h in SPM to perform Western blot and c-ChIP analyses. The data were obtained from two biological replicate experiments. The graph is representative of the quantitative Western blot measurement shown in Fig. S3 B. **(C)** Turnover rate of Spp1 determined at functional elements of the yeast genome. Spp1-only

interactions between Spp1 and Mer2 and former association of Spp1 with methylated H3R2/K4. The question remains open as to whether H3R2 is directly involved in the recognition of the Spp1 PHD motif. It is noteworthy that half of the remaining appearing peaks of the Spp1PHDΔ mutant still overlap Mer2 binding sites, indicating that the Spp1 PHD domain is not an absolute prerequisite. Notwithstanding, Set1-mediated H3K4 methylation might be important for the relocation and Mer2-association of Spp1 (from loops to axes). It should be also noted that loop-tethering to axial sites may be still possible through diffusion driven (random) processes irrespective of H3K4 methylation and Set1; however, this must be far less effective than the general mechanism that is facilitated by Set1, H3K4me3, modifiable histone residues, and to a lesser extent, the H3K4me3-reader PHD finger motif of Spp1.

We previously demonstrated that the Spp1PHDΔ and Spp1CxxCΔ mutations result in decreased DSB formation at the *BUD23* and *CYS3* recombination hot spots (Acquaviva et al., 2013b). Therefore, we propose that local changes in Spp1 turnover status is determinative for downstream biochemical events governing DNA break and crossover formation. Further studies will be needed to understand the functional implications of Spp1 turnover rate on the distribution of meiotic recombination initiation events. It will be particularly interesting to clarify whether higher Spp1 residence times stimulate the loop-tethering process and thus increase the likelihood of Spo11-mediated cleavage at recombination hot spots.

Materials and methods

Yeast strains

All yeast strains are from the SK1 background and are summarized in Table S1. For sporulation, cells were grown in rich medium (YPD) for 24 h, then transferred to SPS and grown overnight to a density of ∼4 × 10^7 cells/ml. Cultures were harvested by centrifugation, washed with one volume of prewarmed 1% potassium acetate and resuspended in SPM (1% potassium acetate supplemented with amino acids and nucleotides according to auxotrophic requirements, and 0.0001% of polypropylene glycol 2000 as an anti-clumping agent) at a density of 2 × 10^7 cells/ml, at 30°C. Meiotic progression and sporulation efficiency was monitored by FACS and fluorescent microscopy of DAPI-stained nuclei. Aberrantly slow or asynchronous sporulation time courses were excluded from further experiments. Spore viability was assessed by tetrad dissection, and it was greater or equal to 90% for all the strains involved in this study.

Chromatin spreads and colocalization statistics

Yeast chromosome spreads were prepared as described (Xaver et al., 2013). Spread nuclei were stained by anti-myc 9E10 mouse antibody (1:50), followed by anti-mouse cy3-conjugated secondary antibody (Jackson ImmunoResearch, 1:400) for Spp1-myc or Bre2-myc, and with rabbit anti-HA antibody (Sigma, 1:100) followed by anti-rabbit FITC-conjugated antibody (Sigma, 1:500) for Mer2-HA. Immunostained, fixed chromosome spreads were analyzed on a ZEISS AXIO Imager M2, with a ZEISS Plan-Neofluar 100×, aperture: 1.3, and a 2× additional magnification by a Zeiss optovar. Specimens were mounted in Vectashield with 0.2 µg/ml DAPI, and well-spread nuclei were selected based on their DNA-morphology (DAPI). Images were taken at a constant exposure time of 2 s for DAPI (BFP channel), CY3 (CY3 channel), and FITC (FITC channel). Light source: Sola SM II (Lumencor); camera: CoolSNAP HQ2 (Visitron Systems GmbH); acquisition software: Visiview (Visitron Systems GmbH). From these records, the nuclear area, foci numbers and foci areas were determined using Fiji software. Signals were counted as overlaps if foci overlapped by >60% of their diameter.

Pictures were optimized for display using linear operations on complete images

To assess the significance of number of Spp1-Mer2 foci overlaps, we performed Monte Carlo simulations (Kurzbauer et al., 2012) of the same foci counts, as well as the same nuclear area and representative foci sizes as measured. Simulated foci were circles which were fully contained within the nuclear ellipse, and overlaps between two foci were counted whenever the pairwise Euclidian distance (d) between their midpoints were smaller than the sum of their radii; i.e., d(m1, m2) < r1 + r2, where m1 and m2 are the midpoints of focus 1 and 2, and r1 and r2 represent the radii. Each nucleus simulation was repeated 200,000× for assessment of mean expectations and quantiles with high precision. For example, if the experimental overlap was greater than the 0.99 percentile of the simulation, the overlap was judged significant at level P = 0.01. Simulated random overlaps were added to each panel shown in Fig. 1 (C–F). For pairwise comparisons between mean values, Welch's *t* test (R, v3.4.3, stats package) was used to calculate significance. Colocalization is presented as overlaps/(all Spp1 foci). This representation can demonstrate changes in the overlap between Spp1 and Mer2 independent of Spp1 foci reduction seen in CxxCΔ, PHDΔ.

FRAP

FRAP measurements were performed in sporulating yeast cells (between 0 and 6 h in SPM) using an Olympus FluoView 1000

sites are shown is blue, while common sites (Spp1 and Bre2) are shown in light brown. Turnover of Spp1-only sites is significantly slower compared with common sites (ANOVA with Tukey HSD; *, p ≤ 0.05; **, p ≤ 0.01, ***, p ≤ 0.001; ns, not signficant). Spp1-only peaks were not detected over RPG genes; therefore, turnover rate estimation is missing for this category. **(D–G)** Turnover rate/occupancy plots showing the binding dynamics of Spp1 over Spp1 peaks (D), Bre2 peaks (E), H3K4me3-enriched regions (F), and Mer2 binding sites (G). Disappearing, appearing, and constant kinetic classes are highlighted in blue, red, and green. y axis: Spp1 turnover rate; x axis: occupancy (ChIP enrichment) of Spp1, Bre2, H3K4me3, and Mer2 sites, respectively. Circles comprise the confidence interval (q5–q95) of pointscatter distributions. The measured parameters are also highlighted as histograms (on the top of and right side of scatter plots), showing the distribution of turnover rates and occupancies, respectively. **(H)** Relationship of Spp1 turnover rate and meiotic gene expression rate. Spp1 peaks were assigned to the closest protein coding ORFs and meiotic mRNA levels were determined for each gene (similarly to Fig. 4 B). Spp1 turnover rates (y axis) were plotted in terms of the rate of mRNA change (x axis). Change of transcription rates were computed from the slope of mRNA expression curves (Fig. 4 B) fitted with the least squares method. The data were obtained from two biological replicate experiments. Sample size (n): appearing (286), disappearing (192), and constant (193).

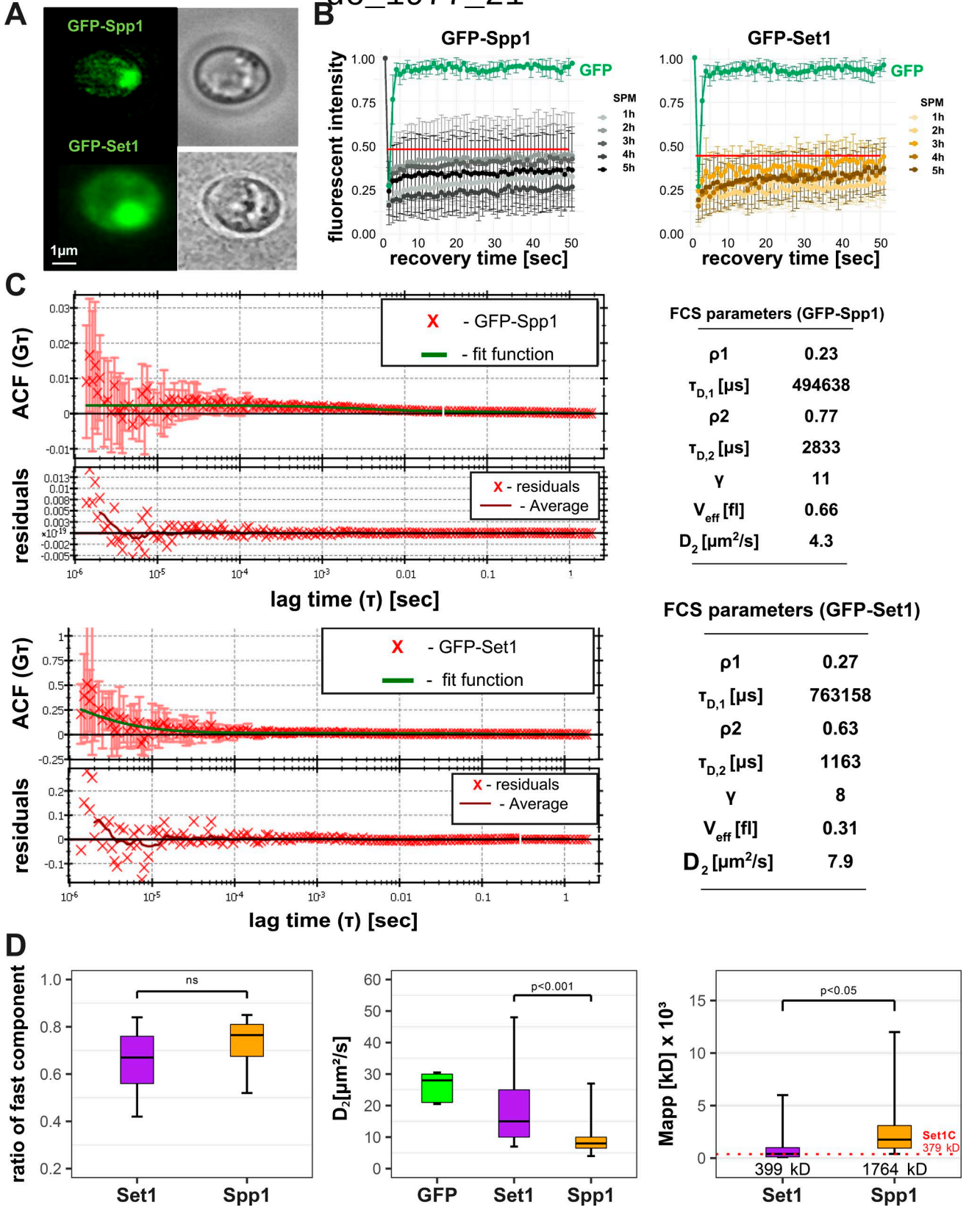


Figure 6. **Quantitative microscopic analysis of Spp1 chromatin binding dynamics based on photobleaching and fluorescence fluctuation. (A)** N-terminally tagged fluorescent Spp1 and Set1 proteins were induced in sporulating yeast cells (between 0 and 5 h in SPM) by adding 100 µM $CuSO_4$. Expression was driven by a p*CUP1* promoter. Both proteins were subjected to FRAP and FCS analyses. **(B)** FRAP curves show the retrieval of GFP-Spp1 (gray), GFP-Set1 (yellow), and GFP-only (green) signals at various meiotic time points. Recoveries reach a plateau phase within 50 s. Horizontal red line indicates the mobile fractions. The FRAP data are representative of two independent biological replicate experiments. Sample size (n) was 30–50 cells per meiotic time point. For each cell, five prebleach images were taken followed by a 500 ms bleach period of 100% laser intensity; then, post-bleach images were recorded in every

confocal microscope, based on an inverted IX-81 stand with an UPlanAPO 60× (NA 1.2) water immersion objective. Samples were taken at every hour from standard liquid sporulation cultures (1% potassium-acetate; imaging medium) and measurements were performed at room temperature (22°C) on microscope slides covered with 1% potassium acetate pad. GFP was excited by the 488-nm argon-ion laser line, and fluorescence was detected through a 500–550-nm band-pass filter. Cells expressing the GFP-Set1 or GFP-Spp1 proteins were randomly selected after $CuSO_4$ induction (100 µM) and five prebleach images were taken (256 × 256–pixel area, 15× zoom, and ∼9-µW laser power at the objective) followed by a 500-ms bleach period of 100% laser power (900 µW). Images were taken in every second up to 1 min. Image acquisition was performed by the FV10-ASW v1.5 software. Subsequently, the Fiji software was used to select rectangular areas for bleach regions of interest and to quantitate fluorescence recovery.

FCS

FCS measurements were all performed at room temperature (22°C) using an Olympus FluoView 1000 confocal microscope. Sporulating cells from taken at every hour from liquid sporulation (1% potassium acetate; imaging medium) cultures and FCS was performed on microscope slides covered with 1% potassium acetate pad. Fluorescence fluctuations were detected by avalanche photodiodes (Perkin-Elmer) and autocorrelation curves were calculated by an ALV-5000E correlation card (ALV Laser) at three randomly selected points of each nuclei, with 10 × 8 s runs. For FCS data processing and autocorrelation curve fitting the QuickFit 3.0 software was used (Krieger and Langowski, 2015) applying a 3D normal diffusion model for two-component fitting:

$$G(\tau) = \frac{1}{N}\left[\rho_1\left(1+\frac{\tau}{\tau_1}\right)^{-1}\left(1+\frac{\tau}{\gamma^2\tau_1}\right)^{-\frac{1}{2}} + \rho_2\left(1+\frac{\tau}{\tau_2}\right)^{-1}\left(1+\frac{\tau}{\gamma^2\tau_2}\right)^{-\frac{1}{2}}\right]$$

where τ is the lag time, τ_{tr} is the triplet correlation time, τ_1 and τ_2 are the diffusion times of the fast and slow species, ρ_1 and $\rho_2 = 1 - \rho_1$ are the fractional amplitudes of the two components, N is the average number of molecules in the detection volume, and γ is the aspect ratio of the ellipsoidal detection volume. Autocorrelation curves distorted by aggregates floating through the focus were excluded from the analysis.

Western blot

Whole cell extracts were prepared from 5 ml cells (4 × 10^7 cells/ml) sampled from the meiotic time course. Cells were lysed in 150 µl of lysis buffer (1.85 M NaOH, 7.5% β-mercaptoethanol at natural pH) and incubated for 10 min on ice. 150 µl of cold 50% trichloroacetic acid was added and incubated for a further 10 min on ice. The precipitated proteins were pelleted, resuspended in 100 µl of 2× Laemmli buffer, and neutralized by adding 5–10 µl of unbuffered 1 M Tris. Separation and blotting were performed according to standard procedures. Spp1-myc and Spp1-GFP proteins were detected using the 9E10 mouse anti-myc (ab56; Abcam) and anti-GFP (ab290; Abcam) primary antibodies and HRP-conjugated goat anti-mouse (1:100,000; Amersham) secondary antibody. Detection was performed by an ECL prime detection system (GE Healthcare).

ChIP and c-ChIP experiments

50 ml of meiotic cells (4 × 10^7 cells/ml) were collected at the indicated time points and cross-linked with 1% formaldehyde for 20 min at room temperature. Formaldehyde was quenched with 125 mM glycine for 5 min at room temperature, and cells were washed three times with ice-cold 1× TBS at pH 7.5 (20 mM Tris-HCl at pH 7.5 and 150 mM NaCl). Cells were resuspended in 500 µl of lysis buffer (50 mM Hepes KOH at pH 7.5, 140 mM NaCl, 1mM EDTA, 1% Triton X-100, 0.1% Na-deoxycholate, and 1 tablet of complete inhibitor cocktail [Roche] in 50 ml solution) and lysed with acid-washed glass beads for 10 min in a FastPrep bead beater machine. Chromatin samples was fragmented to an average size of 300 bp by sonication (Bioruptor; Diagenode). To obtain whole-cell extract, a 50 µl pre-immunoprecipitation (IP) sample was removed and centrifuged at full speed for 10 s to pellet the cell debris (supernatant = whole-cell extract). The rest of the samples were also centrifuged at 12,000 rpm (4°C) for 20 s to pellet the cell debris. IP was performed by adding the 450-µl extract to a pellet of magnetic protein G dynabeads (Dynal), corresponding to 50 µl or 2 × 10^7 beads, which were preincubated with the 9E11 (monoclonal mouse anti-myc, ab56; Abcam) or anti-GFP (polyclonal rabbit, ab290; Abcam) antibodies overnight at 4°C. IP samples were washed twice with lysis buffer, twice with lysis buffer plus 360 mM NaCl, twice with washing buffer (10 mM Tris-HCl at pH 8.0, 250 mM LiCl, 0.5% NP-40, 0.5% Na-deoxycholate, and 1 mM EDTA), and finally once with 1× TE at pH 7.5, using the magnetic device supplied by Dynal. After reversal of cross-linking by heating in TE-1% SDS overnight at 65°C, the proteins were digested with proteinase K (12 µl of 20 mg/ml stock) for 3h at 65°C. Nucleic acids were PCR clean up kit purified and

second up to 1 min. **(C)** Measuring the diffusional properties of GFP-Spp1 and GFP-Set1 by FCS. Representative time-dependent autocorrelation curves with the estimated diffusion parameters and residuals are shown. FCS curves were fitted with a two-component 3D normal diffusion model. ρ1: slow component; ρ2: fast component; $\tau_{D,1}$: diffusion time of slow component; $\tau_{D,2}$: diffusion time of fast component; γ: structure factor of the confocal volume; V_{eff}: effective measurement volume; D_2: diffusion coefficient of the rapid component in µm²/s unit. **(D)** Evaluation of the recorded autocorrelation curves show significant differences between GFP-Spp1 and GFP-Set1 diffusion coefficients (D) and apparent molecular mass (M_{app}) with similar fast component distribution. Left, average fraction of fast components after two-component fitting. ns, no statistically significant difference between GFP-Spp1 and GFP-Set1. Middle, distribution of diffusion coefficients. Significant difference is indicated (Mann-Whitney *U* test). Right, distribution of apparent molecular masses (M_{app}) of the mobile complexes comprising GFP-Spp1 and GFP-Set1. Numbers show median M_{app} values (kD). Significant difference is indicated (Mann-Whitney *U* test). Molecular mass of the full Set1 complex (red dotted line) is also shown on the right side of the cartoon. The FCS data are representative of two independent biological replicate experiments. Sample size (n) represents 10–40 randomly selected cells.

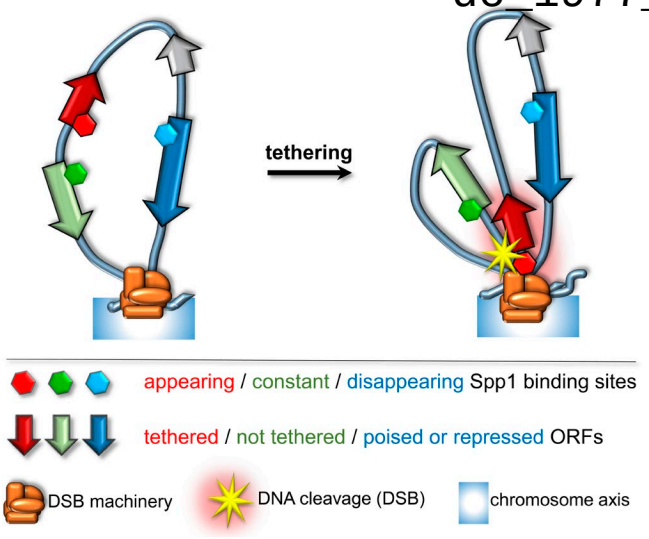


Figure 7. **Loop-axis model showing the dynamic behavior of Spp1 upon tethering.** A subset of Spp1 binding sites (red hexagons) becomes tethered to the chromosome axis (ORFs in red). These tethered sites correspond to the dynamic fraction of Spp1 peaks identified by time resolved ChIP-Seq. Appearing Spp1 sites have the ability to interact with Mer2 and to tether DSB sites to the axis where they undergo Spo11-mediated DNA cleavage (yellow star). These properties depend on the PHD domain and the Mer2 binding motif of Spp1, as well as H3K4 and H3R2 methylation. Constant Spp1 sites (green hexagons) remain unchanged during the meiotic time course. Constant Spp1 sites do not interact with Mer2 and remain colocalized with Set1C (ORFs in green). Disappearing Spp1 sites (blue hexagons) are mainly associated with RPG and SnoRNA genes (ORFs in blue). Spp1 tends to be released from the Set1 holocomplex at the latter sites, reflecting the plasticity of Set1C.

RNA digestion (10 µg RNase) was performed for 1 h at 37°C. The DNA was finally resuspended in 50 µl nuclease-free dH2O.

NGS library preparation and deep sequencing

Sequencing libraries were prepared according to the Illumina's TruSeq ChIP Sample Preparation protocol. In brief, the enriched ChIP DNA was end-repaired and indexed adapters were ligated to the inserts. Purified ligation products were then amplified by PCR. Amplified libraries were prepared in the Genomic Medicine and Bioinformatics Core Facility of the University of Debrecen, Debrecen, Hungary (Halász et al., 2017). The libraries were sequenced using 50 single-end reads with Illumina HiScan SQ (Genomic Medicine and Bioinformatics Core Facility of the University of Debrecen); or with Illumina HiSeq 2500 (EMBL Genomics Core Facility, Heidelberg, Germany).

Raw reads were aligned to the *S. cerevisiae* reference genome (SacCer3; SGD) using the default parameters of Burrows-Wheeler Aligner algorithm (Li and Durbin, 2009) and 38–67% of the sequenced reads were retained after removing low mapping quality (MAPQ < 10) and PCR duplicate reads (Picard).

Enrichment analysis and peak annotation

BayesPeak was used with default parameters to identify ChIP enriched regions (peaks) of the genome compared with input control (Cairns et al., 2011). Peaks sets identified at individual meiotic time points (SPS, 0, 2, 4, and 6 h in SPM) were concatenated and sorted by chromosomal position, and then merged. We used mergeBed (Bedtools; Quinlan and Hall, 2010) to join the overlapping peak positions. The overlap of peak sets detected in different samples was represented by proportional Venn diagrams (Larsson, 2018).

We used deepTools2 (Ramírez et al., 2014) bamCoverage to create Reads Per Kilobase per Million mapped reads (RPKM)–normalized bedgraph files. For each bedgraph we calculated the $\log_2$(IP/INPUT) ratios and used these coverage files for visualization and downstream analysis. Heat maps were generated with computeMatrix and plotHeatmap functions of deepTools2. Read density profiles were generated using HOMER (Li et al., 2009) and plotted in R.

To estimate the enrichment or depletion of Spp1 binding sites within genomic features we created 100–100 randomized peak sets with the shuffleBed Bedtools function (Quinlan and Hall, 2010). Then, we calculated with intersectBed the coverage ratio of observed and randomized peak sets over the relevant annotation categories and over Mer2/Red1 ChIP binding sites. Differences in overlap ratios were then compared by the prop. test function of R.

Identification of dynamic Spp1 clusters

To classify Spp1 binding sites based on their binding dynamics, we first merged every Spp1 binding sites identified at all meiotic time points (union peak set). Next, we mapped the average $\log_2$(IP/INPUT) RPKM ratios of the ChIP samples back to the union peak set. Binding site coverage values were z-transformed across ChIP samples with the scale function in R. Dynamic clusters were identified using a k-means algorithm and plotted with pheatmap (Kolde, 2015).

MDS

Using the union peak set as described above, we applied the cmdscale MDS method in R to visualize the level of similarity between Spp1 datasets. Euclidean distance matrices generated from this table were readily used as an input for cmdscale. The resulting 2D coordinates were plotted in R as a scatter plot.

Turnover rate estimation from c-ChIP data

Average coverage (i.e., the occupancy) of the Spp1 binding sites were calculated using both the GFP- and myc competition ChIP-seq data for each time point separately. Next, GFP/myc occupancy ratios were calculated and the same exponential model was fitted as in Deal et al. (2010):

$$\frac{GFP}{myc} = 1 - e^{-\lambda t}$$

where $\frac{GFP}{myc}$ is the GFP/myc occupancy ratio, e is the mathematical constant ~2.72, t is the time measured from the induction of the GFP-tagged *SPP1* allele (in minutes), and λ is the turnover rate (1/min). This model is identical to the one used by others (Dion et al., 2007; Lickwar et al., 2012). After fitting the model, we estimated the standard error of the estimates and performed a Student's t test to evaluate the goodness of fit of the model. In total, 977 binding sites could be described with the model (P value < 0.05). To assess the consistency of Spp1 peak sets identified by c-ChIP

and ChIP experiments, we calculated the mean differences of IP signals (ChIP – c-ChIP) for all meiotic time points and depicted the results by polar plots (Fig. S3 C) using the radarchart function of R.

Statistical analysis

All statistical analyses were performed in R (version 3.4.4). Group comparisons were performed by ANOVA (aov() R function). Groups were compared with Tukey's post-hoc test (Tukey HSD R function). If the data did not fit the normal distribution, we used Kruskal-Wallis's ANOVA (kruskal.test R function) and the Mann-Whitney *U* test (wilcox.test R function). Probability values of $P \leq 0.001$ were considered as statistically significant. Significance marks: not significant (ns). $P > 0.05$; *, $P \leq 0.05$; **, $P \leq 0.01$; ***, $P \leq 0.001$; ****, $P \leq 0.0001$. The number of cases (n) and P values were indicated in each legend.

External datasets

SacCer3 genome annotation files were obtained from *Saccharomyces* Genome Database. Promoter and downstream regions were defined as the arbitrary extension of TSSs with 500 bp and TTS by 200 bp. RNA-seq (Brar et al., 2012), H3K4me3 ChIP-chip (Borde et al., 2009), Mer2 and Red1 ChIP-Chip (Panizza et al., 2011; Sun et al., 2015), and Spo11-oligo DSB data (Mohibullah and Keeney, 2017) were from the indicated publications.

Data access

Raw sequencing data and processed data files generated in this study have been deposited to Gene Expression Omnibus (GEO) with the accession no. GSE107967.

ChIP-seq data can be accessed and browsed in JBrowse (http://geneart.med.unideb.hu/pub/2018-spp1; login: spp1; password: cerevisiae7).

Online supplemental material

Fig. S1 reports the meiotic progression and sporulation efficiency of yeast strains analyzed by ChIP sequencing. Fig. S2 shows the overlap analysis and ChIP enrichment of Spp1 binding sites in various mutants analyzed in this study. Fig. S3 outlines the experimental design and validation of competition ChIP experiments. Fig. S4 shows the turnover rate estimation from competition ChIP data. Table S1 summarizes the P values related to microscopic colocalization analysis, lists the yeast strains used in this study, and provides basic NGS statistics for ChIP sequencing experiments.

Acknowledgments

We thank Ibolya Fürtös for technical assistance, Christophe de la Roche Saint-André for strains, and the Genomic Medicine and Bioinformatics Core Facility (University of Debrecen) for the NGS service.

L. Székvölgyi received grants from the Hungarian Academy of Sciences (Lendület program, grant LP2015-9/2015), IMéRA-Institut d'Études Avancées - Aix-Marseille University (Marseille, France; grant NKFIH-ERC-HU-117670), and the National Research, Developement, and Innovation Office, Hungary (grant GINOP-2.3.2-15-2016-00024). V. Géli was supported by the Ligue Contre le Cancer, Equipe labelisée Ligue 2017. F. Klein was supported by the Austrian Science Foundation (grants F3405 and F3410).

The authors declare no competing financial interests.

Author contributions: Z. Karányi conceptualized the c-ChIP sequencing experiments, performed, and conducted all bioinformatical and statistical analyses. L. Halász performed bioinformatic analyses (ChIP-seq). L. Acquaviva constructed mutant strains and conceptualized the work. D. Jónás performed bioinformatical analysis (c-ChIP). S. Hetey performed the FCS and FRAP measurements. B. Boros-Oláh performed the ChIP and Western blot experiments. F. Peng constructed strains and performed and analyzed chromosome spread experiments. D. Chen designed and wrote FociSim, a Java program to assess significance of foci colocalization. F. Klein planned and analyzed chromosome spread experiments. V. Géli conceptualized and coordinated the research and wrote the manuscript. L. Székvölgyi conceptualized and coordinated the research, constructed strains, wrote the manuscript, and secured funding.

Submitted: 20 December 2017
Revised: 18 May 2018
Accepted: 9 July 2018

References

Acquaviva, L., J. Drogat, P.M. Dehé, C. de La Roche Saint-André, and V. Géli. 2013a. Spp1 at the crossroads of H3K4me3 regulation and meiotic recombination. *Epigenetics*. 8:355–360. https://doi.org/10.4161/epi.24295

Acquaviva, L., L. Székvölgyi, B. Dichtl, B.S. Dichtl, C. de La Roche Saint André, A. Nicolas, and V. Géli. 2013b. The COMPASS subunit Spp1 links histone methylation to initiation of meiotic recombination. *Science*. 339:215–218. https://doi.org/10.1126/science.1225739

Adam, C., R. Guérois, A. Citarella, L. Verardi, F. Adolphe, C. Béneut, V. Sommermeyer, C. Ramus, J. Govin, Y. Couté, and V. Borde. 2018. The PHD finger protein Spp1 has distinct functions in the Set1 and the meiotic DSB formation complexes. *PLoS Genet*. 14:e1007223. https://doi.org/10.1371/journal.pgen.1007223

Battaglia, S., M. Lidschreiber, C. Baejen, P. Torkler, S.M. Vos, and P. Cramer. 2017. RNA-dependent chromatin association of transcription elongation factors and Pol II CTD kinases. *eLife*. 6:1–26. https://doi.org/10.7554/eLife.25637

Baudat, F., J. Buard, C. Grey, A. Fledel-Alon, C. Ober, M. Przeworski, G. Coop, and B. de Massy. 2010. PRDM9 is a major determinant of meiotic recombination hotspots in humans and mice. *Science*. 327:836–840. https://doi.org/10.1126/science.1183439

Borde, V., and B. de Massy. 2013. Programmed induction of DNA double strand breaks during meiosis: setting up communication between DNA and the chromosome structure. *Curr. Opin. Genet. Dev.* 23:147–155. https://doi.org/10.1016/j.gde.2012.12.002

Borde, V., N. Robine, W. Lin, S. Bonfils, V. Géli, and A. Nicolas. 2009. Histone H3 lysine 4 trimethylation marks meiotic recombination initiation sites. *EMBO J.* 28:99–111. https://doi.org/10.1038/emboj.2008.257

Brar, G.A., M. Yassour, N. Friedman, A. Regev, N.T. Ingolia, and J.S. Weissman. 2012. High-resolution view of the yeast meiotic program revealed by ribosome profiling. *Science*. 335:552–557. https://doi.org/10.1126/science.1215110

Brazda, P., J. Krieger, B. Daniel, D. Jonas, T. Szekeres, J. Langowski, K. Tóth, L. Nagy, and G. Vámosi. 2014. Ligand binding shifts highly mobile retinoid X receptor to the chromatin-bound state in a coactivator-dependent manner, as revealed by single-cell imaging. *Mol. Cell. Biol.* 34:1234–1245. https://doi.org/10.1128/MCB.01097-13

Cairns, J., C. Spyrou, R. Stark, M.L. Smith, A.G. Lynch, and S. Tavaré. 2011. BayesPeak--an R package for analysing ChIP-seq data. *Bioinformatics*. 27:713–714. https://doi.org/10.1093/bioinformatics/btq685

D'Urso, A., Y.H. Takahashi, B. Xiong, J. Marone, R. Coukos, C. Randise-Hinchliff, J.-P. Wang, A. Shilatifard, and J.H. Brickner. 2016. Set1/COMPASS

and Mediator are repurposed to promote epigenetic transcriptional memory. *eLife*. 5:e16691. https://doi.org/10.7554/eLife.16691

Daniel, J.A., M.A. Santos, Z. Wang, C. Zang, K.R. Schwab, M. Jankovic, D. Filsuf, H.-T. Chen, A. Gazumyan, A. Yamane, et al. 2010. PTIP promotes chromatin changes critical for immunoglobulin class switch recombination. *Science*. 329:917–923. https://doi.org/10.1126/science.1187942

Deal, R.B., J.G. Henikoff, and S. Henikoff. 2010. Genome-wide kinetics of nucleosome turnover determined by metabolic labeling of histones. *Science*. 328:1161–1164. https://doi.org/10.1126/science.1186777

Dehé, P.-M., B. Dichtl, D. Schaft, A. Roguev, M. Pamblanco, R. Lebrun, A. Rodríguez-Gil, M. Mkandawire, K. Landsberg, A. Shevchenko, et al. 2006. Protein interactions within the Set1 complex and their roles in the regulation of histone 3 lysine 4 methylation. *J. Biol. Chem.* 281:35404–35412. https://doi.org/10.1074/jbc.M603099200

Dion, M.F., T. Kaplan, M. Kim, S. Buratowski, N. Friedman, and O.J. Rando. 2007. Dynamics of replication-independent histone turnover in budding yeast. *Science*. 315:1405–1408. https://doi.org/10.1126/science.1134053

Ernst, P., and C.R. Vakoc. 2012. WRAD: enabler of the SET1-family of H3K4 methyltransferases. *Brief. Funct. Genomics*. 11:217–226. https://doi.org/10.1093/bfgp/els017

Gothwal, S.K., N.J. Patel, M.M. Colletti, H. Sasanuma, M. Shinohara, A. Hochwagen, and A. Shinohara. 2016. The double-strand break landscape of meiotic chromosomes is shaped by the Paf1 transcription elongation complex in Saccharomyces cerevisiae. *Genetics*. 202:497–512. https://doi.org/10.1534/genetics.115.177287

Halász, L., Z. Karányi, B. Boros-Oláh, T. Kuik-Rózsa, É. Sipos, É. Nagy, Á. Mosolygó-L, A. Mázló, É. Rajnavölgyi, G. Halmos, and L. Székvölgyi. 2017. RNA-DNA hybrid (R-loop) immunoprecipitation mapping: an analytical workflow to evaluate inherent biases. *Genome Res.* 27:1063–1073. https://doi.org/10.1101/gr.219394.116

Halbach, A., H. Zhang, A. Wengi, Z. Jablonska, I.M.L. Gruber, R.E. Halbeisen, P.-M. Dehé, P. Kemmeren, F. Holstege, V. Géli, et al. 2009. Cotranslational assembly of the yeast SET1C histone methyltransferase complex. *EMBO J.* 28:2959–2970. https://doi.org/10.1038/emboj.2009.240

Hetey, S., B. Boros-Oláh, T. Kuik-Rózsa, Q. Li, Z. Karányi, Z. Szabó, J. Roszik, N. Szalóki, G. Vámosi, K. Tóth, and L. Székvölgyi. 2017. Biophysical characterization of histone H3.3 K27M point mutation. *Biochem. Biophys. Res. Commun.* 490:868–875. https://doi.org/10.1016/j.bbrc.2017.06.133

Keeney, S., C.N. Giroux, and N. Kleckner. 1997. Meiosis-specific DNA double-strand breaks are catalyzed by Spo11, a member of a widely conserved protein family. *Cell*. 88:375–384. https://doi.org/10.1016/S0092-8674(00)81876-0

Kim, J., J.A. Kim, R.K. McGinty, U.T.T. Nguyen, T.W. Muir, C.D. Allis, and R.G. Roeder. 2013. The n-SET domain of Set1 regulates H2B ubiquitylation-dependent H3K4 methylation. *Mol. Cell*. 49:1121–1133. https://doi.org/10.1016/j.molcel.2013.01.034

Kirmizis, A., H. Santos-Rosa, C.J. Penkett, M.A. Singer, M. Vermeulen, M. Mann, J. Bähler, R.D. Green, and T. Kouzarides. 2007. Arginine methylation at histone H3R2 controls deposition of H3K4 trimethylation. *Nature*. 449:928–932. https://doi.org/10.1038/nature06160

Kolde, R.2015. pheatmap: Pretty Heatmaps. https://cran.r-project.org/web/packages/pheatmap/index.html (accessed May 19, 2018).

Kraushaar, D.C., W. Jin, A. Maunakea, B. Abraham, M. Ha, and K. Zhao. 2013. Genome-wide incorporation dynamics reveal distinct categories of turnover for the histone variant H3.3. *Genome Biol.* 14:R121. https://doi.org/10.1186/gb-2013-14-10-r121

Krieger, J.W., and J. Langowski. 2015. QuickFit 3.0 (status: beta, compiled: 2015-03-18, SVN: 3891): A data evaluation application for biophysics. http://www.dkfz.de/Macromol/quickfit/ (accessed March 18, 2015).

Krogan, N.J., J. Dover, A. Wood, J. Schneider, J. Heidt, M.A. Boateng, K. Dean, O.W. Ryan, A. Golshani, M. Johnston, et al. 2003. The Paf1 complex is required for histone H3 methylation by COMPASS and Dot1p: linking transcriptional elongation to histone methylation. *Mol. Cell*. 11:721–729. https://doi.org/10.1016/S1097-2765(03)00091-1

Kurzbauer, M.T., C. Uanschou, D. Chen, and P. Schlögelhofer. 2012. The recombinases DMC1 and RAD51 are functionally and spatially separated during meiosis in Arabidopsis. *Plant Cell*. 24:2058–2070. https://doi.org/10.1105/tpc.112.098459

Lam, I., and S. Keeney. 2015. Mechanism and regulation of meiotic recombination initiation. *Cold Spring Harb. Perspect. Biol.* 7:a016634. https://doi.org/10.1101/cshperspect.a016634

Lardenois, A., Y. Liu, T. Walther, F. Chalmel, B. Evrard, M. Granovskaia, A. Chu, R.W. Davis, L.M. Steinmetz, and M. Primig. 2011. Execution of the meiotic noncoding RNA expression program and the onset of gametogenesis in yeast require the conserved exosome subunit Rrp6. *Proc. Natl. Acad. Sci. USA*. 108:1058–1063. https://doi.org/10.1073/pnas.1016459108

Larsson, J.2018. eulerr: Area-Proportional Euler and Venn Diagrams with Ellipses. http://lup.lub.lu.se/student-papers/record/8934042 (accessed February 14, 2018).

Li, H., and R. Durbin. 2009. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics*. 25:1754–1760. https://doi.org/10.1093/bioinformatics/btp324

Li, H., B. Handsaker, A. Wysoker, T. Fennell, J. Ruan, N. Homer, G. Marth, G. Abecasis, and R. Durbin. 1000 Genome Project Data Processing Subgroup. 2009. The Sequence Alignment/Map format and SAMtools. *Bioinformatics*. 25:2078–2079. https://doi.org/10.1093/bioinformatics/btp352

Lickwar, C.R., F. Mueller, S.E. Hanlon, J.G. McNally, and J.D. Lieb. 2012. Genome-wide protein-DNA binding dynamics suggest a molecular clutch for transcription factor function. *Nature*. 484:251–255. https://doi.org/10.1038/nature10985

Lickwar, C.R., F. Mueller, and J.D. Lieb. 2013. Genome-wide measurement of protein-DNA binding dynamics using competition ChIP. *Nat. Protoc.* 8:1337–1353. https://doi.org/10.1038/nprot.2013.077

Liu, H., S. Takeda, R. Kumar, T.D. Westergard, E.J. Brown, T.K. Pandita, E.H.-Y. Cheng, and J.J.-D. Hsieh. 2010. Phosphorylation of MLL by ATR is required for execution of mammalian S-phase checkpoint. *Nature*. 467:343–346. https://doi.org/10.1038/nature09350

Luciano, P., J. Jeon, A. El-Kaoutari, D. Challal, A. Bonnet, M. Barucco, T. Candelli, F. Jourquin, P. Lesage, J. Kim, et al. 2017. Binding to RNA regulates Set1 function. *Cell Discov.* 3:17040. https://doi.org/10.1038/celldisc.2017.40

Miller, T., N.J. Krogan, J. Dover, H. Erdjument-Bromage, P. Tempst, M. Johnston, J.F. Greenblatt, and A. Shilatifard. 2001. COMPASS: a complex of proteins associated with a trithorax-related SET domain protein. *Proc. Natl. Acad. Sci. USA*. 98:12902–12907. https://doi.org/10.1073/pnas.231473398

Mohibullah, N., and S. Keeney. 2017. Numerical and spatial patterning of yeast meiotic DNA breaks by Tel1. *Genome Res.* 27:278–288. https://doi.org/10.1101/gr.213587.116

Nagy, P.L., J. Griesenbeck, R.D. Kornberg, and M.L. Cleary. 2002. A trithorax-group complex purified from Saccharomyces cerevisiae is required for methylation of histone H3. *Proc. Natl. Acad. Sci. USA*. 99:90–94. https://doi.org/10.1073/pnas.221596698

Ng, H.H., S. Dole, and K. Struhl. 2003. The Rtf1 component of the Paf1 transcriptional elongation complex is required for ubiquitination of histone H2B. *J. Biol. Chem.* 278:33625–33628. https://doi.org/10.1074/jbc.C300270200

Pan, J., M. Sasaki, R. Kniewel, H. Murakami, H.G. Blitzblau, S.E. Tischfield, X. Zhu, M.J. Neale, M. Jasin, N.D. Socci, et al. 2011. A hierarchical combination of factors shapes the genome-wide topography of yeast meiotic recombination initiation. *Cell*. 144:719–731. https://doi.org/10.1016/j.cell.2011.02.009

Panizza, S., M.A. Mendoza, M. Berlinger, L. Huang, A. Nicolas, K. Shirahige, and F. Klein. 2011. Spo11-accessory proteins link double-strand break sites to the chromosome axis in early meiotic recombination. *Cell*. 146:372–383. https://doi.org/10.1016/j.cell.2011.07.003

Parvanov, E.D., P.M. Petkov, and K. Paigen. 2010. Prdm9 controls activation of mammalian recombination hotspots. *Science*. 327:835. https://doi.org/10.1126/science.1181495

Primig, M., R.M. Williams, E.A. Winzeler, G.G. Tevzadze, A.R. Conway, S.Y. Hwang, R.W. Davis, and R.E. Esposito. 2000. The core meiotic transcriptome in budding yeasts. *Nat. Genet.* 26:415–423. https://doi.org/10.1038/82539

Quinlan, A.R., and I.M. Hall. 2010. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics*. 26:841–842. https://doi.org/10.1093/bioinformatics/btq033

Ramírez, F., F. Dündar, S. Diehl, B.A. Grüning, and T. Manke. 2014. deepTools: a flexible platform for exploring deep-sequencing data. *Nucleic Acids Res.* 42(W1):W187–W191. https://doi.org/10.1093/nar/gku365

Roguev, A., D. Schaft, A. Shevchenko, W.W.M.P. Pijnappel, M. Wilm, R. Aasland, and A.F. Stewart. 2001. The Saccharomyces cerevisiae Set1 complex includes an Ash2 homologue and methylates histone 3 lysine 4. *EMBO J.* 20:7137–7148. https://doi.org/10.1093/emboj/20.24.7137

Rothbart, S.B., and B.D. Strahl. 2014. Interpreting the language of histone and DNA modifications. *Biochim. Biophys. Acta*. 1839:627–643. https://doi.org/10.1016/j.bbagrm.2014.03.001

Ruthenburg, A.J., C.D. Allis, and J. Wysocka. 2007. Methylation of lysine 4 on histone H3: intricacy of writing and reading a single epigenetic mark. *Mol. Cell*. 25:15–30. https://doi.org/10.1016/j.molcel.2006.12.014

Sayou, C., G. Millán-Zambrano, H. Santos-Rosa, E. Petfalski, S. Robson, J. Houseley, T. Kouzarides, and D. Tollervey. 2017. RNA binding by the histone methyltransferases Set1 and Set2. *Mol. Cell. Biol.* 37:e00165-17. https://doi.org/10.1128/MCB.00165-17

Shilatifard, A. 2012. The COMPASS family of histone H3K4 methylases: mechanisms of regulation in development and disease pathogenesis. *Annu. Rev. Biochem.* 81:65–95. https://doi.org/10.1146/annurev-biochem-051710-134100

Soares, L.M., M. Radman-Livaja, S.G. Lin, O.J. Rando, and S. Buratowski. 2014. Feedback control of Set1 protein levels is important for proper H3K4 methylation patterns. *Cell Reports.* 6:961–972. https://doi.org/10.1016/j.celrep.2014.02.017

Sollier, J., W. Lin, C. Soustelle, K. Suhre, A. Nicolas, V. Géli, and C. de La Roche Saint-André. 2004. Set1 is required for meiotic S-phase onset, double-strand break formation and middle gene expression. *EMBO J.* 23:1957–1967. https://doi.org/10.1038/sj.emboj.7600204

Sommermeyer, V., C. Béneut, E. Chaplais, M.E. Serrentino, and V. Borde. 2013. Spp1, a member of the Set1 Complex, promotes meiotic DSB formation in promoters by tethering histone H3K4 methylation sites to chromosome axes. *Mol. Cell.* 49:43–54. https://doi.org/10.1016/j.molcel.2012.11.008

Stasevich, T.J., F. Mueller, A. Michelman-Ribeiro, T. Rosales, J.R. Knutson, and J.G. McNally. 2010. Cross-validating FRAP and FCS to quantify the impact of photobleaching on in vivo binding estimates. *Biophys. J.* 99:3093–3101. https://doi.org/10.1016/j.bpj.2010.08.059

Sun, X., L. Huang, T.E. Markowitz, H.G. Blitzblau, D. Chen, F. Klein, and A. Hochwagen. 2015. Transcription dynamically patterns the meiotic chromosome-axis interface. *eLife.* 4:1–23. https://doi.org/10.7554/eLife.07424

Székvölgyi, L., and A. Nicolas. 2010. From meiosis to postmeiotic events: homologous recombination is obligatory but flexible. *FEBS J.* 277:571–589. https://doi.org/10.1111/j.1742-4658.2009.07502.x

Székvölgyi, L., K. Ohta, and A. Nicolas. 2015. Initiation of meiotic homologous recombination: flexibility, impact of histone modifications, and chromatin remodeling. *Cold Spring Harb. Perspect. Biol.* 7:a016527. https://doi.org/10.1101/cshperspect.a016527

Tessarz, P., and T. Kouzarides. 2014. Histone core modifications regulating nucleosome structure and dynamics. *Nat. Rev. Mol. Cell Biol.* 15:703–708. https://doi.org/10.1038/nrm3890

Trésaugues, L., P.-M. Dehé, R. Guérois, A. Rodriguez-Gil, I. Varlet, P. Salah, M. Pamblanco, P. Luciano, S. Quevillon-Cheruel, J. Sollier, et al. 2006. Structural characterization of Set1 RNA recognition motifs and their role in histone H3 lysine 4 methylation. *J. Mol. Biol.* 359:1170–1181. https://doi.org/10.1016/j.jmb.2006.04.050

Xaver, M., L. Huang, D. Chen, and F. Klein. 2013. Smc5/6-Mms21 prevents and eliminates inappropriate recombination intermediates in meiosis. *PLoS Genet.* 9:e1004067. https://doi.org/10.1371/journal.pgen.1004067

Yamaguchi, S., K. Hong, R. Liu, L. Shen, A. Inoue, D. Diep, K. Zhang, and Y. Zhang. 2012. Tet1 controls meiosis by regulating meiotic gene expression. *Nature.* 492:443–447. https://doi.org/10.1038/nature11709

Yamashita, K., M. Shinohara, and A. Shinohara. 2004. Rad6-Bre1-mediated histone H2B ubiquitylation modulates the formation of double-strand breaks during meiosis. *Proc. Natl. Acad. Sci. USA.* 101:11380–11385. https://doi.org/10.1073/pnas.0400078101

Yuan, C.C., A.G.W. Matthews, Y. Jin, C.F. Chen, B.A. Chapman, T.K. Ohsumi, K.C. Glass, T.G. Kutateladze, M.L. Borowsky, K. Struhl, and M.A. Oettinger. 2012. Histone H3R2 symmetric dimethylation and histone H3K4 trimethylation are tightly correlated in eukaryotic genomes. *Cell Reports.* 1:83–90. https://doi.org/10.1016/j.celrep.2011.12.008

Supplemental material

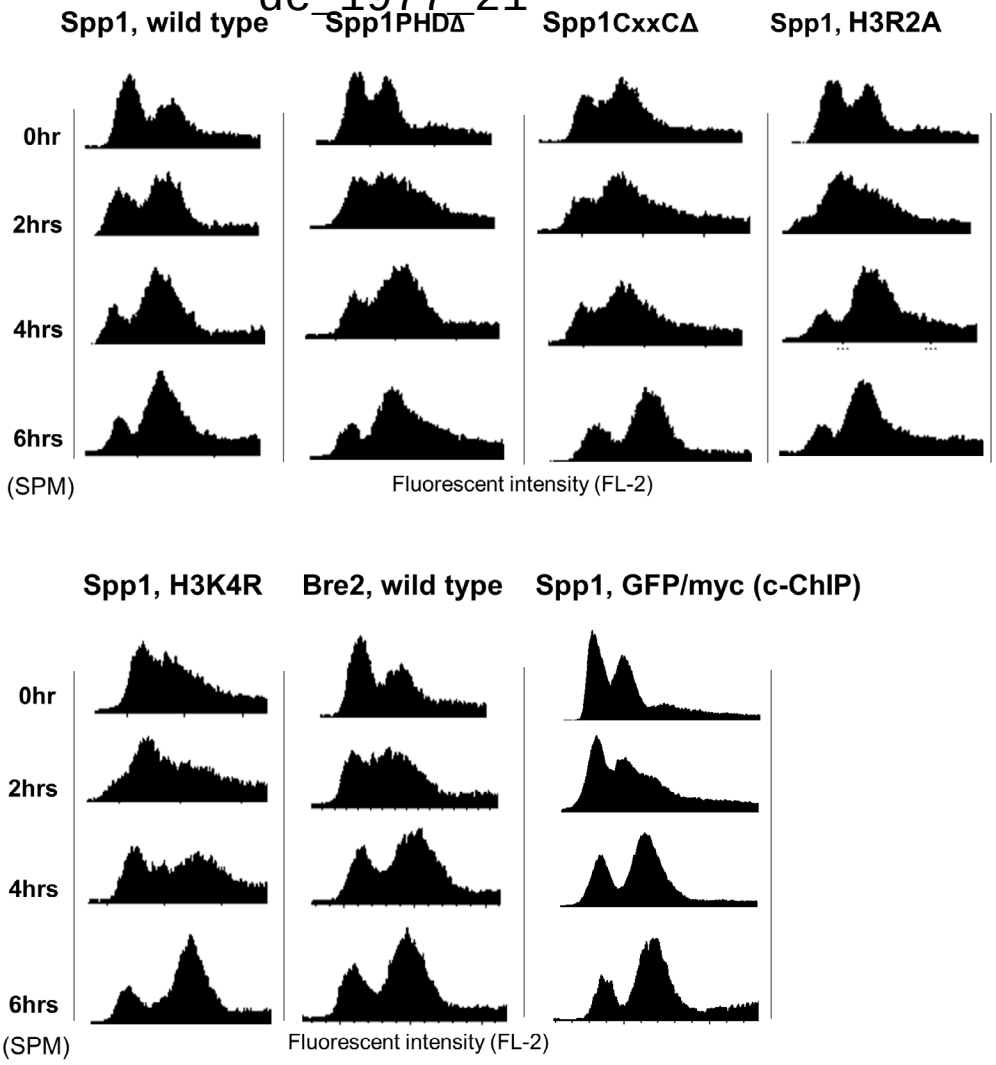
Karányi et al., https://doi.org/10.1083/jcb.201712122

Provided in a separate Excel file is Table S1, summarizing the P values related to microscopic colocalization analysis, listing the yeast strains used in this study, and providing basic NGS statistics for ChIP sequencing experiments.

Reference

Adam, C., R. Guérois, A. Citarella, L. Verardi, F. Adolphe, C. Béneut, V. Sommermeyer, C. Ramus, J. Govin, Y. Couté, and V. Borde. 2018. The PHD finger protein Spp1 has distinct functions in the Set1 and the meiotic DSB formation complexes. *PLoS Genet.* 14:e1007223. https://doi.org/10.1371/journal.pgen.1007223

A



B

strain	spore viability (%)	sporulation efficiency (%)
Bre2-myc	95	88.7
Spp1-myc wild type	94	77.4
Spp1-myc PHDΔ	94	72.5
Spp1-myc CxxCΔ	89	69.5
Spp1-HA H3R2A	69	68.1
Spp1-HA H3K4R	70	72.4
Spp1, GFP/myc (c-ChIP)	90	73.3

Figure S1. **Meiotic progression and sporulation efficiency of yeast strains analyzed by ChIP sequencing. (A)** FACS profile of the sporulating strains measured between 0 and 6 h in SPM. **(B)** Spore viability and sporulation efficiency of the strains. The proportion of viable spores was determined by tetrad dissection (100 tetrads were dissected per strain). Sporulation efficiency was assessed by DAPI staining of nuclei (after 24 h in SPM) and counting the number of full tetrads (200 cells were analyzed in each sample).

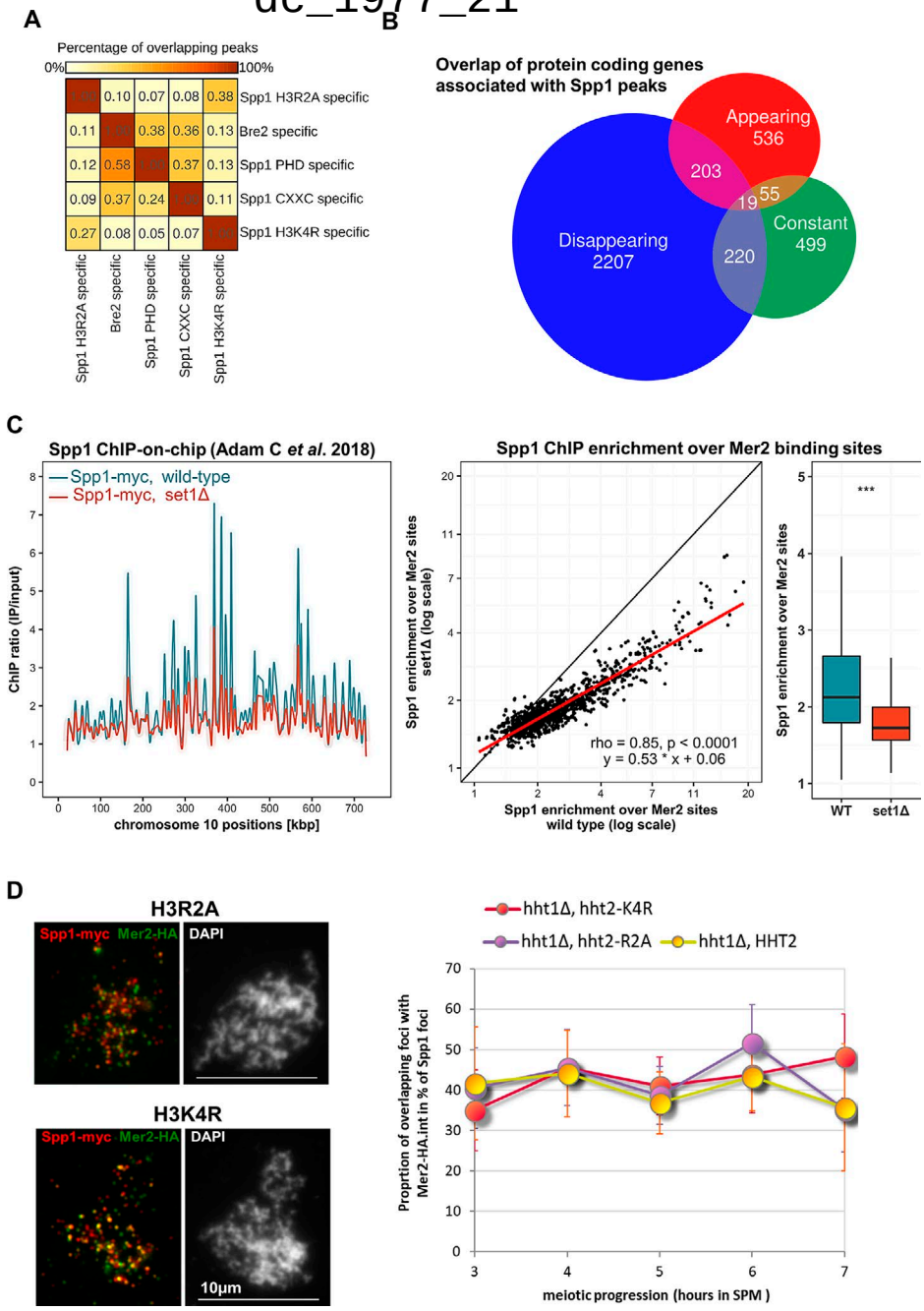


Figure S2. **Overlap analysis and ChIP enrichment of Spp1 binding sites in various mutants analyzed in this study. (A)** Heat map showing the overlap of mutant-specific Spp1 binding sites (ChIP peak positions) identified in each sample (related to Fig. 3 D). Color strength is proportional to the overlap ratio of Spp1 peak sets. The data were obtained from two biological replicate experiments. Sample sizes (n): PHDΔ-specific (183), CXXCΔ-specific (287), Bre2-specific (283), H3R2A-specific (331), and H3K4R-specific (465). **(B)** Overlap of protein coding genes associated with Spp1 peaks (related to Figs. 4, A and B; and Fig. 5 H). ORFs have been assigned to their closest Spp1 binding site and overlap ratios were visualized by proportional Venn diagrams. Number of cases (n) is indicated in the Venn diagram by the amount of differential or common ORFs associated with Spp1. **(C)** Set1-mediated histone methylation facilitates Spp1 localization on chromosomal axes (related to Fig. 3 E). Left, Spp1 ChIP-on-Chip binding profile in wild-type and *set1Δ* cells in meiosis (data are from Adam et al., 2018). ChIP enrichment profiles are shown for chromosome 10 as a representative example. Middle, meiotic Spp1 ChIP profiles over Mer2 axial sites positively correlate in wild-type and *set1Δ* cells (Spearman's rank correlation). However, the slope of regression line (red linear; 0.53) negatively deviates from the diagonal, indicating low Spp1 enrichment in the absence of Set1. Right, Spp1 enrichment at Mer2 binding sites is significantly reduced is *set1Δ* cells compared with wild-type cells. Box-whiskers plots show the medians (with interquartile ranges) of Spp1 ChIP enrichment at Mer2 binding sites. Statistically significant difference is indicated between wild-type and *set1Δ* cells (Mann-Whitney *U* test, ***, P < 0.001). The number of cases (n; Mer2 binding sites) equals to 818. **(D)** Immunofluorescent localization of Spp1-myc and Mer2-HA in the absence of modifiable H3R2 and H3K4 histone residues (related to Fig. 3 E). Left, representative chromatin spread images showing Spp1-myc foci (red) and Mer2-HA foci (green) in an H3R2A mutant (*hht1Δ* and *hht2-R2A*) and an H3K4R mutant (*hht1Δ* and *hht2-K4R*). Cells were from a meiotic time course at 4 h after transfer to SPM (SPM). DNA was stained by DAPI (white). Right, percentage of Spp1-myc foci overlapping with Mer2-HA in *HHT2, hht1Δ* (wild-type, yellow), *hht2-R2A* and *hht1Δ* (violet), and *hht2-K4R* and *hht1Δ* (red) cells. >7,000 foci were analyzed for each of the three analyzed strains. Error bars represent SD.

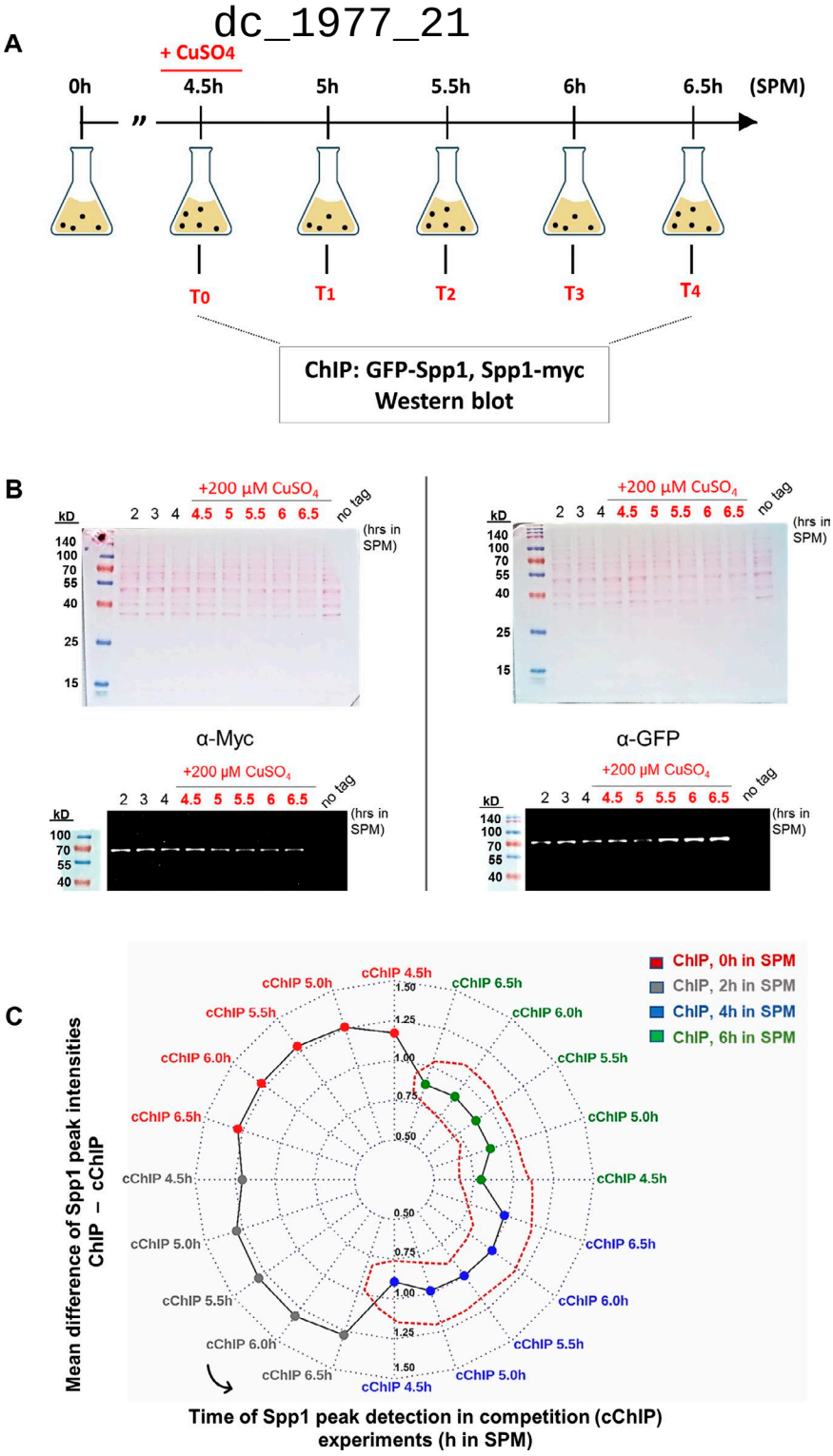


Figure S3. **Experimental design and validation of the competition ChIP experiment. (A)** Flowchart of the c-ChIP experiment. Cells were collected at various meiotic time points after addition of 200 µM $CuSO_4$, switching the inducible *SPP1* allele on (GFP-Spp1). Two independent biological replicate experiments were performed. **(B)** Quantitative Western blot analysis of the proper expression of constitutive Spp1-myc and inducible GFP-Spp1 proteins. Top, Ponceau staining. Bottom, detection by anti-myc and anti-GFP antibodies. The data are representative of one biological replicate experiment. **(C)** Correlation of Spp1 binding sites that were identified by c-ChIP and ChIP, represented on a polar plot. Angular positions indicate the meiotic time points, while color codes highlight the temporally compared peak sets (ChIP time points vs. c-ChIP time points). Radial grid lines define the mean differences of Spp1 IP signals (ChIP – c-ChIP). Distance from the origin is proportional to the variability of Spp1 binding sites (i.e., origin proximal datasets show high similarity between c-ChIP and ChIP). Encircled part of the polar plot (dotted red line) depicts highly ranked datasets, showing the highest similarity between 4 and 6.5 h in SPM. The c-ChIP and ChIP data are representative of two independent biological replicate experiments. Sample size (n; number of peaks): ChIP (4,524); c-ChIP (4,639).

A

	Number of Spp1 peaks	Goodness of fit parameter #
c-ChIP peaks (total)	4639	
c-ChIP peaks (fitted) *	977	1.47 %

* Exponential model: $\frac{GFP}{MYC} = (1 - e^{-\beta_1 * time})$

average Standard Error (SE)

B

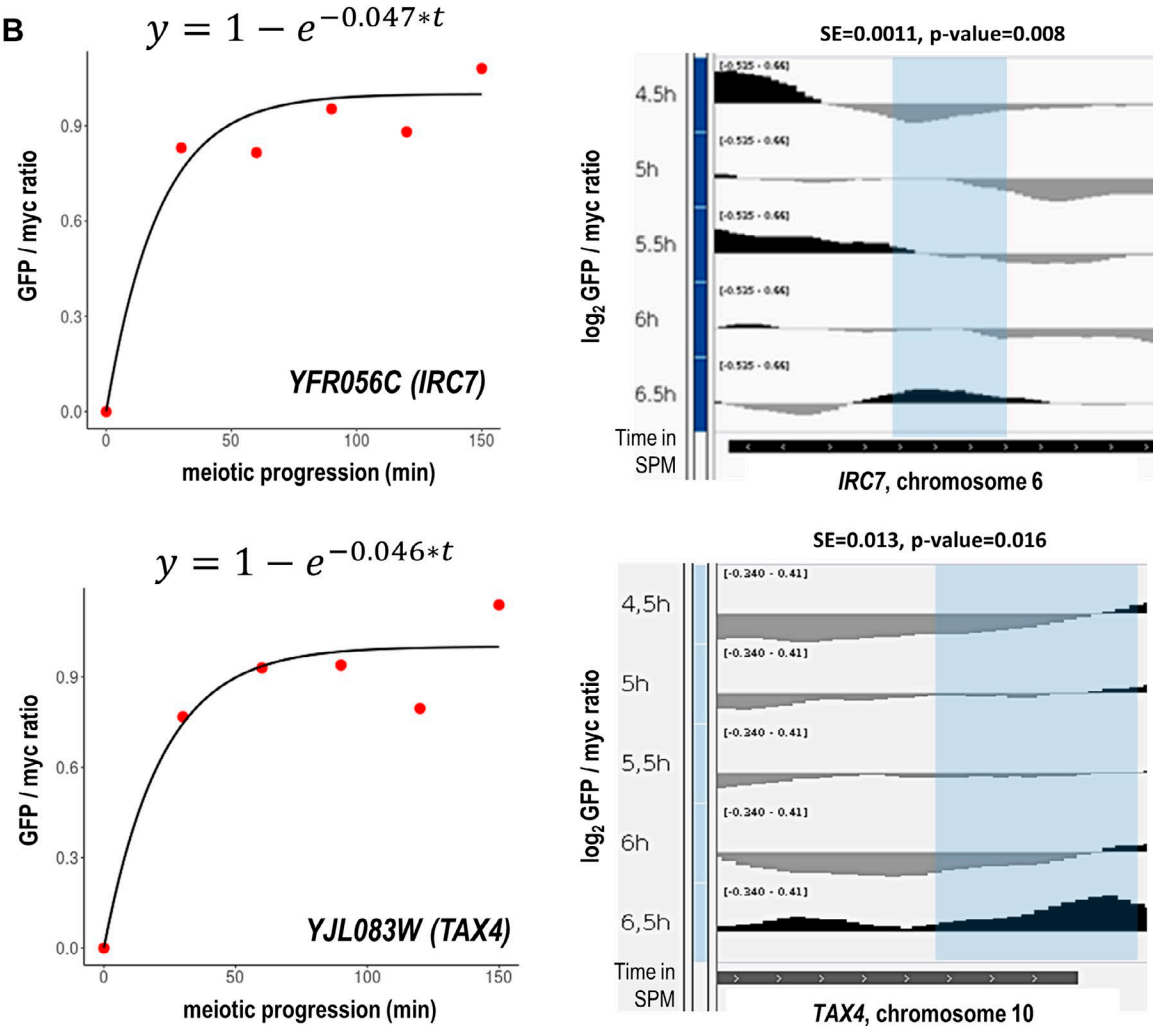


Figure S4. **Turnover rate estimation from competition ChIP data. (A)** Table showing the total number of peaks determined by c-ChIP and the number of Spp1 binding sites that could be fitted with an exponential model. The kinetic model is highlighted with an asterisk (*). Goodness of fit parameters for each binding site were estimated based on the average standard error. **(B)** Example fits of the turnover rate model for two representative Spp1 binding sites (*YFR056C* and *YJL083W*). Model equations (with standard error and P value) and temporal changes of GFP/myc ratios are shown in the left panel. JBrowse snapshots are shown on the right. The c-ChIP data are representative of two independent biological replicate experiments. Sample size (n; number of peaks): c-ChIP total (4,639); c-ChIP fitted (977).

frontiers in Cell and Developmental Biology

dc_1977_21

BRIEF RESEARCH REPORT
published: 10 January 2020
doi: 10.3389/fcell.2019.00364

Histone H3 Lysine 56 Acetylation Is Required for Formation of Normal Levels of Meiotic DNA Breaks in *S. cerevisiae*

Zsolt Karányi[1,2], Lilla Hornyák[1] and Lóránt Székvölgyi[1]†*

[1] MTA-DE Momentum Genome Architecture and Recombination Research Group, Department of Biochemistry and Molecular Biology, Faculty of Medicine, University of Debrecen, Debrecen, Hungary, [2] Department of Internal Medicine, Faculty of Medicine, University of Debrecen, Debrecen, Hungary

Meiotic recombination is initiated by Spo11-catalyzed DNA double-strand breaks (DSBs) that are promoted by histone modifications and histone modifying enzymes. Herein we investigated the role of histone H3 lysine 56 acetylation (H3K56ac) located near the entry/exit points of the DNA in the globular H3 domain. We generated a series of mutant cells (*asf1*Δ, *rtt109*Δ, *hst3/4*Δ, and H3K56A) in which the endogenous level of H3K56ac was manipulated and tracked during meiotic growth. We show that complete loss or increased abundance of H3K56ac in these mutants allows timely entry into meiosis and sporulation and does not impair S phase progression, first and second meiotic cell divisions, and spore viability. In the *asf1*Δ, *rtt109*Δ, *hst3/4*Δ mutants, DSBs and crossovers form normal levels with a short (60-min) delay at the *HIS4-LEU2* artificial recombination hotspot, however, DSB formation shows a ∼threefold decrease in the H3K56A mutant at the natural *BUD23-ARE1* hotspot. The latter DSB phenotype, showing significant DSB reduction in the H3K56A mutant, was also observed at DSB sites using genome-wide mapping of Rfa1-coated single-stranded DNA flanking DSBs (RPA ChIP). Parallel mapping of H3K56-acetylated histones in wild type cells revealed strong depletion of the H3K56ac ChIP signal over Spo11-oligo DSBs, albeit most H3K56-acetylated histones were enriched adjacent to the identified RPA ChIP binding sites. Taken together, these associations demonstrate a prominent role of H3 lysine 56 acetylation in the formation of DNA breaks within recombination hotspot regions.

Keywords: recombination, DNA break, meiosis, histone modification, H3K56 acetylation

OPEN ACCESS

Edited by:
Jean-Philippe Lambert, Laval University, Canada

Reviewed by:
Jeffrey Fillingham, Ryerson University, Canada
Craig Peterson, University of Massachusetts Medical School, United States

****Correspondence:***
Lóránt Székvölgyi
lorantsz@med.unideb.hu

†ORCID:
Lóránt Székvölgyi
orcid.org/0000-0002-7529-0319

Specialty section:
This article was submitted to Epigenomics and Epigenetics, a section of the journal Frontiers in Cell and Developmental Biology

Received: *25 October 2019*
Accepted: *12 December 2019*
Published: *10 January 2020*

Citation:
Karányi Z, Hornyák L and Székvölgyi L (2020) Histone H3 Lysine 56 Acetylation Is Required for Formation of Normal Levels of Meiotic DNA Breaks in S. cerevisiae. Front. Cell Dev. Biol. 7:364. doi: 10.3389/fcell.2019.00364

INTRODUCTION

Meiosis is a cellular differentiation process which is accompanied by high levels of recombination between the homologous chromosomes, initiated by DNA double-strand breaks (DSBs) catalyzed by Spo11 and accessory factors (Keeney et al., 1997; Székvölgyi and Nicolas, 2010; Székvölgyi et al., 2015). In *S. cerevisiae*, meiotic DSBs are controlled by the elements of chromatin structure. This involves a complex interplay between DNA sequence composition, local chromatin status, nucleosome occupancy, and transcription factor binding (Pan et al., 2011). Among these factors, histone modifications represent an important layer which has only recently been explored in detail. The most well-characterized histone modification is H3K4me3 that appears to be essential for recombination sites in most organisms. In the current model, H3K4me3 is deposited by Set1C

and becomes recognized by Spp1 (the PHD finger subunit of Set1C), which leads to tethering of DSB sites to the chromosome axis that undergo Spo11-mediated cleavage (Acquaviva et al., 2013; Sommermeyer et al., 2013; Karányi et al., 2018). A similar mechanism has been proposed in mammals involving the meiosis-specific H3K4 methylase Prdm9 (Baudat et al., 2010; Parpanov et al., 2010), however, CXXC1 (the yeast ortholog of Spp1) is apparently not essential for the association of H3K4 tri-methylated recombination sites with the DSB machinery (Tian et al., 2018).

The widely localized H3K4me3 mark has the virtue of initiating recombination at numerous places in the genome, however, other histone modifications are also needed to keep recombination flexible for the diversity of recombinant haplotypes (Szekvolgyi and Nicolas, 2010). These "alternative" pathways remain to be clarified to better understand the plasticity of crossover patterning. Most chemical modifications are concentrated at the N-termini of histones and are not expected to alter the structure of nucleosomes (White et al., 2001; Biswas et al., 2011). However, modifications of histone core domains can directly change nucleosome structure, which is well-established biochemically (Luger et al., 1997; Biswas et al., 2011) but its functional relevance is less understood.

In the current study, we focused on the role of histone H3 lysine 56 acetylation located near the entry/exit points of the DNA in the globular H3 domain, predicted to destabilize the histone/DNA contact (Buning and Van Noort, 2010; Simon et al., 2011). H3K56ac is a transient chromatin signal showing rapid turnover and is closely linked to DNA replication and histone eviction during transcription (Rufiange et al., 2007; Watanabe et al., 2013). Functional studies in mitotically proliferating yeast cells revealed that H3K56ac enables the assembly and disassembly of nucleosomes during DNA synthesis and upon transcriptional activation. Furthermore, the histone chaperone Asf1 [carrying the H3K56-specific acetyltransferase Rtt109 (Abshiru et al., 2013)] and the histone residue H3K56 were found to be necessary for meiotic S phase progression and for the development of reproductive capacity in yeast and mouse models (Recht et al., 2006; Govin et al., 2010; Messiaen et al., 2016). However, the mechanism of action of H3K56 acetylation has not been fully elucidated and remains to be clarified. Therefore, we applied a functional approach in meiotic *S cerevisiae* cells to modify the natural levels of H3K56ac to study the biochemical phases of meiosis. Our results demonstrate that H3K56ac is necessary for formation of normal levels of DSBs within recombination hotspot regions.

METHODS

All methods are available as **Supplementary Material**.

RESULTS

We generated mutant cells in which the endogenous level of H3K56ac was modified and tracked during meiotic growth (**Figure 1A**). In wild-type cells, the level of H3K56ac reached a peak during DSB formation (at ∼4–5 h in SPM) and then rapidly dropped, similar to earlier western blot results (Recht et al., 2006). The H3K56ac signal disappeared almost completely from *asf1*Δ and *rtt109*Δ deletion mutants during the meiotic time course (∼fourfold reduction compared to wild type), in agreement with the crucial role of Asf1 and Rtt109 in the deposition of this epigenetic tag (Tsubota et al., 2007). The K56ac signal increased and remained high in the *hst3/4*Δ double-mutant that prevents deacetylation of H3K56 (Celic et al., 2006). In mitotic growth conditions, we could not detect an H3K56 acetylation signal, except for the *hst3/4*Δ mutant (**Figure 1A**), in which H3K56ac persists throughout the cell cycle (Celic et al., 2006). This is in accordance with the transient nature of H3K56 acetylation in asynchronously proliferating mitotic cells, mainly associated with newly replicated chromatin regions (Simoneau et al., 2015).

The correlation between H3K56ac dynamics and the phase of meiotic DSBs prompted us to analyze the progress of S phase, DSB formation, sporulation efficiency, and spore viability in cells with various H3K56ac levels. FACS analysis revealed a 30–60 min delay in G1/S phase progression in mutant cells, however, all mutants reached the G2/M phase within 6 h in SPM (**Figures 1B,C**). We then performed a meiotic time course up to 20 h. The results reported in **Figures 1D,E** show that sporulation efficiencies and spore viabilities do not differ between mutant and wild-type cells (less than 1.5-fold change was observed in the number of tetrads and viable spores). We conclude that absence or elevated levels of H3K56ac is dispensable for meiotic S phase progression, MI/MII cell divisions, and spore viability.

We next analyzed whether DSBs were affected in mutants (with $Sae2^{+}$/$Rad50^{+}$ background). DSB levels were monitored at the *HIS4-LEU2* hotspot (Xu and Kleckner, 1995) using Southern blot hybridization (**Figure 2A**). All mutants produced comparable DSB levels relative to wild-type (between 6 and 8% based on the global maxima of the curves), with a ∼60 min delay in *asf1*Δ and *hst3/4*Δ mutants (**Figure 2C**). The delay of DNA breaks in *asf1*Δ and *hst3/4*Δ cells suggests that DSBs appear and disappear with a slightly modified kinetics in these mutants. Maximal DSB reduction was observed in the *rtt109*Δ strain (1.5-fold reduction). Importantly, meiotic DSBs detected in the mutants were highly recombinogenic (R1/R2 recombinants are highlighted in **Figure 2B**) forming wild-type levels of crossovers (∼10–11%, **Figure 2C**). These results suggest that DSBs detected at *HIS4-LEU2* were properly processed to form mature recombination products.

To obtain mechanistic insights whether the modifiable H3K56 residue influences DSB formation, we constructed a plasmid shuffle system that allows the expression of wild-type H3K56 (H3 ctrl) and unmodifiable H3K56A as the only source of histone H3 (**Figure 3A**). Like wild-type cells, H3K56A mutants progressed synchronously through the meiotic S phase (**Figure 3B**) and showed normal transcriptome dynamics with no difference in gene expression during meiosis (**Figure 3C** and **Supplementary Table S1**). "Core"

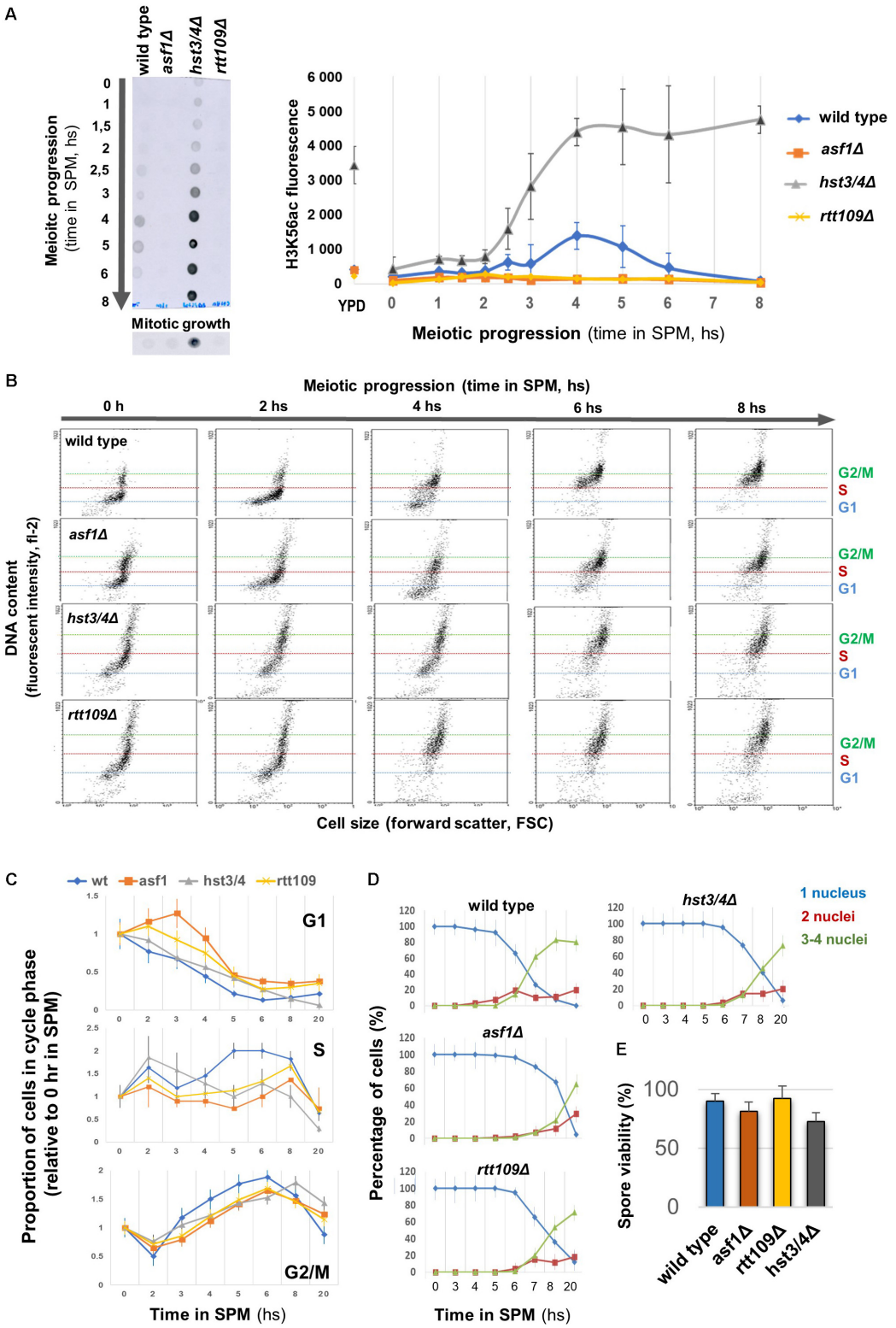


FIGURE 1 | Different H3K56ac levels allow normal meiotic progression and sporulation. **(A)** Left: Representative dot blot showing H3K56ac levels in wild type, *asf1*Δ, *rtt109*Δ, and *hst3/4*Δ cells. Right: Quantification of H3K56ac levels. YPD: mitotic growth. **(B)** Representative FACS profile of sporulating strains, showing DNA content in terms of cell size. G1, S, G2/M phases are indicated. **(C)** Proportion of cells in G1, S, and G2/M based on meiotic FACS profiles. Data were normalized to 0 h in SPM. **(D)** Sporulation efficiency. Nuclei were stained with DAPI and cells were scored for nucleus count. **(E)** Spore viability (50 tetrads per strain were counted). Panels **(A,C–E)** show the mean of three independent replicas. Error bars: SEM.

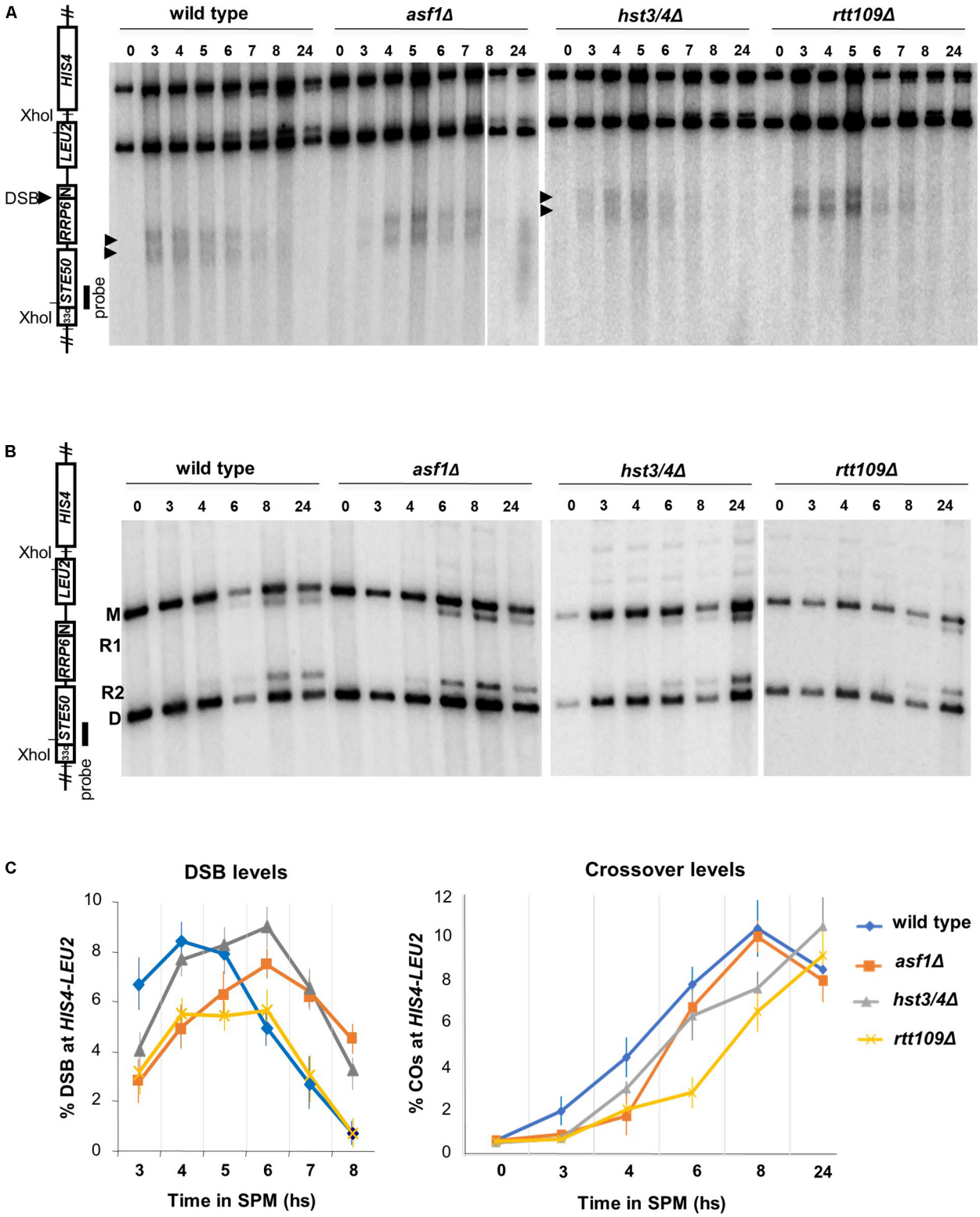


FIGURE 2 | Southern blot analysis of DSB and crossover formation at the *HIS4-LEU2* hotspot. **(A)** Representative blot showing *Xho*I-digested DNA from wild-type, *asf1*Δ, *hst3/4*Δ, and *rtt109*Δ cells. **(B)** Representative blot showing recombinant products (R1, R2) and parental bands (M, D) at *HIS4-LEU2*. **(C)** Quantification of DSB levels (% DSB signal/total lane signal) and recombination frequencies [% (R1 + R2) signal/total lane signal]. Mean values are from three independent replicas. Error bars: SEM.

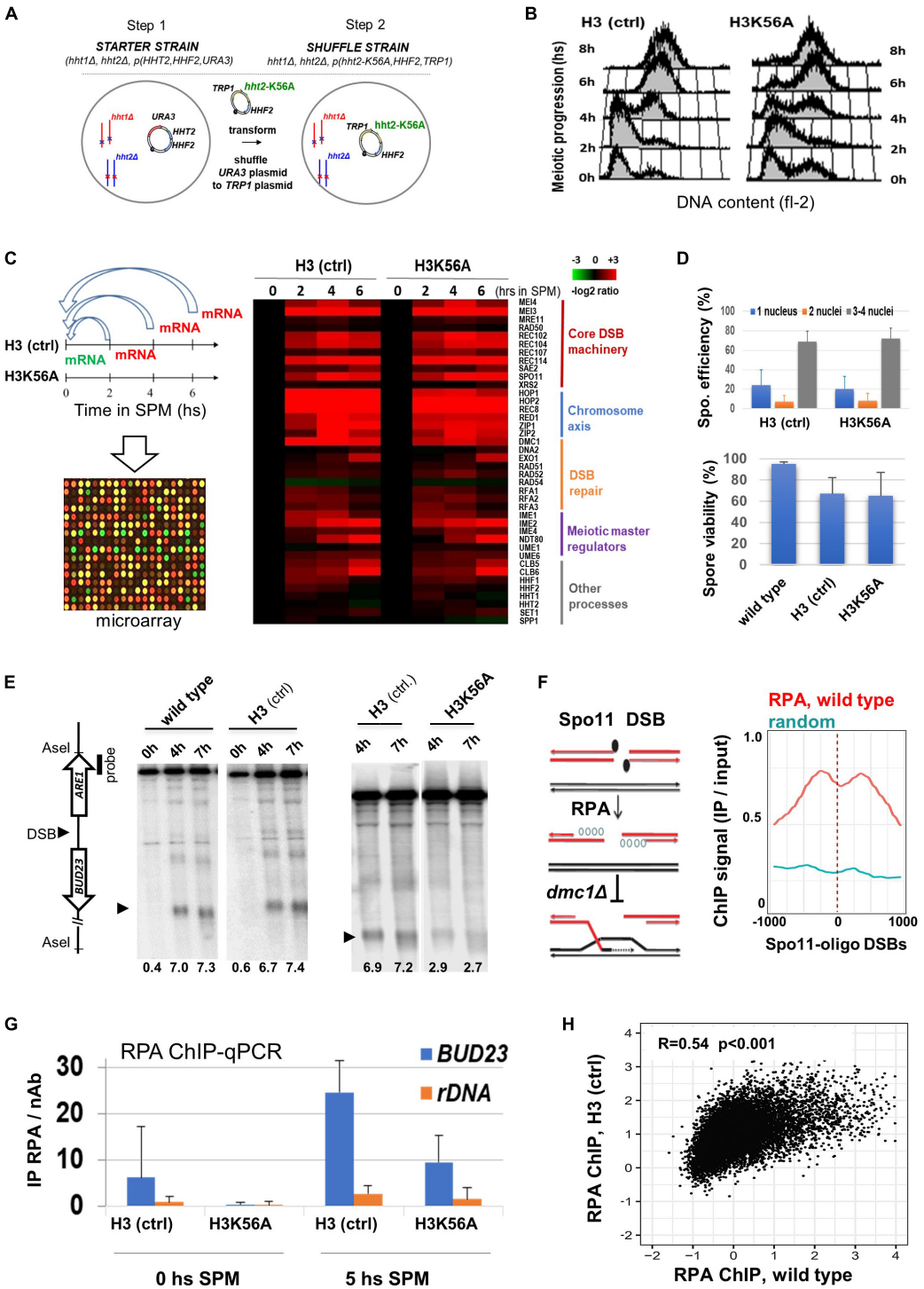


FIGURE 3 | Sporulation and meiotic DSB formation in a H3K56A mutant. **(A)** H3K56A cells were constructed by plasmid shuffling. **(B)** S phase progression in H3K56A and H3 (ctrl) cells measured by FACS. **(C)** Meiotic transcriptome dynamics in H3 (ctrl) and H3K56A strains. Left: Relative mRNA levels as measured by transcriptome microarrays at 2, 4, 6 h in SPM compared to 0 h in SPM. Right: Heatmap highlighting a selection of "core" genes that govern the process meiosis. The warmer the color, the greater the degree of induction. Full details are provided in **Supplementary Table S1**. **(D)** Sporulation efficiency and spore viability of H3K56A, H3 (ctrl), and wild-type strains. Error bars: SEM. **(E)** DSBs within *BUD23-ARE1*. Representative blot shows *Ase*I-digested gDNA from wild type, H3 (ctrl) and H3K56A cells. DSB levels (below the blots): % DSB signal/total lane signal. **(F)** Left: Outline of the RPA ChIP approach, showing resected ssDNAs covered by RPA, which accumulates in *dmc1*Δ cells. Right: RPA enrichment over Spo11-oligo DSB sites and random sites in wild-type SK1 cells. Wild-type RPA data are from Borde et al. (2009). **(G)** ChIP-qPCR analysis of RPA enrichment at 0 and 5 h in SPM in the *BUD23* hotspot and the *rDNA* negative control site (known to lack DSBs). Error bars: SEM. **(H)** Correlation of RPA ChIP profiles in H3 (ctrl) and wild-type cells. R: Pearson correlation coefficient.

dc_1977_21

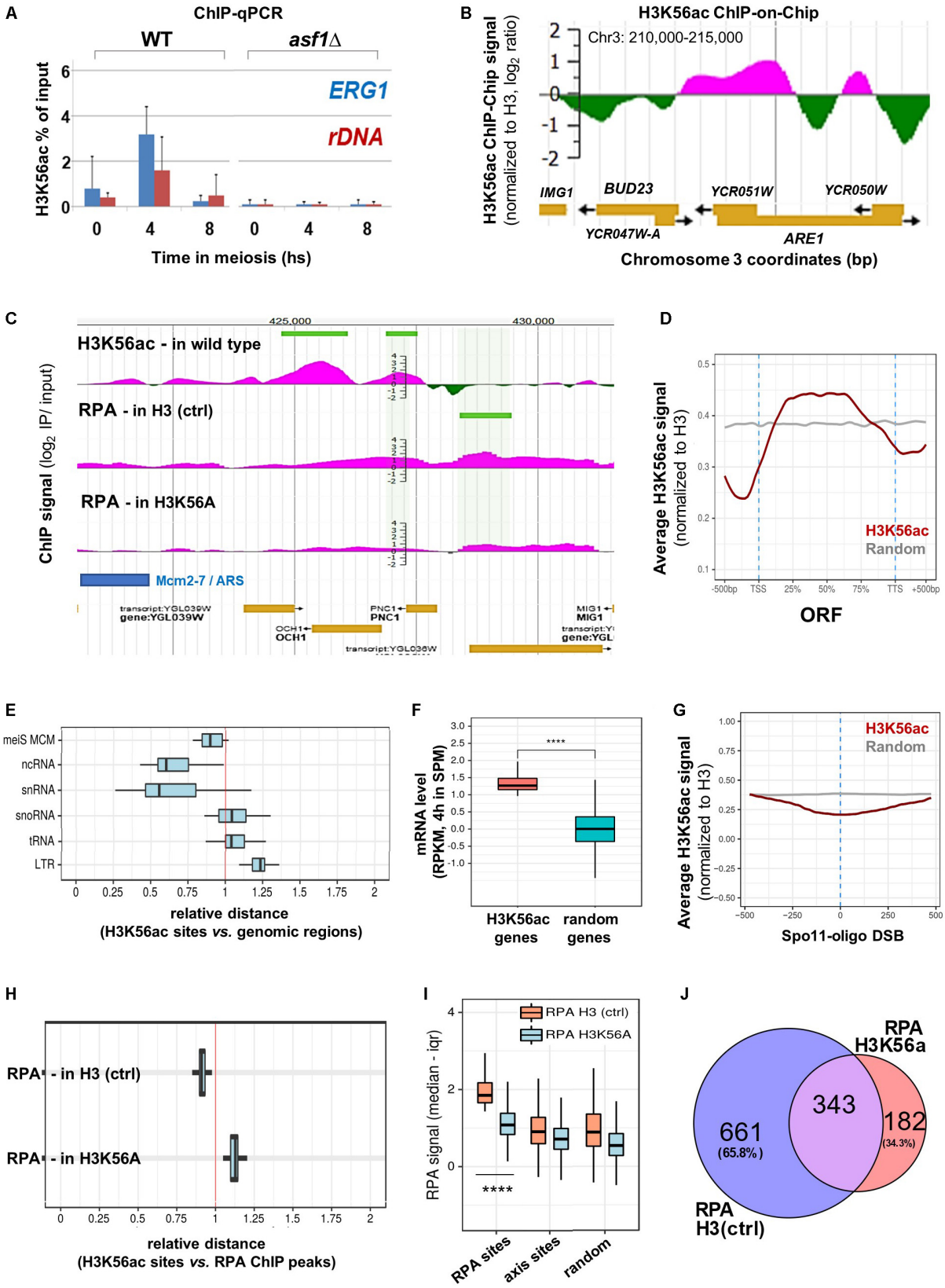


FIGURE 4 | Continued

FIGURE 4 | Genome-wide mapping of H3K56ac and RPA. **(A)** H3K56ac capture using an anti-H3K56ac antibody. H3 (C-ter) ChIP was performed to normalize to H3 occupancy. ChIP-qPCR enrichments were measured at *ERG1* and *rDNA*. Error bars: SEM. **(B)** H3K56ac signal detected by microarray (at 5 h in SPM). IP/input ratios were normalized to H3 occupancy (Borde et al., 2009). **(C)** Genome browser snapshot showing the distribution of H3K56ac in wild type cells and RPA enrichment in H3 (ctrl) and H3K56A cells. Peaks are highlighted by horizontal green lines. **(D)** Metagene profile of H3K56ac over protein-coding genes. Red curve shows the median H3K56ac signal. Gray curve: random signal. **(E)** Relative distance of H3K56ac peaks from various genomic regions. Vertical red line: random distances. Shift toward the left or right sectors: clustering or repulsion between H3K56ac peaks and the studied genomic elements. Mei MCM: meiotic replication origins (Blitzblau et al., 2012). **(F)** Genes associated with high H3K56ac levels show increased mRNA expression compared to random genes. Statistical significance: $p < 0.0001$ (Mann-Whitney *U* test). RNA-seq data are from Brar et al. (2012). **(G)** Depletion of H3K56ac over Spo11-oligo DSBs (Mohibullah and Keeney, 2017). Gray line: random signal. **(H)** Relative distance of H3K56ac peaks from RPA peaks in H3 (ctrl) and H3K56A cells. Vertical red line: random distance. **(I)** Distribution of RPA ChIP signal at RPA binding sites, axis sites (Mer2; Panizza et al., 2011), and random sites in H3 (ctrl) and H3K56A cells. Statistical significance: $p < 0.0001$ (Mann–Whitney *U* test). **(J)** Overlap of RPA peaks in H3 (ctrl) and H3K56A cells.

DSB, chromosome axis, and repair genes were properly transcribed in H3K56A cells, including *SPO11* and *RFA1* (**Supplementary Figure S1**), which were both essential for subsequent meiotic DSB mapping using RPA ChIP (see later). Sporulation efficiency and spore viability of the H3K56A mutant was not different from H3 (ctrl) expressing wild-type H3 (**Figure 3D**).

Next, we analyzed DSB formation in H3 *dmc1*Δ and H3K56A *dmc1*Δ cells that accumulate unrepaired DSB ends. DSB frequencies were determined by Southern blot within the natural hotspot region *BUD23-ARE1* (**Figure 3E**). Quantification of DSBs in wild-type and H3 (ctrl) strains confirmed the correct location, timing and frequency of DSBs in plasmid shuffle cells, demonstrating that our system accurately reports DSB formation. Importantly, a threefold reduction of DSB levels was observed in the H3K56A mutant, which was subsequently confirmed by an independent RPA ChIP method capturing Rfa1-covered ssDNA flanking Spo11-oligo DSBs (**Figures 3F–H**). In plasmid shuffle cells, RPA levels increased in the *BUD23* hotspot region by 5 h in SPM when DSBs are formed, and H3K56A mutants showed a ∼threefold decrease in RPA levels relative to H3 (ctrl) (**Figure 3G**), consistent with our Southern blot results (**Figure 3E**). In addition, the RPA profiles of H3 (ctrl) and wild-type cells (Borde et al., 2009) were positively correlated (**Figure 3H**), whereas the RPA binding sites did not correlate with the binding of Mcm2-7 replicative helicase that marks meiotic DNA replication (**Supplementary Figure S2**). These results collectively demonstrate that RPA enrichment is an adequate indicator of meiotic DSB locations and frequencies.

Notably, the *BUD23* and *ERG1* hotspot regions were flanked by H3K56ac at the time of DSB formation as measured by H3K56ac ChIP-qPCR and ChIP-chip (**Figures 4A,B**). We note that the H3K56ac ChIP signal was undetectable in *asf1*Δ cells that are deficient in H3K56 acetylation (**Figure 1A**), confirming the specificity of our ChIP assay. Furthermore, the H3K56ac signal was also depleted in wild-type cells within the recombinationally cold *rDNA* region (**Figure 4A**). The same associations were observed at the genomic scale (**Figure 4C** and *JBrowse* link), suggesting that the presence of H3K56 residue and/or deposition of H3K56-acetylated nucleosomes near DSB hotspots is required for complete DSB formation. Our H3K56ac ChIP results show that H3K56 acetylation is preferentially associated with protein-coding ORFs (**Figure 4D**) and is depleted from promoters and transcription termination sites [in line with published ChIP-qPCR data in mitotic cells (Schneider et al., 2006)]. The relative distance of H3K56ac from several genomic elements showed a clear non-random distribution (**Figure 4E**) as H3K56ac peaks were preferentially associated with Mcm2-7 helicase binding sites (Blitzblau et al., 2012), ncRNA and snRNA genes, but were further away from snoRNAs, tRNAs, and LTR retrotransposons compared to a computer-randomized chromosomal distribution (vertical red line in the figure). The proximity of H3K56ac peaks to Mcm2-7 binding sites agrees with the suggested role of H3K56 acetylation in marking nascent (replicating) chromatin upon replicative and repair DNA synthesis (Yu et al., 2012). Regarding meiotic gene expression (**Figure 4F**), protein-coding genes enriched in H3K56ac showed significantly higher mRNA expression levels in meiosis (at 4 h in SPM) than random genes [$p < 0.0001$; RNA-seq data are from Brar et al. (2012)]. Increased expression of H3K56-acetylated genes supports previous results in mitotic cells (Schneider et al., 2006; Williams et al., 2008), indicating that H3K56ac is an active chromatin mark associated with transcription.

Regarding meiotic DSB sites, chromosomal distribution of the H3K56ac signal showed a depletion over Spo11-oligo DSBs (**Figure 4G**), which precisely mark recombination hotspots. However, genomic positions of H3K56-acetylated histones were preferentially enriched adjacent to the identified RPA binding sites relative to random distance (**Figure 4H**, RPA H3 ctrl). Importantly, the RPA ChIP signal detected in the H3K56A mutant was strongly reduced at the RPA sites identified in the H3 (ctrl) strain ($p < 0.0001$, Mann–Whitney *U* test), as opposed to chromosome axis sites (Mer2-tagged; Panizza et al., 2011) and randomly selected sites that did not differ between control and mutant strains (**Figure 4I**). The preferential decrease of RPA ChIP signal at RPA binding sites in the H3K56A mutant demonstrates the specificity of H3K56A substitution mutation for recombination hotspot regions, accumulating hyper-resected ssDNA in the vicinity of DSBs. The specific decrease in the number of RPA peaks in the H3K56A mutant is highlighted in **Figure 4J**. This Venn diagram analysis shows that 65.8% of RPA peaks were eliminated by the H3K56A mutation (661 sites out of 1004), whereas 343 RPA sites were not affected or the signal was even increased (182 peaks). The latter RPA binding sites may represent unscheduled DSBs that are not related to H3K56 acetylation, which is clearly apparent from the increased relative distance between H3K56ac histones and RPA ChIP peaks detected in the H3K56A mutant (**Figure 4H**, RPA-H3K56A).

dc_1977_21

DISCUSSION

The above functional results highlight the association of H3K56 acetylation and meiotic DSB formation, suggesting that H3K56-acetylated histones are required to produce normal levels of DSBs within recombination hotspot regions. Nevertheless, the exact molecular mechanism underlying the DSB-promoting effect of H3K56 acetylation has yet to be clarified. The cause of the observed differences between the two mutant systems (histone modifying enzyme deletion vs. histone mutation) is currently not known. We obtained complementary results that are fully consistent with recent data identifying differences in the RNAPII binding profile of *rtt109*Δ and H3K56R mutants (Topal et al., 2019). Based on the genome-scale analysis of H3K56A cells, we propose that lack of H3K56 acetylation affects the stability or turnover rate of the well-positioned first nucleosomes flanking DSB sites (consistent with Kaplan et al. (2008)], and this may reduce the efficiency of Spo11-catalyzed DNA cleavage. Moreover, the absence of H3K56ac mark could reduce the rate of DSB end dissociation from nucleosomes flanking DSB sites, impeding the timely resection and processing of DSB ends. Alternatively, the H3K56A mutation may exert its effect indirectly on DSB formation, however, this is probably independent of changes in transcription since no differential gene expression was detected in the H3K56A mutant. A further possibility could be that H3K56 acetylation promotes the interaction of H3K4me3, Spp1, and Mer2 during the loop tethering process. The recently identified epistatic relationship of H3K56ac and H3K4me3 supports this hypothesis and seems particularly important in this regard (Voichek et al., 2018). H3K56ac was found to act upstream of Set1C and H3K4 methylation, generating complementary H3K56ac/H3K4me3 histone modification patterns along newly replicated chromatin. These functional relationships could readily allow close cooperation between H3K56ac and H3K4me3 during meiotic recombination, especially because newly replicated chromatin is rich in H3K56ac and DNA replication is mechanically coupled to meiotic DSB formation (Murakami and Keeney, 2014). We assume that these complex spatial interactions occur in the context of 3D chromatin structure. This could be detected by chromosome conformation capture methods (Dekker et al., 2017). Future use of these C-based approaches, together with mutant analyses, is expected to provide a deeper understanding of meiotic DSB formation with regards to the role of H3K4 methylation, H3K56 acetylation, and other potentially relevant histone modifications.

DATA AVAILABILITY STATEMENT

Datasets generated for this study can be accessed in **Supplementary Table S1** and via JBrowse (login: h3k56ac, password: Mozaic4, http://geneart.med.unideb.hu/pub/h3k56ac). Raw data are availabe at GEO (GSE37487).

AUTHOR CONTRIBUTIONS

ZK and LH analyzed the genomic data. LS performed the experiments and wrote the manuscript.

FUNDING

LS received funding from HAS-Lendület-LP2015-9/2015, NKFIH-NNE-130913, and GINOP-2.3.2-15-2016-00024.

ACKNOWLEDGMENTS

The authors thank Alain Nicolas (Institut Curie) for the conceptual support and critical comments.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fcell.2019.00364/full#supplementary-material

REFERENCES

Abshiru, N., Ippersiel, K., Tang, Y., Yuan, H., Marmorstein, R., Verreault, A., et al. (2013). Chaperone-mediated acetylation of histones by Rtt109 identified by quantitative proteomics. *J. Proteomics* 81, 80–90. doi: 10.1016/j.jprot.2012.09.026

Acquaviva, L., Székvölgyi, L., Dichtl, B., Dichtl, S. B., de La Roche Saint André, C., Nicolas, A., et al. (2013). The COMPASS subunit Spp1 links histone methylation to initiation of meiotic recombination. *Science* 339, 215–218. doi: 10.1126/science.1225739

Baudat, F., Buard, J., Grey, C., Fledel-Alon, A., Ober, C., Przeworski, M., et al. (2010). PRDM9 is a major determinant of meiotic recombination hotspots in humans and mice. *Science* 327, 836–840. doi: 10.1126/science.1183439

Biswas, M., Voltz, K., Smith, J. C., and Langowski, J. (2011). Role of histone tails in structural stability of the nucleosome. *PLoS Comput. Biol.* 7: e1002279. doi: 10.1371/journal.pcbi.1002279

Blitzblau, H. G., Chan, C. S., Hochwagen, A., and Bell, S. P. (2012). Separation of DNA replication from the assembly of break-competent meiotic chromosomes. *PLoS Genet.* 8:e1002643. doi: 10.1371/journal.pgen.1002643

Borde, V., Robine, N., Lin, W., Bonfils, S., Géli, V., and Nicolas, A. (2009). Histone H3 lysine 4 trimethylation marks meiotic recombination initiation sites. *EMBO J.* 28, 99–111. doi: 10.1038/emboj.2008.257

Brar, G. A., Yassour, M., Friedman, N., Regev, A., Ingolia, N. T., and Weissman, J. S. (2012). High-resolution view of the yeast meiotic program revealed by ribosome profiling. *Science* 335, 552–557. doi: 10.1126/science.1215110

Buning, R., and Van Noort, J. (2010). Single-pair FRET experiments on nucleosome conformational dynamics. *Biochimie* 92, 1729–1740. doi: 10.1016/j.biochi.2010.08.010

Celic, I., Masumoto, H., Griffith, W. P., Meluh, P., Cotter, R. J., Boeke, J. D., et al. (2006). The Sirtuins Hst3 and Hst4p preserve genome integrity by controlling Histone H3 Lysine 56 Deacetylation. *Curr. Biol.* 16, 1280–1289. doi: 10.1016/j.cub.2006.06.023

Dekker, J., Belmont, A. S., Guttman, M., Leshyk, V. O., Lis, J. T., Lomvardas, S., et al. (2017). The 4D nucleome project. *Nature* 549, 219–226. doi: 10.1038/nature23884

Govin, J., Dorsey, J., Gaucher, J., Rousseaux, S., Khochbin, S., and Berger, S. L. (2010). Systematic screen reveals new functional dynamics of histones H3 and H4 during gametogenesis. *Genes Dev.* 24, 1772–1786. doi: 10.1101/gad.1954910

dc_1977_21

Kaplan, T., Liu, C. L., Erkmann, J. A., Holik, J., Grunstein, M., Kaufman, P. D., et al. (2008). Cell cycle- and chaperone-mediated regulation of H3K56ac incorporation in yeast. *PLoS Genet.* 4:e1000270. doi: 10.1371/journal.pgen.1000270

Karányi, Z., Halász, L., Acquaviva, L., Jónás, D., Hetey, S., Boros-Oláh, B., et al. (2018). Nuclear dynamics of the Set1C subunit Spp1 prepares meiotic recombination sites for break formation. *J Cell Biol.* 217, 3398–3415. doi: 10.1083/jcb.201712122

Keeney, S., Giroux, C. N., and Kleckner, N. (1997). Meiosis-specific DNA double-strand breaks are catalyzed by Spo11, a member of a widely conserved protein family. *Cell* 88, 375–384. doi: 10.1016/s0092-8674(00)81876-0

Luger, K., Rechsteiner, T., Flaus, A., Waye, M., and Richmond, T. J. (1997). Characterization of nucleosome core particles containing histone proteins made in bacteria. *J. Mol. Biol.* 272, 301–311. doi: 10.1006/jmbi.1997.1235

Messiaen, S., Guiard, J., Aigueperse, C., Fliniaux, I., Tourpin, S., Barroca, V., et al. (2016). Loss of the histone chaperone ASF1B reduces female reproductive capacity in mice. *Reproduction* 151, 477–489. doi: 10.1530/REP-15-0327

Mohibullah, N., and Keeney, S. (2017). Numerical and spatial patterning of yeast meiotic DNA breaks by Tel1. *Genome Res.* 27, 278–288. doi: 10.1101/gr.213587.116

Murakami, H., and Keeney, S. (2014). Temporospatial coordination of meiotic dna replication and recombination via DDK recruitment to replisomes. *Cell* 158, 861–873. doi: 10.1016/j.cell.2014.06.028

Pan, J., Sasaki, M., Kniewel, R., Murakami, H., Blitzblau, H. G., Tischfield, S. E., et al. (2011). A hierarchical combination of factors shapes the genome-wide topography of yeast meiotic recombination initiation. *Cell* 144, 719–731. doi: 10.1016/j.cell.2011.02.009

Panizza, S., Mendoza, M. A., Berlinger, M., Huang, L., Nicolas, A., Shirahige, K., et al. (2011). Spo11-accessory proteins link double-strand break sites to the chromosome axis in early meiotic recombination. *Cell* 146, 372–383. doi: 10.1016/j.cell.2011.07.003

Parpanov, E. D., Petkov, P. M., and Paigen, K. (2010). Prdm9 controls activation of mammalian recombination hotspots. *Science* 327:835. doi: 10.1126/science.1181495

Recht, J., Tsubota, T., Tanny, J. C., Diaz, R. L., Berger, J. M., Zhang, X., et al. (2006). Histone chaperone Asf1 is required for histone H3 lysine 56 acetylation, a modification associated with S phase in mitosis and meiosis. *Proc. Natl. Acad. Sci. U.S.A.* 103, 6988–6993. doi: 10.1073/pnas.0601676103

Rufiange, A., Jacques, P. E., Bhat, W., Robert, F., and Nourani, A. (2007). Genome-wide replication-independent histone H3 exchange occurs predominantly at promoters and implicates H3 K56 acetylation and Asf1. *Mol. Cell* 27, 393–405. doi: 10.1016/j.molcel.2007.07.011

Schneider, J., Bajwa, P., Johnson, F. C., Bhaumik, S. R., and Shilatifard, A. (2006). Rtt109 is required for proper H3K56 acetylation: a chromatin mark associated with the elongating RNA polymerase II. *J. Biol. Chem.* 281, 37270–37274. doi: 10.1074/jbc.C600265200

Simon, M., North, J. A., Shimko, J. C., Forties, R. A., Ferdinand, M. B., Manohar, M., et al. (2011). Histone fold modifications control nucleosome unwrapping and disassembly. *Proc. Natl. Acad. Sci. U.S.A.* 108, 12711–12716. doi: 10.1073/pnas.1106264108

Simoneau, A., Delgoshaie, N., Celic, I., Dai, J., Abshiru, N., Costantino, S., et al. (2015). Interplay between histone H3 lysine 56 deacetylation and chromatin modifiers in response to DNA damage. *Genetics* 200, 185–205. doi: 10.1534/genetics.115.175919

Sommermeyer, V., Beneut, C., Chaplais, E., Serrentino, M. E., and Borde, V. (2013). Spp1, a member of the Set1 Complex, promotes meiotic DSB formation in promoters by tethering histone H3K4 methylation sites to chromosome axes. *Mol. Cell* 49, 43–54. doi: 10.1016/j.molcel.2012.11.008

Szekvolgyi, L., and Nicolas, A. (2010). From meiosis to postmeiotic events: homologous recombination is obligatory but flexible. *FEBS J.* 277, 571–589. doi: 10.1111/j.1742-4658.2009.07502.x

Székvölgyi, L., and Nicolas, A. (2010). From meiosis to postmeiotic events: homologous recombination is obligatory but flexible. *FEBS J.* 277, 571–589. doi: 10.1111/j.1742-4658.2009.07502.x

Székvölgyi, L., Ohta, K., and Nicolas, A. (2015). Initiation of meiotic homologous recombination: flexibility, Impact of Histone modifications, and chromatin remodelling. *Cold Spring Harb. Perspect. Biol.* 7, 447–462. doi: 10.1101/cshperspect.a016527

Tian, H., Billings, T., and Petkov, P. M. (2018). CXXC1 is redundant for normal DNA double-strand break formation and meiotic recombination in mouse. *PLoS Genet.* 14:e1007657. doi: 10.1371/journal.pgen.1007657

Topal, S., Vasseur, P., Radman-Livaja, M., and Peterson, C. L. (2019). Distinct transcriptional roles for Histone H3-K56 acetylation during the cell cycle in Yeast. *Nat. Commun.* 10, 1–13. doi: 10.1038/s41467-019-12400-5

Tsubota, T., Berndsen, C. E., Erkmann, J. A., Smith, C. L., Yang, L., Freitas, M. A., et al. (2007). Histone H3-K56 acetylation is catalyzed by histone chaperone-dependent complexes. *Mol. Cell* 25, 703–712. doi: 10.1016/j.molcel.2007.02.006

Voichek, Y., Mittelman, K., Gordon, Y., Bar-Ziv, R., Lifshitz Smit, D., Shenhav, R., et al. (2018). Epigenetic control of expression homeostasis during replication is stabilized by the replication checkpoint. *Mol. Cell* 70, 1121.e9–1133.e9. doi: 10.1016/j.molcel.2018.05.015

Watanabe, S., Radman-Livaja, M., Rando, O. J., and Peterson, C. L. (2013). A histone acetylation switch regulates H2A.Z deposition by the SWR-C remodeling enzyme. *Science* 340, 195–199. doi: 10.1126/science.1229758

White, C. L., Suto, R. K., and Luger, K. (2001). Structure of the yeast nucleosome core particle reveals fundamental changes in internucleosome interactions. *EMBO J.* 20, 5207–5218. doi: 10.1093/emboj/20.18.5207

Williams, S. K., Truong, D., and Tyler, J. K. (2008). Acetylation in the globular core of histone H3 on lysine-56 promotes chromatin disassembly during transcriptional activation. *Proc. Natl. Acad. Sci. U.S.A.* 105, 9000–9005. doi: 10.1073/pnas.0800057105

Xu, L., and Kleckner, N. (1995). Sequence non-specific double-strand breaks and interhomolog interactions prior to double-strand break formation at a meiotic recombination hot spot in yeast. *EMBO J.* 14, 5115–5128. doi: 10.1002/j.1460-2075.1995.tb00194.x

Yu, Y., Song, C., Zhang, Q., DiMaggio, P. A., Garcia, B. A., York, A., et al. (2012). Histone H3 lysine 56 methylation regulates DNA replication through its interaction with PCNA. *Mol. Cell* 46, 7–17. doi: 10.1016/j.molcel.2012.01.019

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2020 Karányi, Hornyák and Székvölgyi. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

Downloaded from http://cshperspectives.cshlp.org/ on October 26, 2021 - Published by Cold Spring Harbor Laboratory Press

dc_1977_21

Initiation of Meiotic Homologous Recombination: Flexibility, Impact of Histone Modifications, and Chromatin Remodeling

Lóránt Székvölgyi[1], Kunihiro Ohta[2], and Alain Nicolas[3]

[1]Department of Biophysics and Cell Biology, Faculty of Medicine, University of Debrecen, 4032 Debrecen, Hungary

[2]Department of Life Sciences, The University of Tokyo, 113-8654 Tokyo, Japan

[3]Institut Curie Centre de Recherche, UMR3244 CNRS, Université Pierre et Marie Curie, 75248 Paris CEDEX 05, France

Correspondence: lorantsz@med.unideb.hu; alain.nicolas@curie.fr

Cold Spring Harbor Perspectives in Biology
www.cshperspectives.org

Meiotic recombination is initiated by the formation of DNA double-strand breaks (DSBs) catalyzed by the evolutionary conserved Spo11 protein and accessory factors. DSBs are nonrandomly distributed along the chromosomes displaying a significant (~400-fold) variation of frequencies, which ultimately establishes local and long-range "hot" and "cold" domains for recombination initiation. This remarkable patterning is set up within the chromatin context, involving multiple layers of biochemical activity. Predisposed chromatin accessibility, but also a range of transcription factors, chromatin remodelers, and histone modifiers likely promote local recruitment of DSB proteins, as well as mobilization, sliding, and eviction of nucleosomes before and after the occurrence of meiotic DSBs. Here, we assess our understanding of meiotic DSB formation and methods to change its patterning. We also synthesize current heterogeneous knowledge on how histone modifications and chromatin remodeling may impact this decisive step in meiotic recombination.

Sexual reproduction depends on halving the genome content of germ line cells and faithful chromosome transmission during meiosis to yield viable gametes. Meiosis comprises one round of DNA replication and two successive rounds of chromosome segregation, allowing the reduction of a diploid genome to produce haploid gametes (Fig. 1A).

Central to meiosis is the process of recombination between the paternal and maternal chromosomes (interhomolog recombination), which is crucial to enhance the genetic diversity of the gametes, but also for providing physical connections among homologs. These connections (i.e., chiasmata) ensure proper alignment of homologous chromosome pairs on the spindle, promoting proper reductional segregation following the regulated release of the sister chromatid cohesion among the duplicated chromosomal arms (Watanabe 2012). Defective meiotic recombination is a source of de novo germline mutations, abnormal genome content in gametes (the source of Down's syndrome), and infertility. Not surprisingly, cells have developed a variety of mechanisms and tight controls to ensure sufficient and well-dis-

Editors: Stephen Kowalczykowski, Neil Hunter, and Wolf-Dietrich Heyer
Additional Perspectives on DNA Recombination available at www.cshperspectives.org

Copyright © 2015 Cold Spring Harbor Laboratory Press; all rights reserved; doi: 10.1101/cshperspect.a016527
Cite this article as *Cold Spring Harb Perspect Biol* 2015;7:a016527

Downloaded from http://cshperspectives.cshlp.org/ on October 26, 2021 - Published by Cold Spring Harbor Laboratory Press

Cold Spring Harbor Perspectives in Biology
www.cshperspectives.org

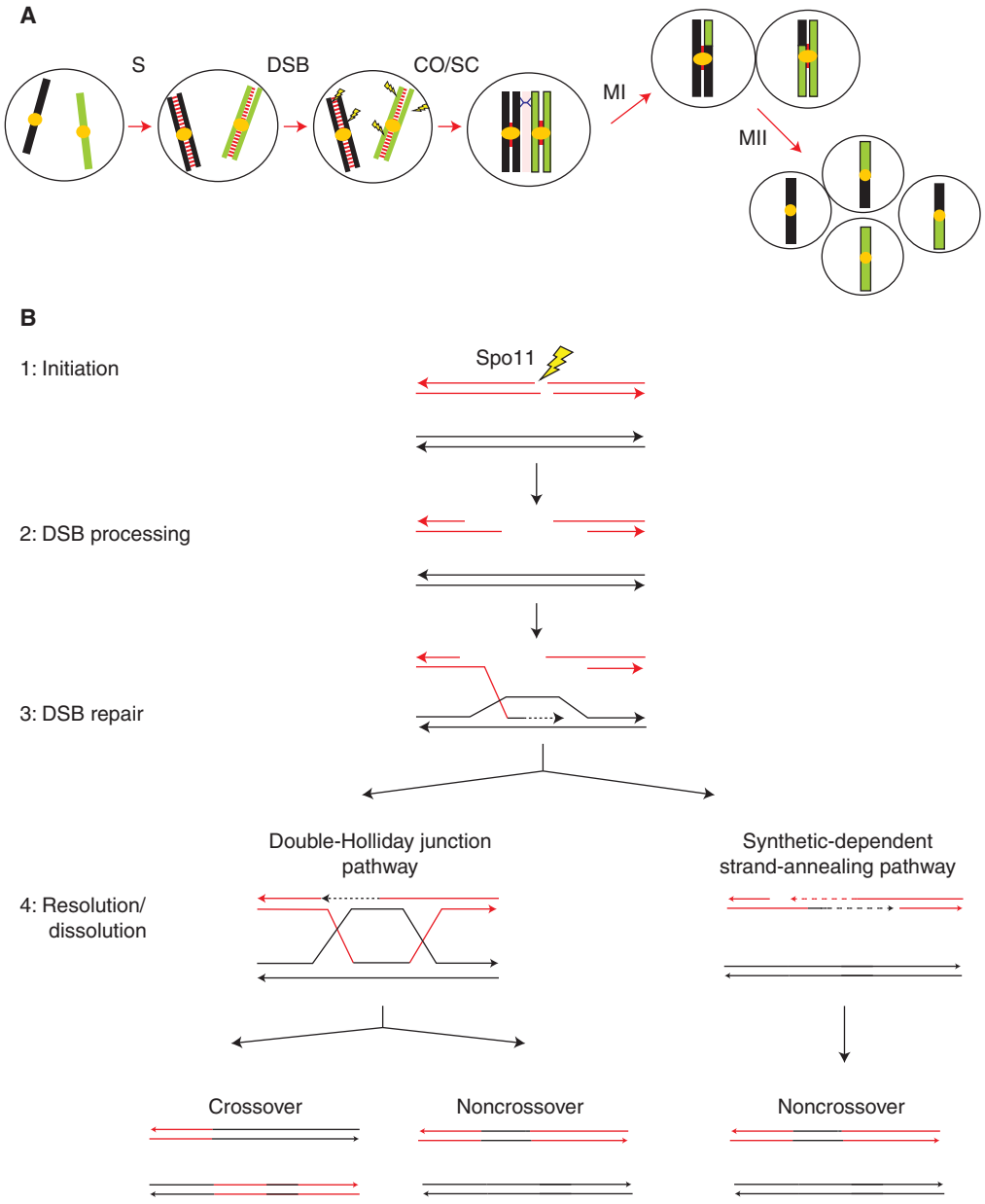


Figure 1. The stages and mechanisms of meiosis. (*A*) Diploid yeast cells initiate meiosis on nutrient depletion and in the presence of a nonfermentable carbon source. In mammals, the process is started by endocrine/paracrine/juxtacrine stimuli from the surrounding cell and tissue environment. Relevant molecular stages are indicated: S, meiotic replication; DSB, double-strand breaks; CO, crossover; SC, synaptonemal complex; MI, and MII (first and second meiotic divisions, respectively). Colors: parental homologous chromosomes (green and black), sister chromatid cohesion (red), SC (pink). (*B*) The mechanism of meiotic recombination. Stage 1: Initiation. DSBs are introduced by the Spo11 protein. Stage 2: DSB processing. Strand resection initiates to yield 3′ single-stranded DNA (ssDNA) overhangs. One of the 3′ ssDNA tails engages in strand invasion and a homology search of the homologous chromosome, resulting in single-end invasion (SEI) and D-loop intermediates. Stage 3: DSB repair. (*Legend continues on following page.*)

Downloaded from http://cshperspectives.cshlp.org/ on October 26, 2021 - Published by Cold Spring Harbor Laboratory Press

dc_1977_21

Cold Spring Harbor Perspectives in Biology
www.cshperspectives.org

tributed meiotic recombination events within their genomes.

At the DNA level, meiotic recombination can be divided in four successive stages (Fig. 1B): (1) initiation, which consists in the formation of programmed DNA DSBs; (2) DSB processing, which yields the recombinogenic single-strand tails; (3) homologous DSB repair, which involves the homologous recombination pathway and several meiosis-specific differentiation modulations, which facilitate interhomolog interactions; and (4) intermediate resolution and dissolution, which allow the formation of local (up to a few kb in length) noncrossover (NCO) gene conversion events and reciprocal crossovers (COs).

Recently, local and genome-wide studies of normal and mutant cells have uncovered a remarkable variability in the number and positioning of recombination events per chromosome and cell, which reveals an impressive level of flexibility (Székvölgyi and Nicolas 2010). In this review, we outline our understanding of the control of the initiation events, and how histone modifications and chromatin remodeling impact this initial step of meiotic recombination. The process of meiotic recombination and its relationship to change in chromosome structures and movements, allowing homolog alignment, pairing, and synapsis, is also reviewed in Zickler and Kleckner (2015) and Lam and Keeney (2015).

MEIOTIC DSBs AND RECOMBINATION EVENTS ARE NONRANDOMLY DISTRIBUTED

In all organisms, meiotic recombination is initiated by the formation of a large number of programmed DNA DSBs per cell, which are repaired primarily via recombination between nonsister chromatids (recombination initiation occurs after replication) to generate NCO and CO recombinant products (Fig. 1B). These DSBs are catalyzed by the evolutionarily conserved Spo11 protein (Bergerat et al. 1997). Besides Spo11, a number of accessory proteins are required for DSB formation. In *Saccharomyces cerevisiae*, 10 DSB proteins are known and functional, but sometimes sequence-divergent orthologs are gradually being identified in other organisms, including mammals. Yeast two-hybrid and co-immunoprecipitation (IP) studies identified several multiprotein subcomplexes, which are successively recruited to the meiotic chromosomes before DSB formation in triggered (Arora et al. 2004; Panizza et al. 2011). The biochemical/structural functions of these DSB protein complexes remains poorly understood. They play a role to select the potential DSB regions along the chromosomes, and ultimately contribute to the recruitment of Spo11 and trigger cleavage.

Spo11 is orthologous to the topoVI family of topoisomerase discovered in archaea, and consistently introduces DSBs by coupled transesterification reactions to form covalent tyrosyl-DNA linkages at the 5′ termini of the broken DNA. Spo11 is then removed by endonucleolytic cleavage (Neale et al. 2005), liberating short Spo11-DNA oligonucleotide complexes and resected strands, which are further extended to generate recombinogenic 3′ single-stranded tails. Over the years, meiotic DSBs have been mapped and quantified in yeast genomic DNA using a variety of approaches, including Southern blot analysis of chromosomal fragments

Figure 1. (*Continued*) In the double-Holliday junction (dHJ) pathway, the opposite DSB end is captured by annealing to the displaced strand of the D-loop, leading to the formation of a dHJ. In the synthesis-dependent strand-annealing (SDSA) pathway, repair of DSBs occurs without the formation of a dHJ. Stage 4: Resolution/dissolution. In the dHJ pathway, after gap-filling DNA synthesis and nick ligation, the dHJ is cleaved on opposing single DNA strands, generating products that can be ligated. Depending on cleavage, the patterns dHJ resolution produces either CO recombinants (associated or not with gene conversion that results from HJ migration) or noncrossover (NCO) (gene conversions). In the SDSA pathway, the SEI intermediate undergoes DNA synthesis by extension of the invading DNA strand with D-loop dissolution, and the extended ssDNA ultimately reanneals to its original complementary ssDNA strand on the opposite side of the DSB. An intact duplex is then produced by gap-filling DNA synthesis and nick ligation, which gives rise to NCO recombinants.

Downloaded from http://cshperspectives.cshlp.org/ on October 26, 2021 - Published by Cold Spring Harbor Laboratory Press

Cold Spring Harbor Perspectives in Biology
www.cshperspectives.org

or full-length chromosomes, PCR, chromatin immunoprecipitation (ChIP)-Chip, and ChIP-seq analyses of enriched single-stranded DNA (ssDNA) (Buhler et al. 2007) and, most recently, via the purification and sequencing of Spo11-associated oligos, providing exquisite resolution both in terms of mapping and quantification (Pan et al. 2011). The most amenable method to estimate the number of DSBs on a per-cell basis is to count the number of Rad51/Dmc1 immunostaining foci on spread meiotic nuclei (Bishop 1994). On a broad scale, these methods have provided consistent general conclusions. They show the evolutionary conservation of the mechanism of DSB formation and the essential role of interhomolog recombination for the production of viable, euploid meiotic products. Depending on the organism and mutant defect, a lack or reduction in DSB frequency can lead to arrest of meiotic progression and apoptosis, or progression and formation of inviable aneuploid gametes.

To address whether Spo11-catalyzed DSBs play a unique role in meiosis, the functionality of DSBs introduced by other nucleases and clastogenic agents has been examined. Interhomolog recombination was promoted by ionizing radiation (Bowring et al. 2006), indicating that Spo11-catalyzed DSBs are not unique in their ability to promote recombination during meiosis. Other studies used the expression of a heterologous endonuclease, such as HO (Kolodkin et al. 1986; Malkova et al. 2000), VDE (Hodgson et al. 2011), or I-SceI (Farah et al. 2009) in yeast, and the Mos1 transposase in *Caenorhabditis elegans* (Robert and Bessereau 2007). In all cases, interhomolog recombination was induced at the break sites. However, in several respects, the DSBs and recombinants induced by these systems are different than those resulting from Spo11. These site-specific DSBs are not bona fide Spo11 breaks because of (1) the uncontrolled timing of DSB induction, (2) the abnormal and dangerous cleavage of both sister chromatids, and (3) alteration in recombination efficiencies when assayed in the wild-type or Spo11-deficient strain background. This relates to the role of the Spo11 DSBs in the pairing of the homologs. The uniqueness of the Spo11 breaks might reside in its mode of cleavage that allows, like in other site-specific recombination processes, to intimately link break formation and processing, and also to avoid extensive DSB signaling associated with accidental DSBs (Borde et al. 2004). Furthermore, Spo11 break formation is intimately linked to the process of chromosome pairing that enforces interhomolog rather than intersister recombinational repair.

Extensive studies with *S. cerevisiae* (Baudat and Nicolas 1997; Gerton et al. 2000) and later with *Schizosaccharomyces pombe* (Cromie et al. 2007), *Arabidopsis* (Drouaud et al. 2013), mice (Smagulova et al. 2011), chimpanzee, and humans (Myers et al. 2010; Auton et al. 2012) showed that the distribution of meiotic DSBs is not random along chromosomes, explaining the longtime noted discrepancy between physical and genetic map distances. Nowadays, genetic distances are measured by high-throughput microarray or genome-wide sequencing analyses of meiotic progenies, starting from diploid heterozygous strains carrying thousands of single-nucleotide polymorphisms (SNPs). In a pioneer study in *S. cerevisiae*, Mancera et al. (2008) genotyped the four spores of several tetrads to reconstitute the meiotic recombination events per meiotic cell and, on average, found ~90 COs and 46 NCOs (66 after correction for undetected NCO) per meiosis. Consistently, these recombination events are located in correlation with the preferred sites of DSBs and, in a number, suggesting that a large majority of the DSBs (estimated to ~150 per meiosis) are repaired on the homologous chromosome. This is consistent with the single-site genetic and physical analysis of recombination intermediates at the most studied artificial *HIS4::LEU2* hotspot, indicating that ~80% of the meiotic DSBs are repaired using a nonsister chromatid rather than the sister chromatid as template (Lao et al. 2013). Genome-wide analysis of meiotic products in other organisms confirms that most DSBs are repaired by interhomolog recombination (Cole et al. 2010), but the large excess of DSBs relative to COs—identified directly or by counting the Rad51 and Dmc1 foci on meiotic prophase spreads over the final recombination products—can be explained also if meiotic DSB

Downloaded from http://cshperspectives.cshlp.org/ on October 26, 2021 - Published by Cold Spring Harbor Laboratory Press

Cold Spring Harbor Perspectives in Biology
www.cshperspectives.org

repair frequently occurs among sister chromatids. This raises the key question of how and to what extent the interhomolog bias is implemented and, globally, to what extent the meiotic versus mitotic differentiation of DSB repair template choice is organism specific.

MULTIPLE LAYERS OF CONTROL SHAPE THE DSB DISTRIBUTION: CLUSTERING, INTERFERENCE, AND REDISTRIBUTION

If recombination-initiation sites were randomly distributed, DSBs would be equally likely to occur at any location along the chromosomes and would not influence one another (Fig. 2A). This is clearly not the case because (1) DSB frequencies vary greatly from site to site (at least 400-fold in *S. cerevisiae*, *S. pombe*, and *Arabidopsis*) (Pan et al. 2011; Choi et al. 2013; Fowler et al. 2014). (2) DSB sites within a hotspot region are strongly localized. In *S. cerevisiae*, they are preferentially localized in intergenic promoter-containing intervals, but depleted from other regions (Baudat and Nicolas 1997). In *S. pombe* and mice, enhanced DSB formation occurs in various regions, but also alternate with regions showing low DSB activity. (3) Over longer distances, DSB-prone and -repressed regions are clustered in subchromosomal domains. The cold regions are found in interstitial coding regions, showing no specific pattern along the chromosomes, but DSB formation is always strongly suppressed in the telomere- and centromere-proximal regions, as well as within the recombinant DNA (rDNA). These quantitative and region-specific variations are indeed strong evidence for the nonrandom nature of the DSB distribution. The contribution of rare and, possibly, random (stochastic) DSBs is not excluded, but remains difficult to measure and map.

Figure 2A illustrates two hypothetical nonrandom modes of DSB distribution: (1) a uniform distribution, when every DSB falls as far from its neighbors as possible, and (2) clustered, when most DSBs are concentrated close together and large regions contain very few, if any, breaks. The observed DSB pattern in *S. cerevisiae* (Borde et al. 2009) illustrated for chromosome III falls between these two extremes; with some deviation, it falls closer to the clustered arrangement, indicating that the contribution of stochasticity might be significant. It is noteworthy that the current genome-wide DSB-mapping techniques (Spo11-oligo sequencing, Dmc1/Rpa/ssDNA ChIP) obscure cell-to-cell variations in DSB formation (as do all mass-biochemical methods that investigate cell populations).

Other strong driver elements that shape the nonrandom distribution of DSBs have been documented in *S. cerevisiae*, but also remain poorly understood. These phenomena are referred to as "DSB interference" and appear to drive even spacing of DSBs and limit simultaneous breakage at allelic positions. First, the occurrence of DSBs reduces the probability of nearby DSBs to occur (Fig. 2B). This is a "*cis*-inhibition" process, in which strong DSB hotspots suppress the activity of nearby recombination-initiation sites on the same homolog (Xu and Kleckner 1995; Fan et al. 1997). This inhibitory effect of natural or targeted DSBs on adjacent hotspots is known to spread over significant distances along the chromosomes (up to 25–100 kbp in *S. cerevisiae*) (Robine et al. 2007). The second interference phenomenon is that DSB formation on one homolog decreases DSB formation on the other homolog (*trans*-inhibition) at the cognate allelic position (Rocco et al. 1992; Zhang et al. 2011a). Whether these *cis*- and *trans*-regulatory events occur independently remains to be elucidated, but, importantly, the *trans*-modulation appears to be genetically controlled, involving the two signal transduction kinases Mec1 (ATR) and Tel1 (ATM) acting as potential direct effectors of the *trans*-DSB interference (Zhang et al. 2011a; Blitzblau and Hochwagen 2013; Gray et al. 2013). The benefit to locally regulate DSB formation between the sister and nonsister chromatids is to prevent simultaneous DSB formation at the same place on more than one chromatid. This would generate poorly interacting broken molecules, perturb the efficiency of DSB repair, and, finally, facilitate mutagenic end-joining and out-of-register interactions between repeats, as well as genome rearrangements on ectopic interactions. Rare but naturally occurring events of this type may contribute to the arising of de novo germline mutations, a

Downloaded from http://cshperspectives.cshlp.org/ on October 26, 2021 - Published by Cold Spring Harbor Laboratory Press

Cold Spring Harbor Perspectives in Biology
www.cshperspectives.org

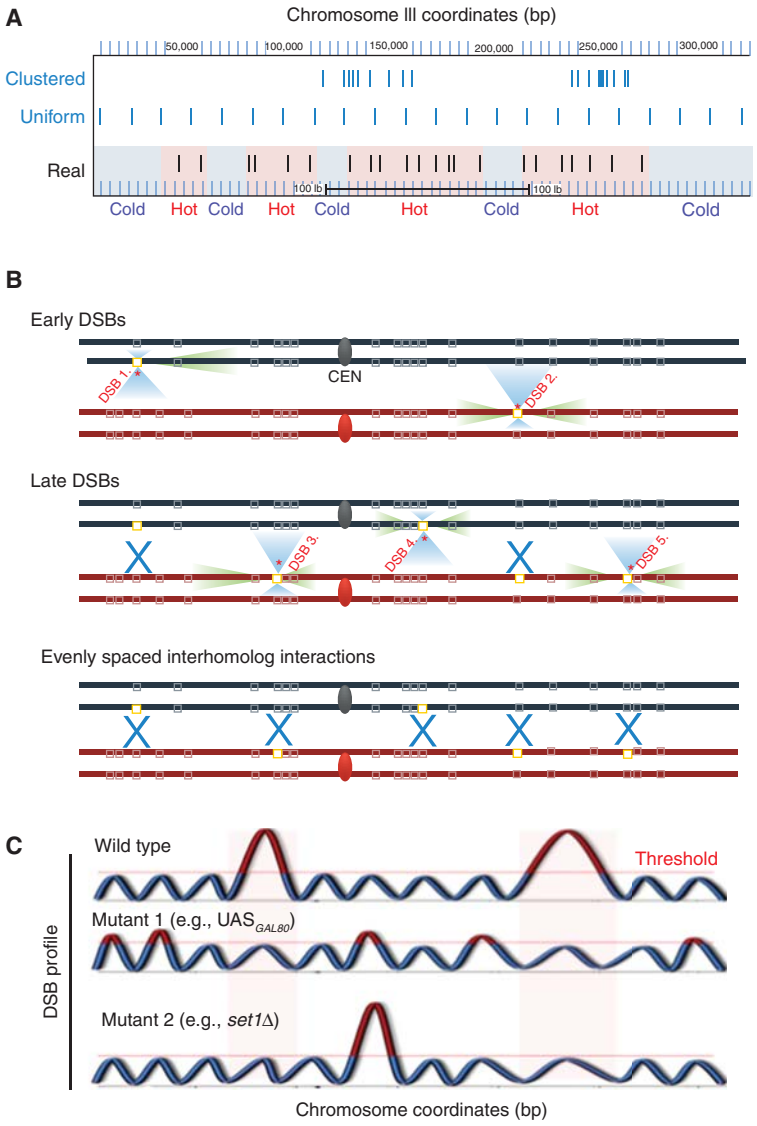


Figure 2. Spatial distribution of meiotic DNA double-strand breaks (DSBs) along the chromosomes. (*A*) Hypothetical (blue) and real (black) DSB patterns, representing random and nonrandom distributions. Real DSBs (Borde et al. 2009) establish "cold" and "hot" domains, which alternate along the chromosomes length. (*B*) DSB interference establishes even spacing among interhomolog crossovers (COs). Interference nucleates at activated DSB sites (yellow rectangles) cleaved by Spo11 (red star) and spreads over significant distances along the sister/nonsister chromatids or homologous chromosome (*trans*-inhibition, blue triangles), or along the same sister chromatid (*cis*-inhibition, green triangles). DSBs form sequentially as meiosis progresses such that "late" DSBs fill in the gaps among "early" DSBs. A subset of DSBs matures into COs (blue crosses). (*C*) DSB redistribution can shuffle recombination-initiation events to reinforce the "obligatory" number of COs per chromosome. Mutant 1 can be, for example, an upstream activating sequence $(UAS)_{GAL80}$, mutant, in which the Gal4 binding site was inactivated to prevent DNA cleavage by the Gal4BD-Spo11 fusion protein. New DSBs readily appear nearby the *UAS*. Mutant 2 corresponds to, for instance, a *set1*Δ mutant, in which the disappearance of "canonical" hotspots often results in the activation of latent hotspots.

Downloaded from http://cshperspectives.cshlp.org/ on October 26, 2021 - Published by Cold Spring Harbor Laboratory Press

Cold Spring Harbor Perspectives in Biology
www.cshperspectives.org

source of human diseases (Campbell and Eichler 2013). Also, cells with broken chromosomes may progress further into the meiotic prophase, and, unless they are arrested by the DNA damage checkpoint before MI, the chromosomes would segregate abnormally, yielding unbalanced genomes.

Another parameter that can result in DSB modulation is the unresolved question of the synchrony/asynchrony of DSB formation in individual meioses. Time-course experiments in synchronously sporulating yeast cell populations and mutant cells that accumulate unrepaired DSBs (*rad50S* or *sae2*Δ) show that a majority of DSBs (~150 in *S. cerevisiae*) tend to occur within a 30- to 60-minute window, but the level of synchrony in these experiments remains insufficient to conclude whether the DSBs take place simultaneously or progressively in individual cells. This issue concomitantly raises the question of how the recruitment and biochemical cleavage activity of Spo11 are functionally and temporally regulated. Although Spo11's covalent attachment to the cleaved strand is a suicide mechanism preventing the turnover of the protein, a regulatory pathway must also exist to locally and/or globally terminate the window for DSB formation. How this happens is not understood. The capacity of certain pachytene arrested mutants (e.g., *ndt80*Δ) to continue DSB formation (Xu and Kleckner 1995), and also the late persistence of Spo11 and other DSB proteins along the chromosomes after DSB formation (Arora et al. 2004; Cole et al. 2010; Gray et al. 2013), are puzzling observations that call to further integrate the connection of DSB formation and repair with changes in chromosome structures and movements and explore the underlying signaling pathway(s) in connection to the meiotic cell-cycle progression.

Another important DSB control can be called "DSB compensation" to denote the homeostatic process that is able to redistribute DSBs along the chromosomes (Fig. 2C). When strong DSB sites become suppressed, locally or genome-wide, as a result of a mutation, a vast number of dormant DSB hotspots get activated to compensate for the loss of DSBs. Consequently, the original DSB distribution can be dramatically remodeled. For instance, when the GAL4BD-Spo11 fusion protein (when the DNA-binding domain of the Gal4 transcription factor is fused with Spo11) (Pecina et al. 2002) targets DSBs near the UAS_{GAL} DNA-binding motives, mutating the UAS_{GAL} site at the *GAL80* locus causes a severe drop of Gal4BD-Spo11-targeted DSB levels (from 10% to 3%) at the *GAL80* promoter, but compensatory "weak" DSBs readily appear elsewhere, within the same intergenic region (Fig. 2C, middle track) (Robine et al. 2007). Also, null mutations of certain histone-modifying enzymes (e.g., *set1*Δ, *sir2*Δ, $Prdm9^{-/-}$) significantly suppress the activity of canonical recombination hotspots, but a number of compensatory DSBs appear elsewhere (Fig. 2C, lower track) (Borde et al. 2009; Brick et al. 2012). For example, in $Prdm9^{-/-}$ knockout mice, lacking the sequence-targeted meiosis-specific histone-K4 methylase PRDM9, DSB sites are redirected to different genomic regions, which are flanked by PRDM9-independent H3K4me3 marks (Brick et al. 2012). In this way, "DSB compensation" can, at least in part, buffer the absence of natural ("frequent") recombination sites to assure the minimal level of DSBs required for homolog pairing and reinforce the "obligatory" COs per chromosome (i.e., $\geq$1 CO/chromosome arm). It is noteworthy that canonical DSBs can be successfully compensated for in the absence of Set1, giving rise to viable spores (Acquaviva et al. 2013). On the contrary, Prdm9 mutants show severe defects in the progress of meiosis (Hayashi et al. 2005; Brick et al. 2012), indicating that these delocalized DSBs are, in some respects, different from naturally occurring DSBs. The differences between the physiological versus pathological outcomes for "relocalized" and "compensatory" DSBs need to be better understood, especially if they used to modify the natural landscape of meiotic recombination.

It is noteworthy that the total number of recombination events (overall recombination rate) and locations of COs (related to hotspot usage) are two distinct features, which are both determined genetically (Kong et al. 2014). For instance, single-nucleotide variants of PRDM9 affect both recombination rate and location.

Downloaded from http://cshperspectives.cshlp.org/ on October 26, 2021 - Published by Cold Spring Harbor Laboratory Press

Cold Spring Harbor Perspectives in Biology
www.cshperspectives.org

Other genetic variants have been identified that alter the genetic map in humans: (1) CTCFL, the testis-specific paralog of CTCF, a Zn-finger protein organizing chromatin loops with cohesins (Sleutels et al. 2012), (2) the meiosis-specific cohesin RAD21L, which plays a role in the formation of meiotic chromosome axis and synapsis, (3) the SUMO ligase RNF212, ubiquitin ligase HEI10 (CCNB1IP1), and MutS homolog MSH4, all related to members of the ZMM class of pro-CO factors identified in *S. cerevisiae*, although with different synapsis defects (Reynolds et al. 2013; Kong et al. 2014; Qiao et al. 2014). Strikingly, significant differences are seen among these genetic variants, as those having a large impact on recombination rates have no effect on hotspot usage. This implies that, in most cases, genetic variants affect the CO or NCO decision rather than directly touching on DSB formation. Accordingly, the yeast Zip3 protein preferentially localizes to DSB hotspots that tend to be repaired as COs (whereas Zip3 is depleted from NCO-biased hotspots) (Serrentino et al. 2013), pointing to the existence of distinct types of DSB sites with regard to CO and NCO formation.

CHROMATIN REMODELING AND HISTONE MODIFICATIONS IMPACT MEIOTIC DSB FORMATION

Overall, heritable polymorphisms can be a major source of quantitative genetic variation shaping the recombination landscape among individuals, but circumstantial and epigenetically inherited elements of the chromatin structure can also significantly contribute to the plasticity of DSB formation. DNA cleavage by Spo11 requires an accessible DNA template; hence, the energy barrier inherently exerted by nucleosomes and other chromatin-packaging proteins must be overcome. Because spontaneous rates of dissociation and sliding of nucleosome core particles vastly exceed biologically relevant timescales, diffusion-driven processes alone cannot account for the extensive reorganization seen at DSB sites in vivo. In mitotically dividing cells, the concerted action of histone-modifying activities and nucleosome remodelers stimulates nucleosome mobilization, end resection, and strand invasion at DSBs to finally restore the original chromatin structure. Although direct evidence is missing, the same processes are expected to hold for meiotic cells.

Histone Modifications

The chromatin-flanking DSB hotspots have a clear spatial organization of histone marks (Fig. 3A) such that H3K9ac falls closest to hotspots, followed by H3K4me3 concentrated at the +1, +2, +3 nucleosomes, whereas H3K4me1/me2, H3K36me3, H3K79me2, and H3R2me/as are mainly present inside or at the 3′ ends of open reading frames (ORFs) (Zhang et al. 2011b). Histone H3K56 acetylation is a transient mark enriched on nucleosomes that show rapid turnover kinetics (Rufiange et al. 2007; Watanabe et al. 2013). These chemical labels can impact meiotic recombination (1) by affecting the structure and mobility of nucleosome core particles, and (2) by interacting with chromatin-signaling proteins that use histone modifications as docking sites. Mishaps in the proper writing, reading, and erasing of these biochemical tags can interfere with recombination, leading to the formation of pathological diseases, such as infertility or carcinogenesis (Schwartzentruber et al. 2012). Here, we mention a few remarkable examples from various model organisms that exemplify the association between histone modifications and meiotic DSBs. (1) In *C. elegans*, acetylation of histone H2A at lysine 5 (H2AK5ac) seems to play a crucial role in meiotic DSB and CO formation (Wagner et al. 2010). Deletion of the *XND-1* gene (X nondisjunction factor 1) disrupts H2AK5 acetylation and induces a significant change in the meiotic DSB and CO landscape, with most COs occurring abnormally within the gene-rich regions of autosomes. (2) In fission yeast, H3K9ac is specifically enriched at recombination hotspots, but the "prototypical" DSB mark H3K4me3 is less relevant (Yamada et al. 2013). Mutating the H3K9 residue mildly reduced levels of Rec12 binding (homologue of Spo11) and DSBs, indicating that H3K9ac may facilitate recombination initiation by stabilizing the contact between

Downloaded from http://cshperspectives.cshlp.org/ on October 26, 2021 - Published by Cold Spring Harbor Laboratory Press

Cold Spring Harbor Perspectives in Biology
www.cshperspectives.org

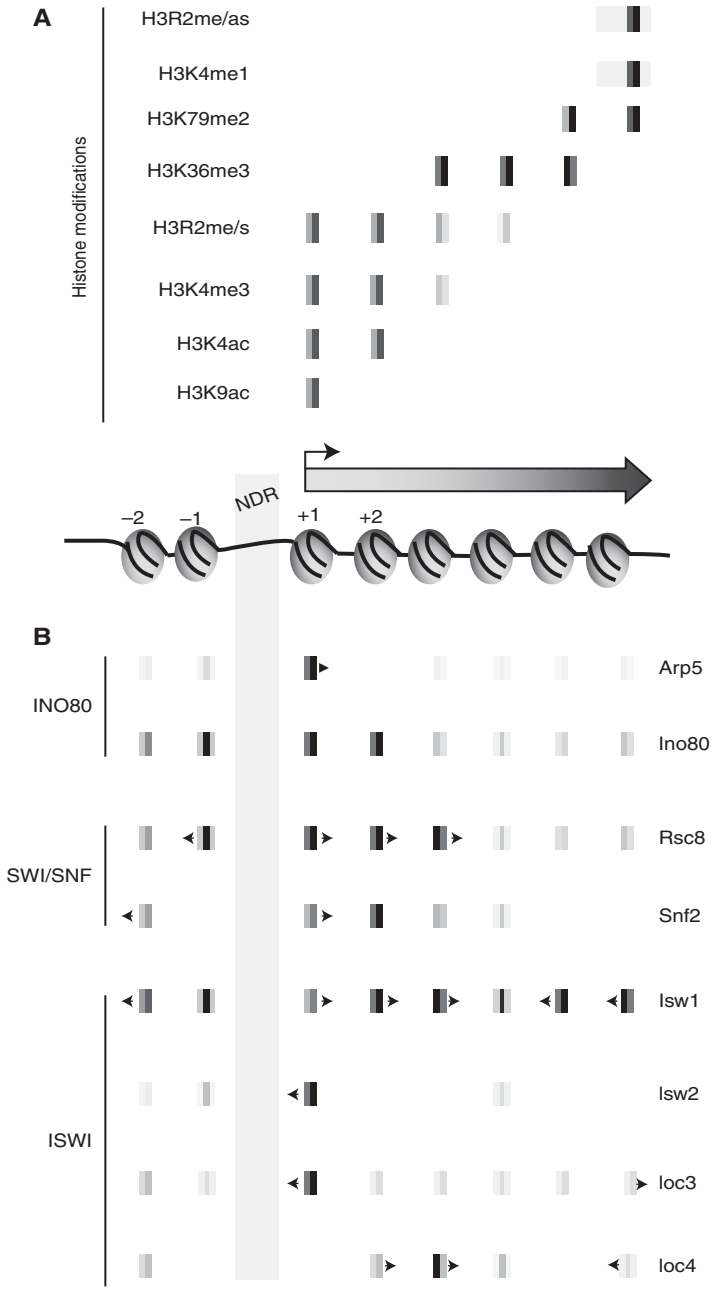


Figure 3. Predominant positions of histone modifications and chromatin remodelers relative to nucleosome depleted regions (NDRs). (*A*) Histone modifications, and (*B*) chromatin remodeling factors flanking NDRs are shown. Arrowheads point to the net directionality of nucleosome movement performed by the corresponding remodeler protein. Ino80 spreads across many positions, but the Arp5 subunit is enriched at the +1 position. Rsc8 and Snf2 are enriched predominantly at the first three genic nucleosomes, but Snf2 is largely depleted at the −1 position and enriched at −2. Isw2 maps mainly to the +1 position and the regulatory subunit Ioc3 (ISW1a) is particularly enriched at the +1 position, whereas Ioc4 (ISW1b) is enriched at positions +2, +3, and +4.

Downloaded from http://cshperspectives.cshlp.org/ on October 26, 2021 - Published by Cold Spring Harbor Laboratory Press

Cold Spring Harbor Perspectives in Biology
www.cshperspectives.org

Rec12 and DSB sites. (3) In budding yeast, deletion of the Sir2 histone deacetylase causes variable, but genome-wide changes in DSB formation, reducing or elevating DSB levels at ~12% of hotspots (Mieczkowski et al. 2007). Sir2 could actively repress DSBs within naturally cold regions (such as the rDNA cluster, telomeres, centromeres) to prevent nonallelic homologous recombination. Interestingly, even in the presence of active Sir2, the edges of the rDNA array remain exceptionally susceptible to meiotic DSBs (Vader et al. 2011). It has turned out that a secondary border-specific system, involving the meiotic ATPase Pch2 and Orc1, operates at heterochromatin–euchromatin junctions to shield the edges of the rDNA array. Pch2 is an evolutionarily conserved AAA-ATPase protein, which appears to influence the initiation of recombination and timely meiotic progression (Borner et al. 2008).

A number of studies indicate that histone H3K4me is a fundamental mark of meiotic DSB formation, which is conserved from yeast to human: (1) deletion of the H3K4 methyltransferase Set1 or point mutation of the modifiable histone H3 lysine 4 residue severely reduce meiotic DSB levels at canonical hotspots in *S. cerevisiae* (Acquaviva et al. 2013; Sommermeyer et al. 2013); (2) absence of Set1 in the distantly related fission yeast partially reduces DSB formation at various loci, in which Set1 and H3K9 acetylation redundantly regulated meiotic DSB formation (Yamada et al. 2013); (3) deletion of *RAD6*, as well as the substitution of the ubiquitylation site on histone H2B, both of which affect H3K4 methylation, decrease DSB frequencies at various hotspots (Yamashita et al. 2004); (4) H3K4me3 levels were significantly enriched at several yeast (Robine et al. 2007) and mouse (Smagulova et al. 2011) DSB hotspots, as well as at *Arabidopsis thaliana* CO hotspots (Choi et al. 2013); and (5) from mice to human, hotspot activity is largely dependent on Prdm9, the sequence-targeted meiosis-specific H3K4 methyltransferase (Grey et al. 2011).

Our latest work (Acquaviva et al. 2013) has revealed a causative and unexpected link between the presence of histone H3 lysine 4 methylation and DSB formation. A genetically engineered Set1 histone methyltransferase (Gal4BD-Set1) targeted to recombination cold regions readily induces DSB formation at these sites (Fig. 4A). The DSB-inducing effect of Gal4BD-Set1 depends on the presence of the modifiable lysine 4 residue (because DSBs at these sites were abolished in the H3K4R mutant), revealing a cause–effect relationship between the presence of histone H3K4me and meiotic DSB formation. Unexpectedly, although Gal4BD-Set1-induced DSBs were strongly dependent on Spp1, the PHD-finger subunit of COMPASS (Fig. 4A), the reverse was not true: tethering of Spp1 to recombination cold spots (Gal4BD-Spp1) strongly induced DSB formation, but these breaks were largely independent of Set1 and histone H3K4me (Fig. 4B). Moreover, the Gal4BD-Spp1-induced DSBs were maintained in both *set1*Δ and H3K4R mutants, indicating that Spp1 on its own is able to initiate meiotic recombination when recruited to the chromosomes. The DSB-promoting effect of Spp1 is mediated via its PHD finger domain, binding to H3K4 trimethylated nucleosomes, and by its physical interaction with the "core" DSB protein Mer2 (Acquaviva et al. 2013; Sommermeyer et al. 2013). Collectively, these findings suggest that Spp1 makes a contact bridge between DSB hotspots and the chromosomal axis via contacts with Mer2 and histone H3K4me3, and contrast the situation in fission yeast, which uses a meiosis-specific bridge protein (liaisonin) to mediate axis–DSB hotspot interaction (Miyoshi et al. 2012).

It should also be noted that a subclass of budding yeast DSB cold spots, localized proximal to chromosomal axes, are associated with lower histone H3K4me3. In budding yeast, binding of Rec8, the kleisin subunit of meiotic cohesin, along chromosome axes plays a critical role in determining the canonical distribution of meiotic DSB sites (Kugou et al. 2009). Regions spanning $\pm$0.8 kbp around axial Rec8 binding sites show lower Spo11-oligo frequency (Ito et al. 2014). Moreover, Spo11 fused with the Gal4 DNA-binding domain (Gal4BD-Spo11) cannot form meiotic DSBs efficiently when targeted to sites adjacent to Rec8 binding sites. In addition, H3K4me3 levels are remarkably lower in Rec8

Downloaded from http://cshperspectives.cshlp.org/ on October 26, 2021 - Published by Cold Spring Harbor Laboratory Press

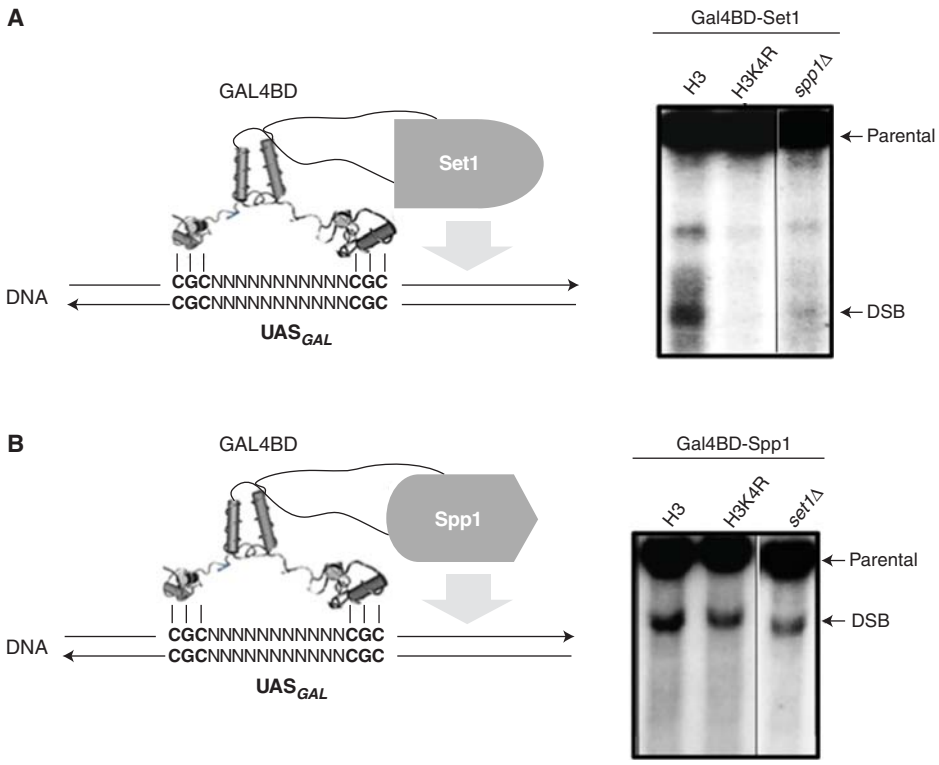


Figure 4. Modulating the double-strand break (DSB) pattern by targeting COMPASS proteins to ectopic chromosomal regions. (*A*) Gal4BD-Set1, and (*B*) Gal4BD-Spp1 fusion proteins induce meiotic DSB formation within recombinationally cold regions. Gal4BD-Set1-targeted DSBs depend on the presence of the modifiable histone H3 lysine 4 residue (H3K4R mutant), as well as the presence of Spp1 (*spp1*Δ mutant), whereas Gal4BD-Spp1-targeted DSBs are independent of the H3 lysine 4 residue and Set1 (*set1*Δ mutant). UAS, upstream activating sequence.

Cold Spring Harbor Perspectives in Biology
www.cshperspectives.org

binding sites. It is, thus, suggested that reduced histone H3K4me3 down-regulates Spo11 activity on sequences proximal to the axes. One possible mechanism is that the absence of H3K4me3 around the axis hampers formation of tethered axis-loop complexes resulting in local inhibition of DSBs.

Collectively, these recent studies strongly implicate COMPASS, H3K4me3, and Mer2 in the determination of DSB sites. However, there are several key unknowns to understand the relationship among meiotic DSB sites, histone modifications, and higher-order genome architecture. For example, (1) existence of axis-loop contacts and their role in DSB formation have not been proven at the molecular levels (physical evidence is missing); (2) it is not known which chromatin proteins and histone modifications, in particular, COMPASS, Mer2, and H3K4me3, make or participate in mediating these contacts (causality is missing); and (3) mechanisms for DSB site selection in the absence of H3K4 methylation and repression of Spo11 activity in Rec8 binding sites with lower H3K4me3 remain unknown. Even if not absolutely required, the broadly localized H3K4me3 modification has the virtue of permitting the initiation of recombination at numerous places of the genome, a molecular strategy that provides flexibility and ensures a large diversity of recombinant haplotypes to be transmitted by the gametes.

CHROMATIN REMODELERS AND HISTONE MODIFICATIONS

Chromatin remodeling factors (Table 1) can significantly accelerate the dynamics of nucleo-

Downloaded from http://cshperspectives.cshlp.org/ on October 26, 2021 - Published by Cold Spring Harbor Laboratory Press

Cold Spring Harbor Perspectives in Biology
www.cshperspectives.org

somes to allow for more rapid and localized access of Spo11 to meiotic DSB sites. For instance, the fission yeast SWI/SNF-type ATP-dependent remodeler SNF22 and CHD-1-type ATP-dependent remodeler Hrp3 activate the *M26* recombination hotspot before DSB formation (Yamada et al. 2004). During DSB repair in mitotically cycling cells, chromatin remodelers exhaustively participate in disrupting and mobilizing nucleosomes (Mellor and Morillon 2004; Seeber et al. 2013). (1) Physical tethering of INO80 (LexA-Ino80) to lacO-tagged DSBs enhances the mobility of the breaks, causing increased gene conversion rates at ectopic donor sites (Neumann et al. 2012); also, INO80 is recruited to DSBs by carboxy-terminally phosphorylated γH2AX, in which it evicts/remodels the H2A.Z variant nucleosomes (Htz1 in yeast) to facilitate end resection. (2) Fun30 mediates the process of end resection after DSB induction (Chen et al. 2012). (3) Rad54 mediates strand-exchange reactions during Holliday junction formation and resolution (Nimonkar et al. 2012). Whether the same remodeling factors function in meiotic recombination remains unclear.

A number observation points toward a functional link between chromatin remodelers and histone modifications (however, these relationships can be rather complex as the latter affects the sites of chromatin remodeling, and vice versa). (1) On meiotic DSB formation, there is the rapid appearance of open chromatin—showing increased DNaseI and MNase sensitivity at hotspots—promptly followed by γH2AX phosphorylation and histone H4 acetylation spreading over hundreds of kbps away from breaks (Ohta et al. 1994; Downs et al. 2004). (2) Histone H3K56 acetylation (H3K56ac) and SWR1 are mechanistically coupled as acetylated H3K56 modulates the specificity of SWR1 to remove the histone variant H2A.Z from gene regulatory regions (Watanabe et al. 2013). It is also known that *A. thaliana* meiotic DSB sites significantly overlap with H2A.Z nucleosomes and H3K4me3 (similarly to *S cerevisiae*) and the SWRI-deposited H2A.Z nucleosomes promote meiotic DSB formation and repair (Choi et al. 2013). (3) Members of the SWI/SNF family have carboxy-terminal bromodomains, which interact with acetylated histones, to preferentially target the acetylated nucleosomes for eviction (Yodh 2013). (4) The CHD remodelers have two tandem chromodomains, recognizing methylated histone H3 tails.

Bai et al. 2011 has screened ~6000 *S. cerevisiae* nucleosome-depleted regions (NDRs) for consensus binding sites of known nucleosome-depleting factors (NDFs), including remodelers (e.g., Rsc3) and transcription factors (e.g., Abf1, Rap1, Reb1). A significant fraction of NDRs (30%) contained at least one NDF binding site. Similar observations were published in *S. pombe* (De et al. 2012), which together suggest that different NDRs depend on the modular binding of NDFs (remodelers, transcription factors, and histone modifiers). As not all NDRs constitute a functional recombination hotspot (Pan et al. 2011), it is easy to envisage that only a combinatorial association of these factors makes an NDR competent for recruiting the Spo11 machinery to initiate homologous recombination.

Table 1. Chromatin remodeling factors

	SWI/SNF	ISWI	CHD	INO80/SWRI
Saccharomyces cerevisiae	SWI/SNF RSC	ISWI1a ISWI1b ISW2	CHD1	INO80 SWRI
Homo sapiens	SWI/SNF	ACF/WCFR CHRAC RSF WICH	NuRD	INO80 SRCAP

Yeast and human chromatin remodeling complexes classified by their ATPase subunits. CHD, chromo-helicase/ATPase DNA binding.

Downloaded from http://cshperspectives.cshlp.org/ on October 26, 2021 - Published by Cold Spring Harbor Laboratory Press

Cold Spring Harbor Perspectives in Biology
www.cshperspectives.org

In a systematic analysis, Yen et al. (2013) has revealed that remodelers bind to DNA in a nucleosome-position and orientation-specific manner. Nucleosomes are mobilized with a predefined net directionality relative to NDRs (toward or away from them), such that (1) remodelers bind predominantly to the +1 nucleosome, or (2) they are located at multiple positions along the ORF (Fig. 3B). For example, Ino80 spreads across many positions, but the Arp5 subunit of the same INO80 complex is enriched at the +1 position, moving this nucleosome 5′ to 3′. In line with the above, the RSC (remodels the structure of chromatin) remodeling complex plays a highly specific role in the precise positioning of nucleosomes over yeast promoter, a function that cannot be replaced by other closely related remodeling enzymes (Wippo et al. 2011). All of the above collectively suggest that ±1 nucleosome-flanking NDRs are differentially processed by remodelers such that the initial positioning of this particular nucleosome may automatically cause the positioning of adjacent nucleosomes. Therefore, NDRs, in which most meiotic DSBs fall in *S. cerevisiae* and *Arabidopsis*, might be the active organizing centers (not just simple bystanders) of DNA cleavage in meiosis.

CONCLUDING REMARKS

Numerous studies have suggested that post-translational histone modifications do not represent a "code" (Sims and Reinberg 2008; Lee et al. 2010), but that the majority of these chemical tags mobilize or immobilize, rather than mark, the nucleosomes. Because of the stringent spatial order of histone modifications that flank meiotic DSB sites, many aspects of meiotic DSB control, including hotspot intensity, clustering/interference, and compensation, might operate at the level of the well-positioned ±1 nucleosomes bordering DSB hotspots. Indeed, binding of Prdm9 to DSB sites actively reorganizes the flanking ±1 nucleosomes, creating a symmetrical pattern of NDRs centered over the 13-mer consensus binding motif of the PRDM9 zinc-finger array (Baker et al. 2014). In this way, H3K4 methylated nucleosomes establish a permissible chromatin structure for meiotic DSB formation, depending on the activity of PRDM9.

In conclusion, the ±1 nucleosome-flanking meiotic DSB hotspots seem to be critical for meiotic recombination initiation and repair. We propose that a combinatorial association of histone modifications and nucleosome remodelers positively (or negatively) regulates the turnover/mobility of the critical ±1 nucleosome. This, in turn, automatically positions the adjacent nucleosomes to establish a permissive (or restrictive) chromatin context for recombination. The resultant effect of these activities is expected to modulate the efficiency of Spo11-mediated DNA cleavage at meiotic recombination hotspots.

ACKNOWLEDGMENTS

Work in the A. Nicolas laboratory was supported by a grant from Meiogenix SA. L.S. was supported by the European Union (FP7/MCA-CIG), CRP-ICGEB (Trieste, Italy), and the Hungarian Scientific Research Fund (OTKA-PD), cofinanced by the European Social Fund in the framework of TÁMOP-4.2.4.A/2-11/1-2012-0001 National Excellence Program. K.O. was supported by a Grant-in-Aid for Scientific Research on Innovative Areas and Platform for Dynamic Approaches to Living System from The Ministry of Education, Culture, Sports, Science and Technology (MEXT) and the Japan Society for the Promotion of Science (JSPS).

REFERENCES

*Reference is also in this collection.

Acquaviva L, Székvölgyi L, Dichtl B, Dichtl BS, de La Roche Saint André C, Nicolas A, Geli V. 2013. The COMPASS subunit Spp1 links histone methylation to initiation of meiotic recombination. *Science* **339:** 215–218.

Arora C, Kee K, Maleki S, Keeney S. 2004. Antiviral protein Ski8 is a direct partner of Spo11 in meiotic DNA break formation, independent of its cytoplasmic role in RNA metabolism. *Mol Cell* **13:** 549–559.

Auton A, Fledel-Alon A, Pfeifer S, Venn O, Segurel L, Street T, Leffler EM, Bowden R, Aneas I, Broxholme J, et al. 2012. A fine-scale chimpanzee genetic map from population sequencing. *Science* **336:** 193–198.

Bai L, Ondracka A, Cross FR. 2011. Multiple sequence-specific factors generate the nucleosome-depleted region on CLN2 promoter. *Mol Cell* **42:** 465–476.

Downloaded from http://cshperspectives.cshlp.org/ on October 26, 2021 - Published by Cold Spring Harbor Laboratory Press

Baker CL, Walker M, Kajita S, Petkov PM, Paigen K. 2014. PRDM9 binding organizes hotspot nucleosomes and limits Holliday junction migration. *Genome Res* **24:** 724–732.

Baudat F, Nicolas A. 1997. Clustering of meiotic double-strand breaks on yeast chromosome III. *Proc Natl Acad Sci* **94:** 5213–5218.

Bergerat A, de Massy B, Gadelle D, Varoutas PC, Nicolas A, Forterre P. 1997. An atypical topoisomerase II from Archaea with implications for meiotic recombination. *Nature* **386:** 414–417.

Bishop DK. 1994. RecA homologs Dmc1 and Rad51 interact to form multiple nuclear complexes prior to meiotic chromosome synapsis. *Cell* **79:** 1081–1092.

Blitzblau HG, Hochwagen A. 2013. ATR/Mec1 prevents lethal meiotic recombination initiation on partially replicated chromosomes in budding yeast. *eLife* **2:** e00844.

Borde V, Lin W, Novikov E, Petrini JH, Lichten M, Nicolas A. 2004. Association of Mre11p with double-strand break sites during yeast meiosis. *Mol Cell* **13:** 389–401.

Borde V, Robine N, Lin W, Bonfils S, Geli V, Nicolas A. 2009. Histone H3 lysine 4 trimethylation marks meiotic recombination initiation sites. *EMBO J* **28:** 99–111.

Borner GV, Barot A, Kleckner N. 2008. Yeast Pch2 promotes domainal axis organization, timely recombination progression, and arrest of defective recombinosomes during meiosis. *Proc Natl Acad Sci* **105:** 3327–3332.

Bowring FJ, Yeadon PJ, Stainer RG, Catcheside DE. 2006. Chromosome pairing and meiotic recombination in *Neurospora crassa* spo11 mutants. *Curr Genet* **50:** 115–123.

Brick K, Smagulova F, Khil P, Camerini-Otero RD, Petukhova GV. 2012. Genetic recombination is directed away from functional genomic elements in mice. *Nature* **485:** 642–645.

Buhler C, Borde V, Lichten M. 2007. Mapping meiotic single-strand DNA reveals a new landscape of DNA double-strand breaks in *Saccharomyces cerevisiae*. *PLoS Biol* **5:** e324.

Campbell CD, Eichler EE. 2013. Properties and rates of germline mutations in humans. *Trends Genet* **29:** 575–584.

Chen X, Cui D, Papusha A, Zhang X, Chu CD, Tang J, Chen K, Pan X, Ira G. 2012. The Fun30 nucleosome remodeller promotes resection of DNA double-strand break ends. *Nature* **489:** 576–580.

Choi K, Zhao X, Kelly KA, Venn O, Higgins JD, Yelina NE, Hardcastle TJ, Ziolkowski PA, Copenhaver GP, Franklin FC, et al. 2013. Arabidopsis meiotic crossover hot spots overlap with H2A.Z nucleosomes at gene promoters. *Nat Genet* **45:** 1327–1336.

Cole F, Keeney S, Jasin M. 2010. Comprehensive, fine-scale dissection of homologous recombination outcomes at a hot spot in mouse meiosis. *Mol Cell* **39:** 700–710.

Cromie GA, Hyppa RW, Cam HP, Farah JA, Grewal SI, Smith GR. 2007. A discrete class of intergenic DNA dictates meiotic DNA break hotspots in fission yeast. *PLoS Genet* **3:** e141.

de Castro E, Soriano I, Marin L, Serrano R, Quintales L, Antequera F. 2012. Nucleosomal organization of replication origins and meiotic recombination hotspots in fission yeast. *EMBO J* **31:** 124–137.

Downs JA, Allard S, Jobin-Robitaille O, Javaheri A, Auger A, Bouchard N, Kron SJ, Jackson SP, Cote J. 2004. Binding of chromatin-modifying activities to phosphorylated histone H2A at DNA damage sites. *Mol Cell* **16:** 979–990.

Drouaud J, Khademian H, Giraut L, Zanni V, Bellalou S, Henderson IR, Falque M, Mezard C. 2013. Contrasted patterns of crossover and non-crossover at *Arabidopsis thaliana* meiotic recombination hotspots. *PLoS Genet* **9:** e1003922.

Fan QQ, Xu F, White MA, Petes TD. 1997. Competition between adjacent meiotic recombination hotspots in the yeast *Saccharomyces cerevisiae*. *Genetics* **145:** 661–670.

Farah JA, Cromie GA, Smith GR. 2009. Ctp1 and Exonuclease 1, alternative nucleases regulated by the MRN complex, are required for efficient meiotic recombination. *Proc Natl Acad Sci* **106:** 9356–9361.

Fowler KR, Sasaki M, Milman N, Keeney S, Smith GR. 2014. Evolutionarily diverse determinants of meiotic DNA break and recombination landscapes across the genome. *Genome Res* **24:** 1650–1664.

Gerton JL, DeRisi J, Shroff R, Lichten M, Brown PO, Petes TD. 2000. Global mapping of meiotic recombination hotspots and coldspots in the yeast *Saccharomyces cerevisiae*. *Proc Natl Acad Sci* **97:** 11383–11390.

Gray S, Allison RM, Garcia V, Goldman AS, Neale MJ. 2013. Positive regulation of meiotic DNA double-strand break formation by activation of the DNA damage checkpoint kinase Mec1(ATR). *Open Biol* **3:** 130019.

Grey C, Barthes P, Chauveau-Le FG, Langa F, Baudat F, de Massy B. 2011. Mouse PRDM9 DNA-binding specificity determines sites of histone H3 lysine 4 trimethylation for initiation of meiotic recombination. *PLoS Biol* **9:** e1001176.

Hayashi K, Yoshida K, Matsui Y. 2005. A histone H3 methyltransferase controls epigenetic events required for meiotic prophase. *Nature* **438:** 374–378.

Hodgson A, Terentyev Y, Johnson RA, Bishop-Bailey A, Angevin T, Croucher A, Goldman AS. 2011. Mre11 and Exo1 contribute to the initiation and processivity of resection at meiotic double-strand breaks made independently of Spo11. *DNA Repair (Amst)* **10:** 138–148.

Ito M, Kugou K, Fawcett JA, Mura S, Ikeda S, Innan H, Ohta K. 2014. Meiotic recombination cold spots in chromosomal cohesion sites. *Genes Cells* **19:** 359–373.

Kolodkin AL, Klar AJ, Stahl FW. 1986. Double-strand breaks can initiate meiotic recombination in *S. cerevisiae*. *Cell* **46:** 733–740.

Kong A, Thorleifsson G, Frigge ML, Masson G, Gudbjartsson DF, Villemoes R, Magnusdottir E, Olafsdottir SB, Thorsteinsdottir U, Stefansson K. 2014. Common and low-frequency variants associated with genome-wide recombination rate. *Nat Genet* **46:** 11–16.

Kugou K, Fukuda T, Yamada S, Ito M, Sasanuma H, Mori S, Katou Y, Itoh T, Matsumoto K, Shibata T, et al. 2009. Rec8 guides canonical Spo11 distribution along yeast meiotic chromosomes. *Mol Biol Cell* **20:** 3064–3076.

Cold Spring Harbor Perspectives in Biology
www.cshperspectives.org

Downloaded from http://cshperspectives.cshlp.org/ on October 26, 2021 - Published by Cold Spring Harbor Laboratory Press

Cold Spring Harbor Perspectives in Biology
www.cshperspectives.org

* Lam I, Keeney S. 2015. Mechanism and regulation of meiotic recombination initiation. *Cold Spring Harb Perspect Biol* **7:** a016634.

Lao JP, Cloud V, Huang CC, Grubb J, Thacker D, Lee CY, Dresser ME, Hunter N, Bishop DK. 2013. Meiotic crossover control by concerted action of Rad51-Dmc1 in homolog template bias and robust homeostatic regulation. *PLoS Genet* **9:** e1003978.

Lee JS, Smith E, Shilatifard A. 2010. The language of histone crosstalk. *Cell* **142:** 682–685.

Malkova A, Klein F, Leung WY, Haber JE. 2000. HO endonuclease-induced recombination in yeast meiosis resembles Spo11-induced events. *Proc Natl Acad Sci* **97:** 14500–14505.

Mancera E, Bourgon R, Brozzi A, Huber W, Steinmetz LM. 2008. High-resolution mapping of meiotic crossovers and non-crossovers in yeast. *Nature* **454:** 479–485.

Mellor J, Morillon A. 2004. ISWI complexes in *Saccharomyces cerevisiae*. *Biochim Biophys Acta* **1677:** 100–112.

Mieczkowski PA, Dominska M, Buck MJ, Lieb JD, Petes TD. 2007. Loss of a histone deacetylase dramatically alters the genomic distribution of Spo11p-catalyzed DNA breaks in *Saccharomyces cerevisiae*. *Proc Natl Acad Sci* **104:** 3955–3960.

Miyoshi T, Ito M, Kugou K, Yamada S, Furuichi M, Oda A, Yamada T, Hirota K, Masai H, Ohta K. 2012. A central coupler for recombination initiation linking chromosome architecture to S phase checkpoint. *Mol Cell* **47:** 722–733.

Myers S, Bowden R, Tumian A, Bontrop RE, Freeman C, MacFie TS, McVean G, Donnelly P. 2010. Drive against hotspot motifs in primates implicates the PRDM9 gene in meiotic recombination. *Science* **327:** 876–879.

Neale MJ, Pan J, Keeney S. 2005. Endonucleolytic processing of covalent protein-linked DNA double-strand breaks. *Nature* **436:** 1053–1057.

Neumann FR, Dion V, Gehlen LR, Tsai-Pflugfelder M, Schmid R, Taddei A, Gasser SM. 2012. Targeted INO80 enhances subnuclear chromatin movement and ectopic homologous recombination. *Genes Dev* **26:** 369–383.

Nimonkar AV, Dombrowski CC, Siino JS, Stasiak AZ, Stasiak A, Kowalczykowski SC. 2012. *Saccharomyces cerevisiae* Dmc1 and Rad51 proteins preferentially function with Tid1 and Rad54 proteins, respectively, to promote DNA strand invasion during genetic recombination. *J Biol Chem* **287:** 28727–28737.

Ohta K, Shibata T, Nicolas A. 1994. Changes in chromatin structure at recombination initiation sites during yeast meiosis. *EMBO J* **13:** 5754–5763.

Pan J, Sasaki M, Kniewel R, Murakami H, Blitzblau HG, Tischfield SE, Zhu X, Neale MJ, Jasin M, Socci ND, et al. 2011. A hierarchical combination of factors shapes the genome-wide topography of yeast meiotic recombination initiation. *Cell* **144:** 719–731.

Panizza S, Mendoza MA, Berlinger M, Huang L, Nicolas A, Shirahige K, Klein F. 2011. Spo11-accessory proteins link double-strand break sites to the chromosome axis in early meiotic recombination. *Cell* **146:** 372–383.

Pecina A, Smith KN, Mezard C, Murakami H, Ohta K, Nicolas A. 2002. Targeted stimulation of meiotic recombination. *Cell* **111:** 173–184.

Qiao H, Prasada Rao HB, Yang Y, Fong JH, Cloutier JM, Deacon DC, Nagel KE, Swartz RK, Strong E, et al. 2014. Antagonistic roles of ubiquitin ligase HEI10 and SUMO ligase RNF212 regulate meiotic recombination. *Nat Genet* **46:** 194–199.

Reynolds A, Qiao H, Yang Y, Chen JK, Jackson N, Biswas K, Holloway JK, Baudat F, de Massy B, Wang J, et al. 2013. RNF212 is a dosage-sensitive regulator of crossing-over during mammalian meiosis. *Nat Genet* **45:** 269–278.

Robert V, Bessereau JL. 2007. Targeted engineering of the *Caenorhabditis elegans* genome following Mos1-triggered chromosomal breaks. *EMBO J* **26:** 170–183.

Robine N, Uematsu N, Amiot F, Gidrol X, Barillot E, Nicolas A, Borde V. 2007. Genome-wide redistribution of meiotic double-strand breaks in *Saccharomyces cerevisiae*. *Mol Cell Biol* **27:** 1868–1880.

Rocco V, de Massy B, Nicolas A. 1992. The *Saccharomyces cerevisiae* ARG4 initiator of meiotic gene conversion and its associated double-strand DNA breaks can be inhibited by transcriptional interference. *Proc Natl Acad Sci* **89:** 12068–12072.

Rufiange A, Jacques P.E., Bhat W, Robert F, Nourani A. 2007. Genome-wide replication-independent histone H3 exchange occurs predominantly at promoters and implicates H3 K56 acetylation and Asf1. *Mol Cell* **27:** 393–405.

Schwartzentruber J, Korshunov A, Liu XY, Jones DT, Pfaff E, Jacob K, Sturm D, Fontebasso AM, Quang DA, Tonjes M, et al. 2012. Driver mutations in histone H3.3 and chromatin remodelling genes in paediatric glioblastoma. *Nature* **482:** 226–231.

Seeber A, Hauer M, Gasser SM. 2013. Nucleosome remodelers in double-strand break repair. *Curr Opin Genet Dev* **23:** 174–184.

Serrentino ME, Chaplais E, Sommermeyer V, Borde V. 2013. Differential association of the conserved SUMO ligase Zip3 with meiotic double-strand break sites reveals regional variations in the outcome of meiotic recombination. *PLoS Genet* **9:** e1003416.

Sims RJ III, Reinberg D. 2008. Is there a code embedded in proteins that is based on post-translational modifications? *Nat Rev Mol Cell Biol* **9:** 815–820.

Sleutels F, Soochit W, Bartkuhn M, Heath H, Dienstbach S, Bergmaier P, Franke V, Rosa-Garrido M, van de Nobelen S, Caesar L, et al. 2012. The male germ cell gene regulator CTCFL is functionally different from CTCF and binds CTCF-like consensus sites in a nucleosome composition-dependent manner. *Epigenetics Chromatin* **5:** 8.

Smagulova F, Gregoretti IV, Brick K, Khil P, Camerini-Otero RD, Petukhova GV. 2011. Genome-wide analysis reveals novel molecular features of mouse recombination hotspots. *Nature* **472:** 375–378.

Sommermeyer V, Beneut C, Chaplais E, Serrentino ME, Borde V. 2013. Spp1, a member of the Set1 Complex, promotes meiotic DSB formation in promoters by tethering histone H3K4 methylation sites to chromosome axes. *Mol Cell* **49:** 43–54.

Székvölgyi L, Nicolas A. 2010. From meiosis to postmeiotic events: Homologous recombination is obligatory but flexible. *FEBS J* **277:** 571–589.

Vader G, Blitzblau HG, Tame MA, Falk JE, Curtin L, Hochwagen A. 2011. Protection of repetitive DNA borders

Downloaded from http://cshperspectives.cshlp.org/ on October 26, 2021 - Published by Cold Spring Harbor Laboratory Press

from self-induced meiotic instability. *Nature* **477:** 115–119.

Wagner CR, Kuervers L, Baillie DL, Yanowitz JL. 2010. xnd-1 regulates the global recombination landscape in *Caenorhabditis elegans*. *Nature* **467:** 839–843.

Watanabe Y. 2012. Geometry and force behind kinetochore orientation: Lessons from meiosis. *Nat Rev Mol Cell Biol* **13:** 370–382.

Watanabe S, Radman-Livaja M, Rando OJ, Peterson CL. 2013. A histone acetylation switch regulates H2A.Z deposition by the SWR-C remodeling enzyme. *Science* **340:** 195–199.

Wippo CJ, Israel L, Watanabe S, Hochheimer A, Peterson CL, Korber P. 2011. The RSC chromatin remodelling enzyme has a unique role in directing the accurate positioning of nucleosomes. *EMBO J* **30:** 1277–1288.

Xu L, Kleckner N. 1995. Sequence non-specific double-strand breaks and interhomolog interactions prior to double-strand break formation at a meiotic recombination hot spot in yeast. *EMBO J* **14:** 5115–5128.

Yamada T, Mizuno K, Hirota K, Kon N, Wahls WP, Hartsuiker E, Murofushi H, Shibata T, Ohta K. 2004. Roles of histone acetylation and chromatin remodeling factor in a meiotic recombination hotspot. *EMBO J* **23:** 1792–1803.

Yamada S, Ohta K, Yamada T. 2013. Acetylated histone H3K9 is associated with meiotic recombination hotspots, and plays a role in recombination redundantly with other factors including the H3K4 methylase Set1 in fission yeast. *Nucleic Acids Res* **41:** 3504–3517.

Yamashita K, Shinohara M, Shinohara A. 2004. Rad6-Bre1-mediated histone H2B ubiquitylation modulates the formation of double-strand breaks during meiosis. *Proc Natl Acad Sci* **101:** 11380–11385.

Yen K, Vinayachandran V, Pugh BF. 2013. SWR-C and INO80 chromatin remodelers recognize nucleosome-free regions near +1 nucleosomes. *Cell* **154:** 1246–1256.

Yodh J. 2013. ATP-dependent chromatin remodeling. *Adv Exp Med Biol* **767:** 263–295.

Zhang L, Kim KP, Kleckner NE, Storlazzi A. 2011a. Meiotic double-strand breaks occur once per pair of (sister) chromatids and, via Mec1/ATR and Tel1/ATM, once per quartet of chromatids. *Proc Natl Acad Sci* **108:** 20036–20041.

Zhang L, Ma H, Pugh BF. 2011b. Stable and dynamic nucleosome states during a meiotic developmental process. *Genome Res* **21:** 875–884.

* Zickler D, Kleckner N. 2015. Recombinatin, pairing, and synapsis of homologs during meiosis. *Cold Spring Harb Perspect Biol* doi: 10.1101/cshperspect.a016626.

Cold Spring Harbor Perspectives in Biology
www.cshperspectives.org

Downloaded from http://cshperspectives.cshlp.org/ on October 26, 2021 - Published by Cold Spring Harbor Laboratory Press

dc_1977_21

Cold Spring Harbor Perspectives in Biology

Initiation of Meiotic Homologous Recombination: Flexibility, Impact of Histone Modifications, and Chromatin Remodeling

Lóránt Székvölgyi, Kunihiro Ohta and Alain Nicolas

Cold Spring Harb Perspect Biol 2015; doi: 10.1101/cshperspect.a016527

Copyright © 2015 Cold Spring Harbor Laboratory Press; all rights reserved

Downloaded from http://cshperspectives.cshlp.org/ on October 26, 2021 - Published by Cold Spring Harbor Laboratory Press

dc_1977_21

Mediators of Homologous DNA Pairing
Alex Zelensky, Roland Kanaar and Claire Wyman

Regulation of DNA Pairing in Homologous Recombination
James M. Daley, William A. Gaines, YoungHo Kwon, et al.

For additional articles in this collection, see http://cshperspectives.cshlp.org/cgi/collection/

Copyright © 2015 Cold Spring Harbor Laboratory Press; all rights reserved

dc_1977_21

the FEBS Journal

MINIREVIEW

From meiosis to postmeiotic events: Homologous recombination is obligatory but flexible

Lóránt Székvölgyi and Alain Nicolas

Recombination and Genome Instability Unit, Institut Curie, Centre de Recherche, UMR 3244 CNRS, Universite Pierre et Marie Curie, Paris, France

Keywords
double-strand break; histone modification; recombination; sister chromatid cohesion; Spo11

Correspondence
A. Nicolas, 26 rue d'Ulm, 75248 Paris Cedex 05, France
Fax: +33 0 1 56 24 66 44
Tel: +33 0 1 56 24 65 20
E-mail: alain.nicolas@curie.fr

(Received 12 September 2009, revised 9 November 2009, accepted 17 November 2009)

doi:10.1111/j.1742-4658.2009.07502.x

Sexual reproduction depends on the success of faithful chromosome transmission during meiosis to yield viable gametes. Central to meiosis is the process of recombination between paternal and maternal chromosomes, which boosts the genetic diversity of progeny and ensures normal homologous chromosome segregation. Imperfections in meiotic recombination are the source of *de novo* germline mutations, abnormal gametes, and infertility. Thus, not surprisingly, cells have developed a variety of mechanisms and tight controls to ensure sufficient and well-distributed recombination events within their genomes, the details of which remain to be fully elucidated. Local and genome-wide studies of normal and genetically engineered cells have uncovered a remarkable stochasticity in the number and positioning of recombination events per chromosome and per cell, which reveals an impressive level of flexibility. In this minireview, we summarize our contemporary understanding of meiotic recombination and its control mechanisms, and address the seemingly paradoxical and poorly understood diversity of recombination sites. Flexibility in the distribution of meiotic recombination events within genomes may reside in regulation at the chromatin level, with histone modifications playing a recently recognized role.

Introduction

The means by which sexual reproduction emerged some 2 Ga and spread in eukaryotes, conferring a likely evolutionary advantage, is a challenging subject of debate [1]. Central to this phenomenon is meiosis, the unique differentiation process in which the number of chromosomes in diploid germ cells is halved to generate haploid gametes. Then, during fertilization, the fusion of male and female gametes creates a new diploid genome, while the reduction of chromosome number during meiosis keeps the genome size constant over successive generations.

Embedded in the process of meiosis, and essential for its evolutionary role, is the production of genetic diversity in the offspring, upon which selection will act. Meiosis creates new genomic variation in two ways. First, each gamete transmits either chromosome of a given parental pair to offspring, and second, during meiotic prophase, homologous chromosome pairs undergo recombination, which shuffles their polymorphic information. Thus, gametes are genetically diverse. Furthermore, the randomness of fecundation expands diversity in the offspring. Another essential role of meiotic recombination is to ensure proper chromosome segregation into the meiotic products, such as spores in fungi or gametes in other organisms. Halving the chromosome content in the gametes is achieved by

Abbreviations
CO, crossover; dHJ, double Holliday junction; DSB, double-strand break; DSBR, double-strand break repair; HJ, Holliday junction; MI, meiosis I; MII, meiosis II; NCO, noncrossover; POF, premature ovarian failure; SDSA, synthesis-dependent strand-annealing; SEI, single-end invasion; SNP, single-nucleotide polymorphism.

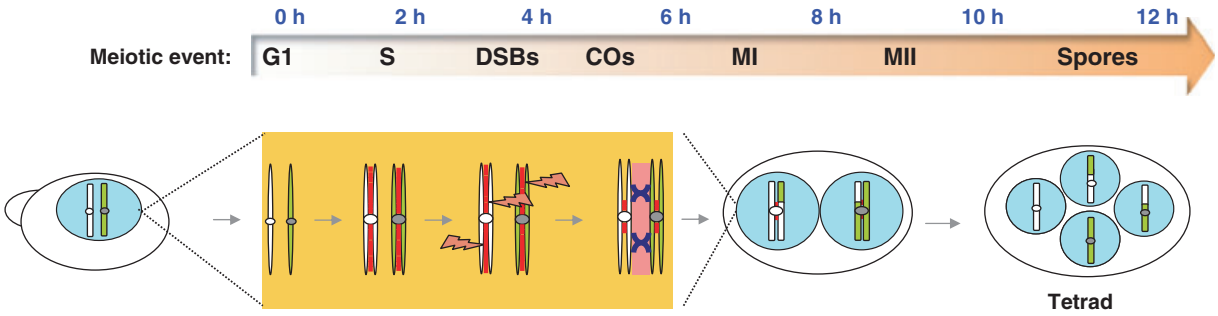


Fig. 1. Meiosis and sporulation in *S. cerevisiae*. Upon nutrient depletion and in the presence of a nonfermentable carbon source, diploid yeast cells initiate meiosis and generate four haploid spores. *S. cerevisiae* strains of the SK1 background are widely used for meiotic studies, because sporulation is rapid (about 12 h) and very efficient (> 90% of cells complete meiosis). Synchronized meiotic samples are easily obtained for time-course physical analyses of premeiotic replication, recombination intermediates, cell division phases (MI and MII), and spore formation. The relevant meiotic events are indicated. Colors: green and white, parental homologous chromosomes; red, sister chromatid cohesion; pink, synaptonemal complex.

a modified version of the mitotic cell cycle (Fig. 1). After one round of DNA replication (also called premeiotic replication) and recombination between homologous chromosomes, during meiosis I (MI, the reductional division), homologous chromosomes segregate from each other, and during meiosis II (MII, the equational division), sister chromatids segregate from each other. These two rounds of chromosomal disjunction yield haploid nuclei that are ultimately packaged into gametes, with or without additional clonal expansions. Central to the process of homologous chromosome segregation is its intimate relationship with recombination, which ensures that chromosomes are held together at metaphase of MI through the formation of at least one crossover (CO) per pair of homologs. To achieve this synaptic relationship – errors in recombination yield a variety of genome abnormalities – and at the same time distribute recombination events along chromosomes, organisms described to various extents (e.g. yeasts, mammals) have developed specific strategies that have begun to be characterized. Also important for meiosis are the dynamics of meiotic chromosome structures and movements that occur during the extended meiotic prophase I, and in particular homolog pairing, which culminates with the formation of the synaptonemal complex, a highly conserved proteinaceous structure that forms between the homologs along their entire lengths. The synaptonemal complex is important for the normal formation of COs [2,3]. In many (but not all) organisms, the homology search that occurs during recombination mediated by DNA–DNA interactions is also intimately associated with the movement of homologous chromosomes to bring them into close juxtaposition. All of these topics have been the subjects of several reviews [4–8]. Methods with which to study meiosis have been recently reviewed in *Methods in Molecular Biology* series *Meiosis* volumes (Springer Protocols, 2009).

Herein, we focus on current knowledge and outstanding questions regarding the mechanisms that control the frequency, location and nature of genomic recombination events. We also consider defects in meiosis that lead to genome alterations, and emphasize recent studies illustrating that, besides its obligate role in proper chromosome segregation during meiosis, homologous recombination is not corseted but is instead flexible. These issues underlie a fascinating cell-to-cell variation in the numbers and positions of recombination events per chromosome and per cell that remains to be mechanistically described. We review the significant progress in unraveling the intimate links between recombination and chromosome segregation, and in uncovering the layers of factors that control local and genome-wide levels of recombination (including histone modifications). Notably, our current knowledge has inspired methods with which to locally and globally modulate the initiation of recombination and thereby to modify the chromosomal distribution of meiotic recombination events.

The mechanism of meiotic recombination

A large body of genetic, molecular, cytological and biochemical studies have identified numerous steps of meiotic recombination, including the principal DNA intermediates and proteins. These studies have confirmed several key features of the double-strand break repair (DSBR) model [9], and modified some aspects of it, in particular the mode of processing and

resolution of intermediates to yield gene conversion/noncrossover (NCO) and recombinant CO products [10]. Meiotic recombination events have been extensively described for the yeast *Saccharomyces cerevisiae*, and numerous data support the conclusion that the key recombination intermediates and enzymes are similar in all eukaryotes, although there is organism-specific variation [11]. Meiotic recombination involves the formation and repair of 'self-inflicted' DNA double-strand breaks (DSBs) catalyzed by the evolutionarily conserved Spo11 enzyme [12,13] (Fig. 2). Globally, Spo11 proteins have no target sequence specificity, except for a biased nucleotide preference at the cleavage site [14]. Spo11 forms a dimer that cuts the DNA duplex in a transesterification-like reaction that generates covalent 5′-protein–DNA linkages on either side of the break. Then, and probably tightly coupled to DSB formation, Spo11 monomers are removed from the DSB ends as oligonucleotide-bound covalent complexes, leaving behind single-stranded tails [15]. Intriguingly, two populations of Spo11-bound oligonucleotides have been isolated from sporulating

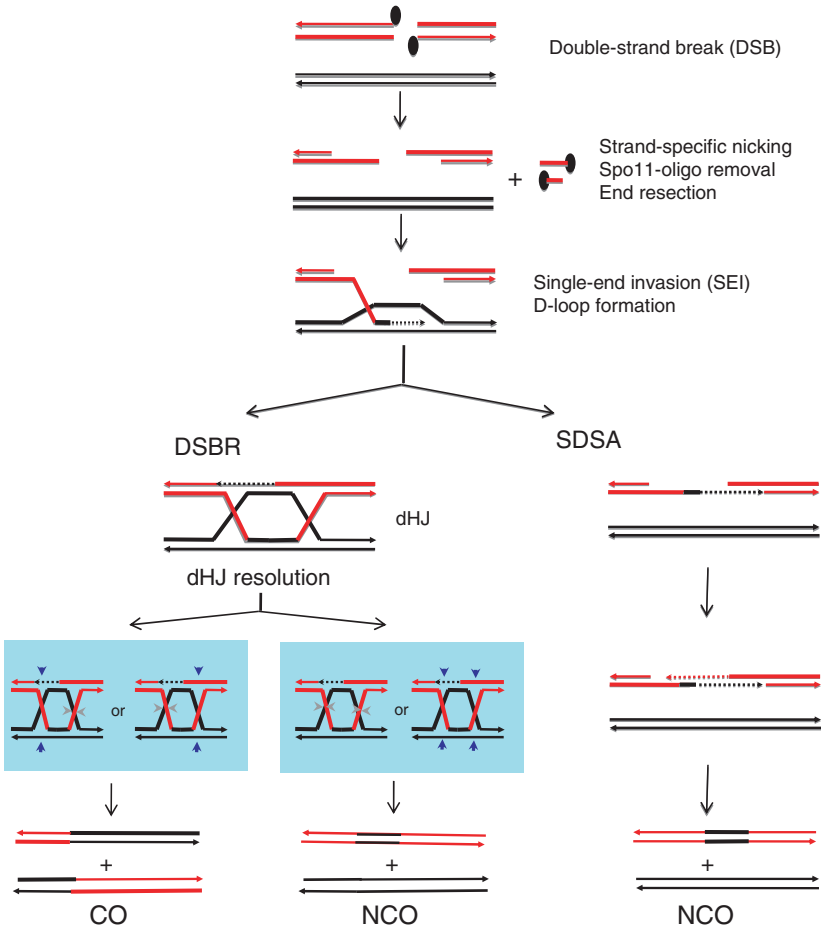


Fig. 2. The mechanism of meiotic recombination. DSBs are formed by the Spo11 protein and associated factors in a topoisomerase II-related reaction. Single-stranded nicks are asymmetrically introduced on either side of the DSB ends, liberating Spo11 subunits covalently attached to a short or a long oligonucleotide. Strand resection is then initiated at these nicks to yield 3′-ssDNA overhangs. One of the 3′-ssDNA tails engages in strand invasion and a homology search of the homologous chromosome, resulting in an SEI intermediate. After D-loop formation, repair follows one of two alternative pathways. In the DSBR pathway, the opposite DSB end is captured by annealing to the displaced strand of the D-loop, leading to the formation of a dHJ. After gap-filling DNA synthesis and nick ligation, the dHJ is symmetrically cleaved on opposing single DNA strands (vertical and horizontal arrowheads), generating products that can be ligated. Depending on cleavage patterns, dHJ resolution produces either CO recombinants or NCO products. In the SDSA pathway, homology-mediated repair of DSBs occurs without the formation of a dHJ. The SEI intermediate undergoes DNA synthesis by extension of the invading DNA strand with D-loop dissolution, and the extended ssDNA ultimately reanneals to its original complementary ssDNA strand on the opposite side of the DSB. An intact duplex is then produced by gap-filling DNA synthesis and nick ligation, which gives rise to an NCO product.

S. cerevisiae cells (10–15 nucleotides and 24–40 nucleotides) and mouse testis (12–26 nucleotides and 28–34 nucleotides). It is not known whether the short and long oligonucleotides reflect different classes of DSBs, either with symmetrical cleavage or controlled asymmetric cleavage. In either case, this conserved heterogeneity might be used to differentially load the Rad51 and Dmc1 recombinases. After cleavage, the DSB ends are further degraded at their 5′-termini by nucleolytic resection to produce recombinogenic 3′-single-stranded tails of heterogeneous length (> 100 nucleotides). This process is mediated by the Mre11–Rad50–Xrs2 complex, which is also involved in DSB formation [16], and by Sae2–Com1 (not required for DSB formation). The nuclease(s) that acts subsequent to Spo11–oligonucleotide formation to produce the 3′-single-stranded tails has not been definitively identified. In addition to Mre11–Rad50–Xrs2 and Sae2–Com1, Exo1, Sgs1 and Dna2, recently characterized for their role in mitotic DSB processing [17–20], remain candidates; whether they are all involved in resection or simply provide potentially overlapping functions in the mutant context is important to determine.

Once sufficient 3′-overhangs are formed, DSBR by homologous recombination is primed to occur with a partner DNA duplex in a strand exchange reaction catalyzed by the Rad51 (which functions during normal DSBR in all cell types) and the meiosis-specific Dmc1 recombinases to yield joint molecule intermediates [21]. Since the identification of Dmc1 [22], the molecular role of this widely (but not always) conserved meiosis-specific strand exchange protein and how it differs from Rad51 have been extensively investigated *in vitro* and *in vivo* [21] (W. Kagawa and H. Kurumisaka, this issue [22a]). This issue is still unresolved but, importantly, it is known that the two proteins do not have redundant functions, as each single mutant exhibits unrepaired DSBs, and each protein's activities are modulated by distinct accessory factors [23]. What are the specific substrates of Rad51 and Dmc1, how do they work in a coordinated way, how does their role extend to controlling other key aspects of meiotic recombination such as partner choice and the NCO/CO decision, and how is recombination driven in organisms such aslike *Schizosaccharomyces pombe*, which lacks a Dmc1 homolog? These are major challenges for the future. Two types of joint molecules have been characterized: the single-end invasion (SEI) intermediate, in which only one end of the DSB is engaged in strand exchange, and the double Holliday junction (dHJ) intermediate, which involves both DSB ends. Strand exchange generating SEI and dHJ intermediates produces heteroduplex DNA containing strand information from both parents, and therefore creates mismatches that are subjected to repair when divergent parental sequences are involved [24]. Finally, the resolution of intermediates ensues, with the restoration of intact and unlinked duplexes. Holliday proposed that symmetric incisions across the bimolecular junction produce ligatable nicks, and that cleavage of alternative pairs of strands produces NCO and CO recombinant products in equal amounts [25] (Fig. 2). However, classic tetrad analyses of linked genetic markers in fungi showed that parity was rarely observed: gene conversions not associated with the exchange of flanking markers (NCOs) were generally in excess over gene conversions associated with an adjacent CO, representing up to 80% of all events at some loci. The CO/NCO ratio varies among diverse subclasses of recombination events: for example, COs are rarely associated with 5 : 3 postmeiotic segregations, which represent unrepaired heteroduplex intermediates [26]. The emergence of alternative models of initiation and strand exchange [27] and the long-standing failure to unambiguously identify 'the' eukaryotic Holliday junction (HJ) resolvase raise the question of how meiotic (as well as mitotic) strand exchange intermediates are resolved. On the basis of powerful molecular analyses of recombination intermediates extracted from synchronized meiotic yeast cells, the contemporary view is that NCOs and COs are derived from alternative processing of early recombination intermediates (Fig. 2). Additional evidence for a mechanistic separation of NCO and CO recombination comes from the molecular study of mutants that block CO formation without reducing that of NCOs [28]. NCO formation involves a synthesis-dependent strand-annealing (SDSA) mechanism in which one DSB end invades the homologous chromosome to prime DNA synthesis, but the nascent DNA strand is then displaced, and, if sufficiently elongated, anneals to the complementary ssDNA tail associated with the other end of the resected DSB. The reaction terminates with gap-filling DNA synthesis and nick ligation, which gives rise only to NCO products [29]. The net product is the transfer of information from the partner chromosome to the repaired DSB chromatid. In contrast, fully ligated SEIs and/or HJs can be resolved to give NCO and/or CO products. Four pathways with evolutionarily conserved orthologous proteins might participate in cleaving HJs: resolution by the BLM–TOPIII–RMI1 helicase–toposiomerase complex [30] and/or the MUS81–EME1 [31], GEN1–YEN1 [32] and SLX1–SLX4 [33,34] pathways. Whether multiple pathways act redundantly or overlap to resolve the same set of HJ-containing intermediates or are specialized for different subsets of intermediates

are essential issues to be addressed. Nonetheless, the obligation that a minimum of one CO per bivalent must be yielded implies that the final number of COs, and therefore the NCO/CO ratio, must be tightly controlled.

Meiotic recombination is obligatory for faithful meiosis

The process of homologous recombination is intrinsic to the success of meiosis. After DNA replication and before chromosome segregation, homologous recombination is recruited to efficiently and faithfully repair an overwhelming burst of self-inflicted DSBs made by Spo11 on every chromosome (Figs 1 and 2). Physical DSB detection and enumeration of Rad51 and γH2AX foci in several organisms have provided an estimate of 150–300 DSBs per meiotic cell, a variable fraction of which will end up as COs if DSBR involves a nonsister chromatid. This outcome entails a selective search for the homologous chromosome and an enhanced risk of nonallelic recombination. Additionally, and not least dauntingly, meiotic recombination events should be properly distributed so as to yield at least one CO per chromosome pair, in order to ensure proper homolog disjunction at MI. Thus, not surprisingly, meiotic recombination is tightly controlled. Mishaps are potentially deleterious and prone to induce *de novo* mutations and other chromosomal abnormalities in progeny, as well as to trigger arrest of the meiotic program. Both kinds of imperfection are sources of infertility.

Errors in the transmission of chromosomes during meiosis can lead to alterations in chromosome number (aneuploidy) in gametes. Upon fecundation, this leads to unbalanced genomes (monosomies or triploidies) in zygotes. In most organisms, owing to physiological selection from the time of parental meiosis through progeny development, the absolute frequencies of unbalanced gametes and the germline rates of *de novo* mutations are difficult to assess. Nonetheless, they are certainly high. In *S. cerevisiae*, the spontaneous frequency of mis-segregation of an individual chromosome is approximately 1 in 10 000, yielding $\sim 0.5\%$ aneuploid spores. In *Drosophila*, where X chromosome nondisjunction in the female has been estimated, there are up to 1 in 1700 spontaneous nondisjunction events per meiosis [35]. The vast majority ($> 90\%$) of these nondisjunctions occur in MI. In the mouse, the overall incidence of monosomies and triploidies among fertilized eggs is $\sim$ 1–2%. For humans, where miscarriage is frequent, the incidence of aneuploidy is 0.3% of live births and 4% among stillbirths. The source of trisomy 21 (Down syndrome) has been well studied [36,37]. We know that: (a) $\sim 80\%$ of segregation errors occur during MI, and 20% result from MII nondisjunction; (b) over 90% of all trisomy 21 cases are of maternal origin, being due to errors in oogenesis, and originate equally from MI and MII nondisjunction events; and (c) the probability of meiotic chromosome segregation errors increases with maternal age, starting around 35 years. Two likely leading causes of mis-segregation in meiosis are abnormalities in sister chromatid cohesion and in recombination.

Chromosome mis-segregation and sister chromatid cohesion

As illustrated in Figs 1 and 3, sister chromatid cohesion allows orderly segregation by holding sister chromatids together from the time of their generation by DNA replication until MI. At this time, chromosomal arm cohesion is removed by separase but maintained at centromeres, protected by the shugoshin protein (Sgo1), the 'guardian spirit at the centromere', and sister kinetochores, which are mono-oriented by the monopolin complex [38]. Thus, sister chromatids continue to associate until the metaphase II to anaphase II transition. The remaining cohesion sites then dissociate, and sister chromatids can be incorporated into haploid gametes. Defects in these processes can result in the premature separation of sister chromatids and chromosome mis-segregation. Another unique aspect of meiotic differentiation is the replacement of the mitotic Scc1–Mcd1 cohesin subunit by the evolutionarily conserved meiosis-specific subunit Rec8 [39]. At MI, activated separase cleaves most Rec8 proteins, causing loss of cohesin from chromosome arms, but not at the centromere, where Rec8 is protected by Sgo1 and additional factors. Thus, cells deleted for Rec8 display defects in chromosome segregation. Interestingly, studies of null and separation-of-function *rec8* mutants and post-translation phosphorylation have revealed that Rec8 is required for the completion of recombinant products [39] and that it is implicated in homolog pairing (by defining the initial alignment of homologous chromosomes) and synaptonemal complex formation [40,41]. These results place Rec8 in the center of multiple meiotic prophase events. Hence, the loading of Rec8 onto chromatin during replication provides meiosis-specific sister chromatid cohesion, and it also permits cells to anticipate and regulate the subsequent cascade of interdependent recombination and chromosomal events. The roles of cohesions in postreplicative DSBR [42] and in chromosomal transactions [43] provide additional reasons

why compromised sister chromatid cohesion may lead to meiotic abnormalities.

Chromosome mis-segregation and recombination

Proper transmission of chromosomes during meiosis also depends on reciprocal recombination, as a CO occurring between homologous, nonsister chromatids is required to provide a physical link between the paternal and maternal chromosomes prior to their bi-orientation on the first meiotic spindle (Fig. 3). The CO generates tension, allowing recombined chromosomes to be pulled away on the metaphase I spindle, while cohesion between sister chromatids distal to chiasmata serves as a 'glue' that holds them together [44]. The essential role of homologous recombination has been demonstrated in numerous studies. In all organisms, when DSB formation is abolished, e.g. by inactivation of Spo11, COs do not form, and homologous chromosomes segregate at random (Fig. 3). When a partial complement of chromosomes is packaged in spores, as in yeasts, spore inviability results. In other organisms, such as *Caenorhabditis elegans* and the mouse, random segregation triggers apoptosis [45,46].

Reduced frequencies of COs and the positions of exchanges can also lead to nondisjunctions. Accordingly, classic genetic mapping techniques for studying the inheritance of DNA polymorphisms in human trisomy 21 patients have allowed the recombinational events that led to trisomy-generating meioses to be recapitulated [36,47]. An estimated 40% of maternal MI-derived cases of trisomy 21 involved an achiasmate bivalent, and in a remaining case, a single CO located near the centromere or in the distal part of the chromosome occurred. Similar observations have been made for wild-type *S. cerevisiae* cells [48], yeast artificial chromosomes [49,50], and *Drosophila* oogenesis [35].

Sex-specific differences and aging are also risk factors. The length of time over which cohesin complexes and chiasmata hold meiotic chromosomes together in mammals varies greatly between males and females. In males, meiosis is a repetitive process over a lifetime, starting at puberty. In females, oocytes start undergoing meiosis during fetal development. Recombination is initiated, but cells enter a period of prolonged diplotene arrest (before MI). Then, meiosis resumes years later at puberty, and continues until menopause. This probably explains why maternal age over 35 years is clearly an important factor in the etiology of human aneuploidy [47]. Over time, the dissolution of sister chromatid cohesion or chiasmata can significantly

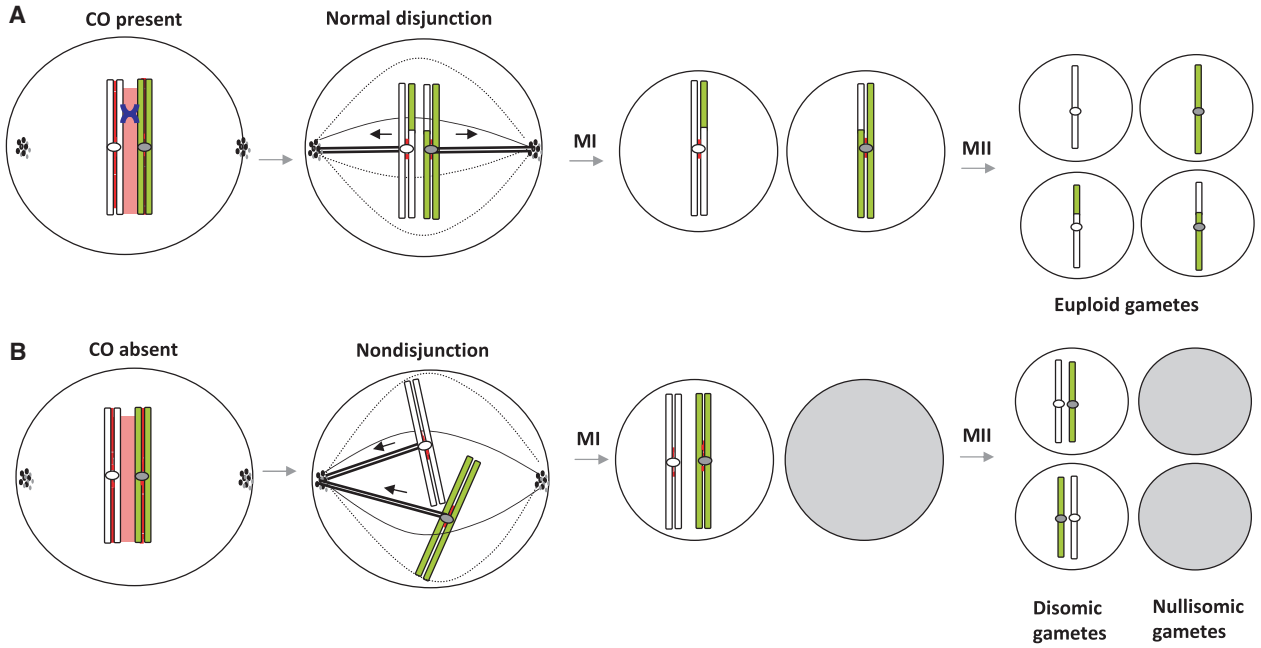


Fig. 3. COs create the connections between homologous chromosomes required for accurate segregation. (A) A CO establishes a physical link between a pair of homologous chromosomes. In MI, the two homologs move towards opposite poles. Sister chromatids separate during MII, leading to the formation of euploid gametes. (B) In the absence of COs, homologous chromosomes are not properly paired. They randomly segregate in MI, generating disomic and nullisomic nuclei. Separation of sister chromatids in MII yields aneuploid gametes.

weaken the links between chromatids and homologs, perturbing meiotic outcomes. Aging also affects meiosis in budding yeast [51]. The consequences of a yeast cell's age are reduced spore viability and failure to enter the meiotic program, in part due to the inability to express the Ime1 master transcription factor and increased chromosome mis-segregation both in MI and in MII. Remarkably, the inability of senescent cells to sporulate can be genetically bypassed by deleting the Sir2 histone deacetylase, suggesting that replicative lifespan controls meiosis, at least in part, through epigenetic mechanisms. In support of this interpretation, a novel Sir2-related aging pathway has been identified that regulates cellular aging in a manner dependent on acetylated histone H4K16 [52]. It would be interesting to examine whether Sir2p orthologs play a role in gametogenesis of other organisms.

Imperfections in meiotic recombination can yield genome rearrangements

Although DSBR by homologous recombination is generally considered to be nonmutagenic, and *de novo* mutations are rare, germline recombination errors occur and can generate genetic diseases [53]. As in somatic cells, nonallelic homologous recombination can generate deletions, inversions, duplications, and translocations. For example, nonallelic homologous recombination is the source of Charcot–Marie–Tooth disease type 1A, hereditary neuropathy with liability to pressure palsies [54], Smith–Magenis syndrome and other syndromes [55,56], and velocardiofacial syndrome [57,58]. Estimates of the meiotic rates of these rearrangements have typically relied on the identification of individuals with the dominant disease phenotype, and are thus likely to be underestimates. Indeed, direct inspection of meiotic products in male germ cells revealed that the above syndromes are undiagnosed in the majority of cases [59]. In the future, improved methods allowing for the direct recognition of these genomic imbalances (e.g. by comparative genomic hybridization arrays) will be useful. Error-prone DSBR mechanisms [60] and the activity of error-prone polymerases in the germline may also contribute to mutations [61]. For instance, microsatellite-related diseases originate in the human germline probably through replication slippage, whereas the frequent contraction and expansion of human minisatellite loci is a consequence of their fortuitous location near natural meiotic recombination initiation sites and the repair of overlapping recombination intermediates by SDSA [62]. The extent of small indel and single-nucleotide polymorphism (SNP) mutagenesis in meiosis is still unknown, but the power of next-generation sequencing technologies should allow precise estimates.

The genetic basis of infertility

In humans, approximately 15% of couples consult for infertility. The underlying causes are heterogeneous, and to a large extent the contribution of genetic factors is unknown. Premature ovarian failure (POF) is a frequent cause of female infertility due to the loss of normal ovarian function in women under 40 years. Several imperfections are probably involved in POF pathogenesis, such as viral or autoimmune inflammatory disease, environmental toxins, and radiation or chemotherapy, but the genetic contribution is also a potential etiological component. Several genes have been suspected of carrying mutations responsible for POF [63], but causal relationships remain difficult to establish in humans, and their significance relies on the number of cases and control samples analyzed [64]. Numerous genes characterized in model organisms have provided valid candidates for mammalian infertility, but, altogether, screening for human infertility mutations remains limited in comparison to that for other prevalent human diseases. In our pilot attempt, we used a sequencing approach to identify mutations of five evolutionarily conserved genes (*DMC1*, *SPO11*, *MSH4*, *MSH5*, and *CCNA1*) in DNA samples from 145 clinically well-characterized patients who presented with unexplained infertility. The panel was composed of 44 samples from infertile women with POF, and 101 men with azoospermia and without a Y microdeletion [65]. Most interestingly, we identified one patient presenting POF with a homozygous mutation of the *DMC1* recombinase (W. Kagawa and H. Kurumizaka, this issue [22a]). Subsequent structural, biochemical and genetic analyses revealed that the responsible M200V mutation partially affects strand exchange activity and reduces meiotic recombination in fission yeast [66]. Altogether, these results suggest that the M200V polymorphism present in heterozygote form in the human population could be a source of infertility, but causality remains to be established. Whether *DMC1* mutations contribute to human male infertility is also an open question. Sex-specific differences in the phenotypes of knockout genes in the mouse are not rare, and, intriguingly, a dominant, recombination-defective allele of *DMC1* causing male-specific sterility has been isolated [67]. How a significant portion of murine female oocytes can compensate for the *DMC1* deficiency to undergo crossing over and complete gametogenesis will be interesting to determine. To pursue high-throughput approaches in humans for candi-

date gene mutations or conduct fruitful association mapping studies, a large collection of DNA from infertile patients needs to be obtained.

Distribution and control of meiotic recombination events

Distribution of DSBs, NCOs and COs

In recent years, the cartography of recombination events in several model organisms (yeasts, plants, nematode, mouse and human) has reached the chromosomal and genome-wide scales. The methods involved include high-density microarray analysis to detect initiating DSBs and recombination products using polymorphic markers, high-throughput determination of linkage disequilibrium in humans, the detection of rare recombinant DNA molecules at hotspots by sperm genotyping, and cytological immunolocalization approaches that allow visualization of CO points in spread pachytene cells. Clearly, the frequencies and the spatial positions of recombination events are not uniform along chromosomes, with the accepted view being that most recombination events occur at highly localized hotspots, whereas large chromosomal regions are cold [68,69]. In yeasts, the frequency of DSBs ranges over a few orders of magnitude throughout the genome. Hotspots have a 10–100-fold higher propensity to form DSBs than do other sites [70]. In *S. cerevisiae*, at the 'strongest' natural hotspots (e.g. *YCR048W/BUD23*), the frequency of DSBs per chromatid can reach up to 10% of DNA molecules [7], and this can rise to 25% at artificially created hotspots [2,14], implying that, in these cases, essentially every meiotic cell has a DSB at that hotspot region (DSBs occur at the four-chromatid stage).

At broad scales, hotspots are distributed on every chromosome and contribute to a large fraction of the total number of COs per genome. However, at the population level, many rarely used sites probably contribute to recombination. In humans, COs appear to cluster within approximately 2 kbp-wide regions, spaced, on average, every 50–100 kbp [71,72], and it is estimated that 72% of human COs overlap a nearby (30 kbp window) recombination hotspot [73]. The other COs probably result from dispersed and rarely used initiation sites. Similar hotspot distribution properties appear to occur in mice [74], *Arabidopsis* [75] and *Sc. pombe*, in which meiotic DSBs are located in large intergenic regions separated by long distances ($\sim$ 65 kbp on average [76]), whereas in *S. cerevisiae*, the DSBs that are located in intergenic regions near promoters are more evenly distributed [77]. In certain chromosomal regions, DSBs form in every promoter, with variable frequencies, whereas DSBs are rare in other large interstitial chromosomal regions, as well as near centromeres and telomeres [78–81].

High-resolution mapping of meiotic recombination events in the progeny of hybrid *S. cerevisiae* diploids carrying high-density SNP differences, but not so high to act as a barrier to recombination, has allowed the recombination landscape of a single meiotic cell to be reconstituted, and thus has allowed both NCO conversion tracts and COs to be examined [82–84]. Microarrays allowing the genotyping of $\sim$ 52 000 SNPs distributed on the 16 chromosomes in 56 tetrads have permitted a resolution with a median distance of 78 bp between constitutive markers. Remarkably, the recombination landscape is different from one meiosis to another, and yet the number of recombination events per tetrad remains constant, with an average of $\sim$ 90 COs and $\sim$ 66 NCOs per meiosis. NCO tracts are typically 1–2 kbp long, and are slightly longer when associated with a nearby CO, in agreement with observations in mice and humans [85,86]. Thus, in budding yeast, the total number of recombination events per meiosis observed on a cell-to-cell basis is similar to the estimate of 150–170 DSBs per meiosis established for a population of cells [80], and consistent with the observation that a majority ($\sim$ 80%) of the DSBs are repaired using the nonsister chromatid as template [87]. Several other important findings have emerged from these approaches. First, the heterogeneous spatial distribution of recombination events along chromosomes correlates well with the heterogeneous distribution of DSBs [77–81,88], and explains discrepancies between genetic and physical distances. A low DSB frequency accounts for the rarity of recombination events near centromeres and subtelomeric regions [77,82,83]. Second, all chromosomes have at least one CO, in agreement with its essential role in chromosome segregation. The average number of COs is linearly related to chromosome length, with an intercept of 1.0 corresponding to the obligate number of COs, plus an additional 6.1 COs per Mbp. In contrast, NCOs occur at an average density of 3.4 NCOs per Mbp, with a low intercept (0.3), consistent with the fact that they do not play a role in chromosome segregation but nevertheless contribute substantially to genetic diversity. These data have allowed the determination of whether COs and NCOs always occur in similar proportions or whether there are CO and NCO hotspots in the genome. Interestingly, approximately 60 regions favorable to COs and $\sim$ 170 favorable to NCOs, spanning 1.4% of the genome, have been identified. In the NCO-biased regions,

the enrichment of genes transcribed at the time of DSB formation is intriguing, and suggests a mechanistic relationship between genetic control of the NCO/CO outcome and transcription factor binding, with its consequences for chromatin accessibility.

The view that the spatial distribution of COs is tightly controlled on a single-cell basis is emphasized by three other manifestations of CO control, illustrated in Fig. 4. The first is a recently discovered process known as DSB interference [88], in which the targeted induction of DSBs by GAL4BD–Spo11 was found to reduce the DSB frequencies at nearby natural hotspots (Fig. 4A).

The second manifestation is the process known as CO interference [89] (Fig. 4B). Interference refers to the observation that a CO in one chromosomal region reduces the probability that a CO will occur simultaneously in an adjacent region, therefore creating a

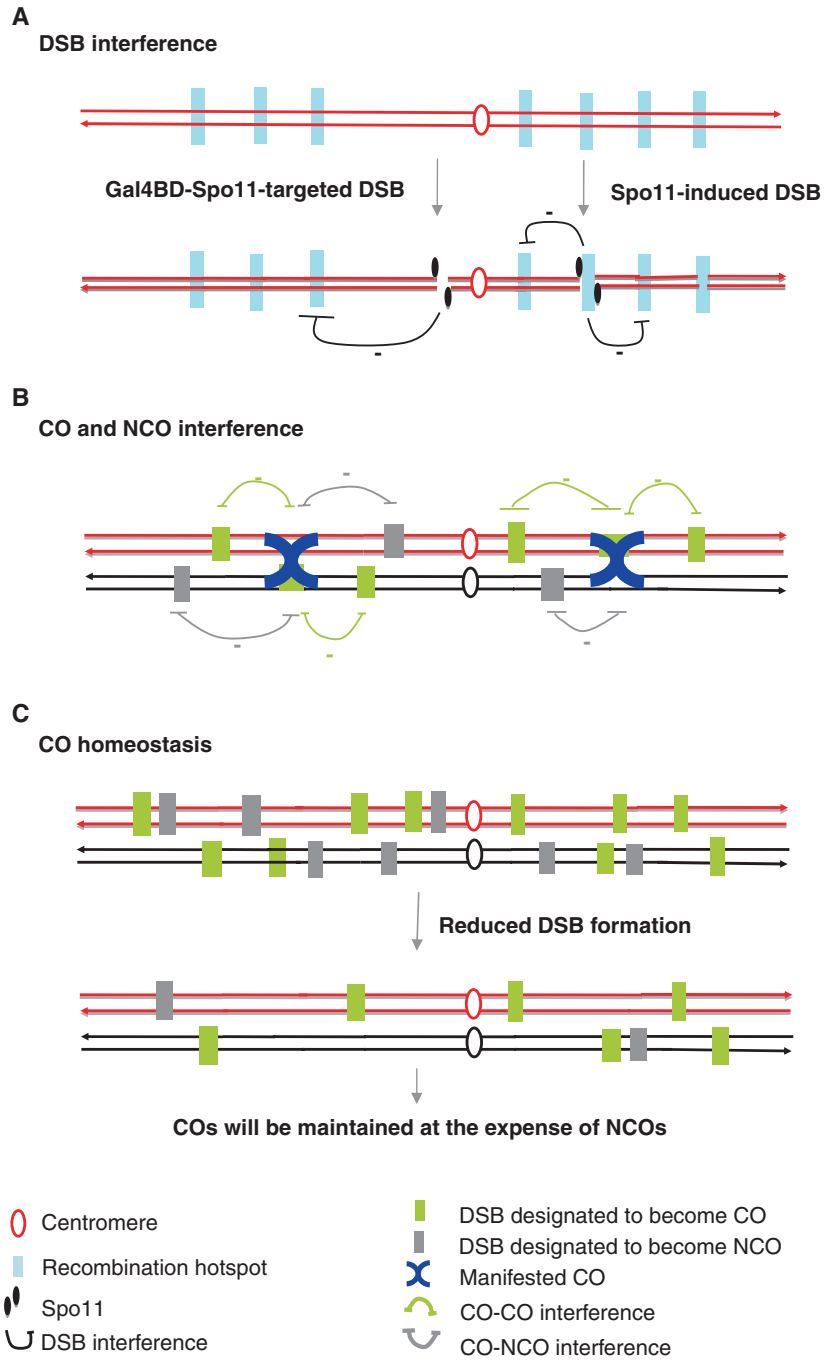


Fig. 4. The control of meiotic recombination. (A) DSB interference. Along chromosomes, only a fraction of recombination hotspots undergo Spo11-dependent DSB formation. Interference between DSBs occurring at targeted sites and natural hotspots shapes the chromosomal DSB profile [88]. (B) CO and NCO interference. A subset of chromosomal DSBs is designated to become COs and NCOs. The presence of one CO inhibits the coincident occurrence of another CO in its vicinity (CO–CO interference), causing them to be widely spaced [89]. As the average distance between COs and NCOs is significantly greater than expected from chance [83], COs and NCOs also appear to interfere with each other (CO–NCO interference). (C) CO homeostasis. A reduction in the number of DSBs does not lead to a correlated decrease in the number of COs [91]. The CO/NCO ratio increases, maintaining COs at the expense of NCOs.

more regular spacing between COs than would be expected on the basis of a random distribution. CO interference is commonly visualized either genetically by monitoring the distribution of the CO events on multiply marked chromosomes, or by cytological methods to visualize chiasmata or recombination-related foci of CO-specific proteins such as Mlh1 [90]. With a few exceptions, most organisms exhibit CO interference, which acts strongly over short distances and decreases in intensity with increasing distances along a chromosome, but which still extends over large physical distances (> 100 Mbp in mammals). The mechanism of CO interference has still not been elucidated but it is clearly genetically controlled: numerous mutations that reduce or abolish CO interference have been identified. Interestingly, several mutations disturb initiation of the SC (*zip2* and *zip4*) and also DNA strand exchange structures (ZMM mutants), raising the hypothesis that the decision for a CO rather than an NCO outcome might be made early, around the time of DSB formation. The third manifestation of CO control is CO homeostasis (Fig. 4C) [91]. This process maintains COs at a relatively constant number per cell when the number of DSBs is reduced (as, for example, in Spo11-leaky mutants). The benefit of this control is to reinforce the obligatory outcome of one CO per chromosome pair and thus reroute NCOs into COs. The establishment of the global recombination landscape in yeast has confirmed this phenomenon, and led to two new and unexpected findings. First, ZMM mutants defective in CO interference (*zip2* and *zip4*) also exhibit a reduced level of CO homeostasis, a genetic linkage that promises an interesting expansion of our understanding of the underlying molecular events [82]. Second, in contrast to a previous assumption that only COs are subject to interference, COs and NCOs also interfere with each other (the median distance between these sites is greater than expected from a random distribution). This result is further substantiated by the observation that both CO–CO and CO–NCO interferences are absent in the *msh4* mutant. Altogether, these results add to the view that CO control in meiosis is a key to enforcing the 'obligatory' CO per chromosome and, at the same time, it illustrates the somehow paradoxical observation that, on a cell-to-cell basis, the distribution of recombination events remains remarkably flexible.

What makes a recombination site?

The distribution of recombination events varies significantly in each cell and between individuals. The obligatory CO per chromosome in a flexible context raises the question of what makes a recombination site – a DNA sequence, a specific DNA–protein interaction, and/or a chromatin structure – and how one site differs from another.

Modifying and targeting meiotic recombination

The manifestations of recombination flexibility are numerous. Key observations include the large number of recombination sites per genome and the apparent stochasticity of their activity. The high density of potential sites is well suited to produce extensive and finely scaled genetic diversity within a population, whereas partial activity at each site preserves the haplotypic structure of the species. Besides chromosomal and genome-wide *cis*-acting and *trans*-acting factors, local factors that predispose a specific region or site to DSB formation (and hence recombination) are likely to play a significant role in creating recombination-competent sites. Regarding genome-wide *trans*-acting factors, a number of genes in various organisms, from fungi to mammals, have been identified that, when mutated, confer recombination defects from initiation to resolution [92]. In *S. cerevisiae*, extensive efforts have been made to characterize the proteins that promote DSB formation [21]. To date, 10 proteins, mostly expressed early and specifically in meiotic prophase, including Spo11, are required for DSB formation. They are related by a network of physical and functional interactions and have been schematically structured into four multiprotein subcomplexes, namely Spo11–Ski8, Rec102–Rec104, Rec114–Mer2–Mei4, and Mre11–Rad50–Xrs2 (NBS1), which is also involved in mitotic DSB repair. Null mutation of any of these proteins leads to the absence of DSBs, abnormal synaptonemal complexes, and complete spore inviability. Little is known about their molecular functions. Beyond the well-established role of Spo11 in DSB induction, Ski8 helps recruit Rec102–Rec104 to chromosomes [93], and Mer2, which is phosphorylated by Cdc7–Dbf4 and the cyclin-dependent kinase Cdc28 in complex with the B-type cyclin Clb5–Clb6, provides a functional link between replication and DSB formation [94] by modulating the loading of interacting proteins onto chromatin. Notably, aside from Spo11, which is evolutionarily conserved, several of the DSB proteins identified in *S. cerevisiae* have no obvious orthologs in *Sc. pombe* or in other organisms, and, conversely, some *Sc. pombe* DSB proteins are apparently unrepresented in *S. cerevisiae* [95]. In the future, functional orthologs without recognizable sequence homology may be uncovered, but it is also meaningful to consider that the defining characteristics of a DSB

site are embedded in species-specific features. In this respect, environmental factors (temperature or chemical composition of media, for example) that trigger a large spectrum of physiological changes have been found to modulate meiotic recombination [96]. Such external alterations may causes molecular changes that affect the activity and/or substrate specificity of transcription factors, or modify chromatin structures, and thus contribute to the activation of dormant recombination sites.

The possibility of artificially targeting meiotic recombination to naturally cold regions has also revealed the existence of rarely used but potentially competent recombination sites. In *S. cerevisiae*, the fusion of Spo11 or other DSB proteins to the sequence-specific DNA-binding domain of Gal4 (Gal4BD–Spo11) or to the synthetic zinc-finger motif (QQR–Spo11) is sufficient to target DSB formation to regions containing the consensus binding sequence of Gal4, in the former case, and to create recombination hotspots [97,98] (V. Borde & N. Uematsu, personal communication). As the Gal4–Spo11 fusion protein binds to approximately 500 sites in the *S. cerevisiae* genome, the genome-wide mapping of Gal4BD–Spo11 cleavage sites revealed that DSB formation could be stimulated in numerous naturally 'cold' regions, leading to a substantial modification of its natural distribution [88]. The DSB profiles in Gal4BD–Spo11 and QQR–Spo11 strains are different from one another (V. Borde, personal communication), owing to the distinct locations of the targeted sites and of long-range (> 100 kbp) repression effects in the chromosomal regions next to the newly induced hotspots (Fig. 4A) [88]. Importantly, it should be noted that Gal4BD–Spo11 binding to meiotic chromatin is not sufficient for Spo11 cleavage, leading to the idea that chromosomal regions can be categorized as 'naturally permissive', 'cold' but having the potential to become activated, and 'refractory' for DSB formation owing to chromosomal context, as in the case of Gal4BD–Spo11 binding in a centromere-proximal region.

In fungi and higher organisms, recombination is also modulated by *cis*-acting factors. In *S. cerevisiae*, local modification of Spo11-dependent DSB frequencies is obtained by: (a) deletion of a *cis*-acting element locally controlling DSB formation [99]; (b) insertion/substitution of ectopic or foreign DNA fragments [100]; (c) transcription across the DSB region [101]; or (d) modification of chromatin associated-factors, including transcription factors [102]. A single-nucleotide change can also create or inactivate a hotspot, as in the case of the *ade6-M26* mutation in *Sc. pombe*, which is a single G/T transversion, sufficient to create the cAMP-responsive element-like heptanucleotide binding sequence for the Atf1–Pcr1 transcription factor, which locally induces a favorable chromatin reorganization and allows the initiation of recombination [103]. As similar nucleotide motifs are present in other regions of the genome, some are natural recombination hotspots [104,105]. However, in *Sc. pombe*, *S. cerevisiae*, mice and humans, most hotspots do not share substantial sequence homology, or at best, share only weak homology [71]. A unique 'recombination site' consensus sequence is not in prospect, but subsets of motifs dependent on the same sequence-specific regulatory factors can be expected. DSBs preferentially occur in intergenic regions near promoters in *S. cerevisiae*, and in long, intergenic regions in *Sc. pombe*, but this relationship with gene organization may be indirect. Instead of, or in addition to, primary DNA sequences, it is more likely that elements of chromatin structure define recombination sites.

The role of chromatin remodeling and histone modifications

Experiments in yeasts have indicated that: (a) hotspots exhibit nuclease (MNase and DNase I) hypersensitivity [106,107]; (b) an open chromatin configuration is insufficient for DSB formation [108]; (c) some, but not all, loci undergo meiosis-specific alterations in nuclease sensitivity prior to DSB formation under the dependence of some DSB proteins (as, for example, Mre11, Rad50, Xrs2, Mre2) [109]; (d) the insertion of a nucleosome-excluding sequence into the genome creates a recombination hotspot [110]; and (e) chromatin modifications associated with transcription factor binding stimulate hotspot activity [106].

Covalent post-translational modifications of histones are numerous and are known to have important functions in replication, transcription, repair and other aspects of eukaryotic chromosome dynamics in somatic cells [111]. Their roles in meiosis have not been extensively examined. Table 1 lists studies in various organisms that have addressed the roles of histone acetylation, methylation, ubiquitinylation and phosphorylation upon mutation of histone amino acids or histone-modifying enzymes. The replacement of histones during mammalian late spermatogenesis is reviewed by Gaucher *et al.* (this issue [111a]). Perturbation of histone modifications affects meiotic replication, DSB formation, DSBR, and chromosome condensation, and leads to reduced sporulation and infertility. The effects on DSB formation can be global or local. For example, deletion of the gene encoding the GCN5 histone acetyltransferase, which acetylates

Table 1. Studies of histone modifications in meiosis. Sc, *S. cerevisiae*; Sp, *Sc. pombe*; Ce, *C. elegans*; Mm, *Mus musculus*.

Modification studied	Mutation	Species	Main effect	Reference
Acetylation				
H3K9/14/18ac	*gcn5-21*	Sc	Replication defect; reduced DSBs at *THR4*	123
H3K27ac, H4K12ac	*rpd3Δ*, *hda1Δ*	Sc	Increased DSBs at *HIS4*	112
H3K56ac	*asf1Δ*, *h3k56Q/R*	Sc	Reduced sporulation	126
H4K16ac	*sir2Δ*	Sc	Increased and reduced DSB levels at various sites	114
H3ac, H4ac	*gcn5Δ*	Sp	Delay in chromatin remodeling and partial reduction of meiotic recombination frequency at *M26*	130
H3K9ac, H4ac	–	Mm	Specific enrichment of H3K9ac and H4ac at active Psmb9 and Hlx1 hotspots	119
Methylation				
H3K4me	*set1Δ*	Sc	Global reduction of DSB formation; altered meiotic gene expression	116
H3K36me	*set2Δ*	Sc	Increased DSBs at *HIS4*; temperature-sensitive reduced sporulation	112
H3K79me	*dot1Δ*	Sc	No effect on meiosis	112
H3K9me	*him-17*	Ce	Reduced DSB formation; delay in the accumulation of H3MeK9 on germline chromatin	127
H3K4me3	*prmd9Δ*	Mm	Impaired DSBR; infertility	120
H3K4me2/3	–	Mm	Specific enrichment of H3K4me3 at the Psmb9 hotspot	119
Ubiquitination				
H2B123ub	*rad6Δ*, *h2B123R*	Sc	Reduced DSB formation and sporulation	115
H2B123ub	*rhp6Δ*	Sp	Reduced sporulation	128
H2B123ub	*hr6BΔ*	Mm	Increased apoptosis of primary spermatocytes; damaged synaptonemal complexes; male infertility	129
Phosphorylation				
H3S1ph	*h3S1A*	Sc	Reduced sporulation	124
H3S10ph	*h3S10A*	Sc	No effect on meiosis	122
H2AS139ph (γ-H2AX)	–	Mm	Colocalization of γ-H2AX with Rad51 and Dmc1 foci during meiotic prophase	125

N-terminal lysines on histones H2B and H3, decreases recombination at the *S. cerevisiae HIS4* and *Sc. pombe ade-M26* hotspots, respectively. Mutations of the histone deacetylases Rpd3 and Hda1 strongly stimulate DSB formation and recombination at the *HIS4* locus [112], probably upon acetylation of histone H3K27 and histone H4K12. Also, inactivation of the Set2 methylase results in stimulation of DSBs at *HIS4*, suggesting that Set2-mediated histone H3K36 methylation leads to recruitment of Rpd3 to its sites of action [113]. The loss of methylated histone H3K36-dependent recruitment of Rpd3 (in a *set2* strain) or suppression of histone deacetylation (in an *rpd3* strain) results in hyperacetylated chromatin at the *HIS4* region, which might facilitate the entry of the Spo11 complex and give rise to more DSBs. Sir2 is another histone deacetylase in *S. cerevisiae*. Deletion of the *SIR2* gene has a broad but still uneven effect on DSB formation: elevating DSB frequencies in 5% of the genes, and reducing them in 7% [114]. Increased frequencies of DSBs were clearly detected in naturally cold regions, such as centromere-adjacent and telomere-adjacent regions (within 10 kbp), within the rRNA gene cluster, and in other genes scattered throughout the genome. In the absence of Sir2, elevated levels of histone H3K16 acetylation may lead to a more open chromatin structure that allows Spo11 access to DNA.

Another interesting link between the control of DSB formation and histone modifications has been uncovered by a study of the *rad6* and *set1* mutants in *S. cerevisiae*. *RAD6* encodes an E2 ubiquitin-conjugating enzyme that is targeted by the E3 ubiquitin ligase Bre1 and ubiquitinylates histone H2BK123. The deletion of *RAD6* as well as the histone H2B K123R mutation were found to severely reduce DSB frequencies along chromosome III without changing their distribution [115]. This effect is probably mediated through histone H3K4 methylation, as histone H2BK123 ubiquitination promotes histone H3K4 methylation, and deletion of the *SET1* gene, which encodes the only histone H3K4 methyltransferase, severely reduces meiotic DSB formation in 84% of hotspots [116,117] (Fig. 5A,B). At some sites (e.g. PES4), however, DSBs are strongly stimulated in the absence of methylated histone H3K4, which is another sign of flexibility in the distribution of recombination initiation events (Fig. 5C).

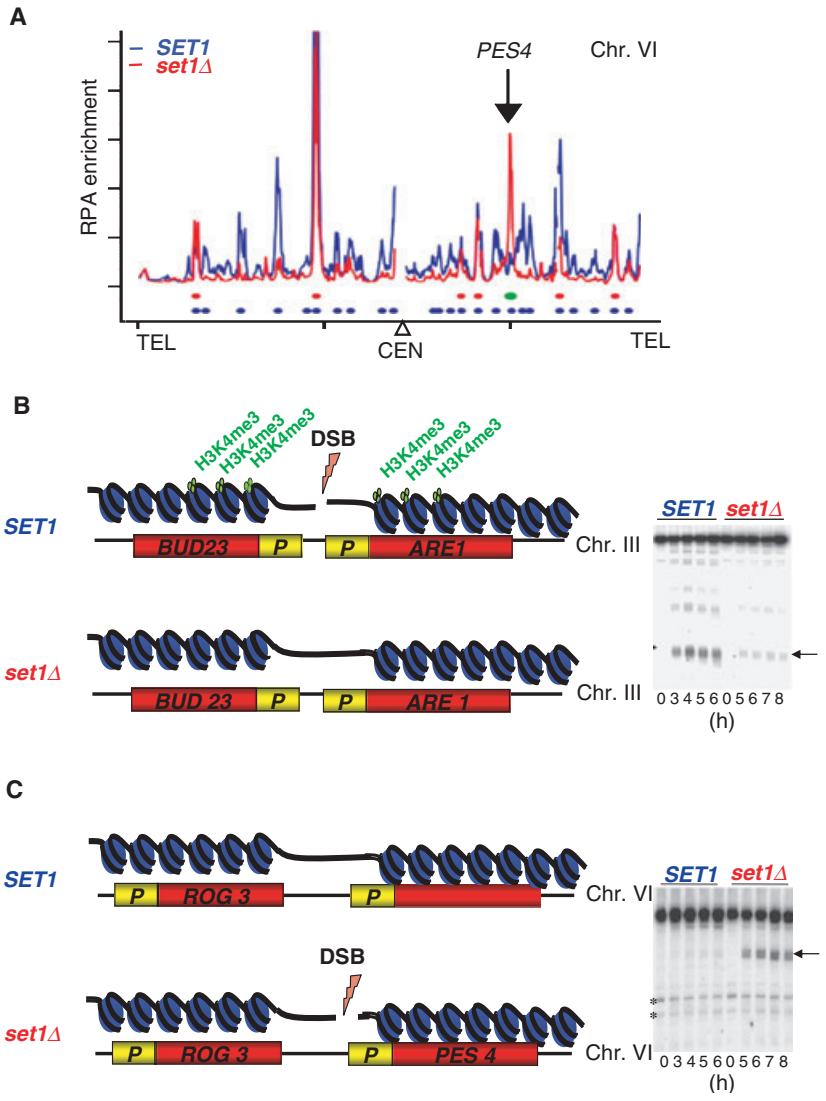


Fig. 5. Histone H3K4 methylation affects the localization and frequency of meiotic DSBs. (A) Profile of meiotic DSB-associated ssDNA enrichment in *S. cerevisiae* wild-type (*SET1*) and *set1Δ* strains. Meiotic DSB profiles of *SET1* and *set1Δ* strains were determined by RPA chromatin immunoprecipitation analysis [116], revealing a global reduction in DSBs in the absence of histone H3K4 methylation. Blue line, wild-type *SET1* strain; red line, *set1Δ* strain; *y*-axis, level of DSB-associated RPA enrichment; *x*-axis, chromosomal coordinates; blue (wild type) and red (*set1Δ*) circles, DSB peaks. (B) The number of DSBs decreases in the absence of histone H3K4 methylation. In the *set1Δ* strain, DSB formation is strongly reduced within the intergenic region of the *BUD23* and *ARE1* loci. At right: Southern blot analysis of DSBs at the *BUD23* hotspot in *SET1* and *set1Δ* cells [116]. Arrow: DSBs. (C) Stimulation of DSBs in the absence of histone H3K4 methylation. Enhanced DSB formation occurs at the *PES4* locus in the *set1Δ* strain. At right: Southern blot analysis of DSB formation at *PES4* in *SET1* and *set1Δ* cells [116].

Genome-wide analyses revealed that the level of trimethylated histone H4K4 is constitutively higher close to DSB sites, independently of local gene expression levels. As this differential histone marker is present in vegetative cells, and at higher levels in DSB-prone regions than in regions with no or few DSBs, H3K4 trimethylation may set the stage for future meiotic DNA breaks [116,118]. Consistently, an enrichment of dimethylated histone H3K4 has been recently observed at two active mouse hotspots [119], and may be dependent on the *MEISETZ/PRMD9* gene, which encodes a histone methyltransferase specifically expressed in meiotic cells. In *meisetz*$^{-/-}$ spermatocytes, the level of trimethylated histone H3K4 is reduced as compared with that in wild-type cells, and gametogenesis is perturbed at the pachytene stage [120]. In conclusion, a growing body of evidence showing the influence of histone modification and chromatin dynamics on recombination initiation has been accumulated, but how the DSB-forming machinery is influenced remains to be elucidated [118].

Concluding remarks

Here, we have reviewed recent advances in our understanding of how meiotic cells create and position the obligatory COs that ensure correct chromosome disjunction without relying on the fortuitous distribution of chromosomes to create a balanced genome in gametes. At the same time, it is somehow paradoxical that, on a cell-to-cell basis, the chromosomal profile of recombination events is not constrained, but instead is flexible. This is well suited to generate genetic diversity, but what defines recombination sites at the molecular

level and explains their large number and diversity in the same organism and in distinct organisms from fungi to humans remains to be determined. For example, a surprising observation is that, in contrast to the situation in the rest of their genomes, human and chimpanzee recombination hotspots are not well conserved, indicating that the recombination landscape has changed markedly between the two species [121], and underscoring the fascinating issue of genome nature and plasticity. The lack of hotspot activity at the 13 bp 'consensus motif' in the chimpanzee suggests that distinct recombination-promoting sequence features operate in the two species.

The growing evidence indicates that instead of, or in addition to, primary DNA sequences, elements of chromatin structure are more likely the common denominators of recombination initiation sites and provides a novel framework to draw hypothesis: (a) the apparent stochasticity of meiotic recombination initiation may also reflect pre-existing cell-to-cell variation of chromatin structure in the mitotic lineages, which is then passively used in meiosis; and (b) over time, regulating chromatin structures (in particular, chromatin opening) might be easier than changing the DNA sequence. For organisms subjected to environmental fluctuations (like the single-cell eukaryote *S. cerevisiae*, in which the entry into meiosis results from nutrient starvation, and meiotic cells can return to mitotic growth even after the induction of high levels of homologous recombination [7]), 'obligatory flexibility' of recombination initiation site distribution may be a transient and rapid strategy to create genetic diversity in diploid cells.

Acknowledgements

We are grateful to K. Smith for critically reading the manuscript. This work was supported by grants from the ANR (BLANC06-3-150811) and the CNRS-GDR2585-CNRS. L. Székvölgyi has received funding from the European Union in terms of the Seventh Framework Program (FP7/People/Marie Curie Actions/IEF).

References

1 Zimmer C (2009) Origins. On the origin of sexual reproduction. *Science* **324**, 1254–1256.
2 Borner GV, Kleckner N & Hunter N (2004) Crossover/noncrossover differentiation, synaptonemal complex formation, and regulatory surveillance at the leptotene/zygotene transition of meiosis. *Cell* **117**, 29–45.
3 Zickler D & Kleckner N (1999) Meiotic chromosomes: integrating structure and function. *Annu Rev Genet* **33**, 603–754.
4 Keeney S (2001) Mechanism and control of meiotic recombination initiation. *Curr Top Dev Biol* **52**, 1–53.
5 Keeney S & Neale MJ (2006) Initiation of meiotic recombination by formation of DNA double-strand breaks: mechanism and regulation. *Biochem Soc Trans* **34**, 523–525.
6 Longhese MP, Bonetti D, Guerini I, Manfrini N & Clerici M (2009) DNA double-strand breaks in meiosis: checking their formation, processing and repair. *DNA Repair (Amst)* **8**, 1127–1138.
7 Simchen G (2009) Commitment to meiosis: what determines the mode of division in budding yeast? *BioEssays* **31**, 169–177.
8 Smith KN & Nicolas A (1998) Recombination at work for meiosis. *Curr Opin Genet Dev* **8**, 200–211.
9 Szostak JW, Orr-Weaver TL, Rothstein RJ & Stahl FW (1983) The double-strand-break repair model for recombination. *Cell* **33**, 25–35.
10 Allers T & Lichten M (2001) Intermediates of yeast meiotic recombination contain heteroduplex DNA. *Mol Cell* **8**, 225–231.
11 Cromie GA, Hyppa RW, Taylor AF, Zakharyevich K, Hunter N & Smith GR (2006) Single Holliday junctions are intermediates of meiotic recombination. *Cell* **127**, 1167–1178.
12 Bergerat A, de Massy B, Gadelle D, Varoutas PC, Nicolas A & Forterre P (1997) An atypical topoisomerase II from Archaea with implications for meiotic recombination. *Nature* **386**, 414–417.
13 Keeney S, Giroux CN & Kleckner N (1997) Meiosis-specific DNA double-strand breaks are catalyzed by Spo11, a member of a widely conserved protein family. *Cell* **88**, 375–384.
14 Murakami H & Nicolas A (2009) Locally, meiotic double-strand breaks targeted by Gal4BD–Spo11 occur at discrete sites with a sequence preference. *Mol Cell Biol* **29**, 3500–3516.
15 Neale MJ, Pan J & Keeney S (2005) Endonucleolytic processing of covalent protein-linked DNA double-strand breaks. *Nature* **436**, 1053–1057.
16 Borde V & Cobb J (2009) Double functions for the Mre11 complex during DNA double-strand break repair and replication. *Int J Biochem Cell Biol* **41**, 1249–1253.
17 Gravel S, Chapman JR, Magill C & Jackson SP (2008) DNA helicases Sgs1 and BLM promote DNA double-strand break resection. *Genes Dev* **22**, 2767–2772.
18 Mimitou EP & Symington LS (2008) Sae2, Exo1 and Sgs1 collaborate in DNA double-strand break processing. *Nature* **455**, 770–774.
19 Nimonkar AV, Ozsoy AZ, Genschel J, Modrich P & Kowalczykowski SC (2008) Human exonuclease 1 and

BLM helicase interact to resect DNA and initiate DNA repair. *Proc Natl Acad Sci USA* **105**, 16906–16911.
20 Zhu Z, Chung WH, Shim EY, Lee SE & Ira G (2008) Sgs1 helicase and two nucleases Dna2 and Exo1 resect DNA double-strand break ends. *Cell* **134**, 981–994.
21 Neale MJ & Keeney S (2006) Clarifying the mechanics of DNA strand exchange in meiotic recombination. *Nature* **442**, 153–158.
22 Bishop DK, Park D, Xu L & Kleckner N (1992) DMC1: a meiosis-specific yeast homolog of E. coli recA required for recombination, synaptonemal complex formation, and cell cycle progression. *Cell* **69**, 439–456.
22a Kagawa W & Kurumizaka H (2009) From meiosis to postmeiotic events: Uncovering the molecular roles of the meiosis-specific recombinase Dmc1. *FEBS J* **277**, 590–598.
23 Shinohara M, Oh SD, Hunter N & Shinohara A (2008) Crossover assurance and crossover interference are distinctly regulated by the ZMM proteins during yeast meiosis. *Nat Genet* **40**, 299–309.
24 Lichten M, Goyon C, Schultes NP, Treco D, Szostak JW, Haber JE & Nicolas A (1990) Detection of heteroduplex DNA molecules among the products of *Saccharomyces cerevisiae* meiosis. *Proc Natl Acad Sci USA* **87**, 7653–7657.
25 Holliday R (1964) A mechanism for gene conversion in fungi. *Genet Res* **5**, 282–304.
26 Nicolas A & Rossignol JL (1983) Gene conversion: point-mutation heterozygosities lower heteroduplex formation. *EMBO J* **2**, 2265–2270.
27 Meselson MS & Radding CM (1975) A general model for genetic recombination. *Proc Natl Acad Sci USA* **72**, 358–361.
28 Bishop DK & Zickler D (2004) Early decision; meiotic crossover interference prior to stable strand exchange and synapsis. *Cell* **117**, 9–15.
29 Paques F & Haber JE (1999) Multiple pathways of recombination induced by double-strand breaks in *Saccharomyces cerevisiae*. *Microbiol Mol Biol Rev* **63**, 349–404.
30 Wu L & Hickson ID (2006) DNA helicases required for homologous recombination and repair of damaged replication forks. *Annu Rev Genet* **40**, 279–306.
31 Constantinou A, Chen XB, McGowan CH & West SC (2002) Holliday junction resolution in human cells: two junction endonucleases with distinct substrate specificities. *EMBO J* **21**, 5577–5585.
32 Ip SC, Rass U, Blanco MG, Flynn HR, Skehel JM & West SC (2008) Identification of Holliday junction resolvases from humans and yeast. *Nature* **456**, 357–361.
33 Fekairi S, Scaglione S, Chahwan C, Taylor ER, Tissier A, Coulon S, Dong MQ, Ruse C, Yates JR III, Russell P *et al*. (2009) Human SLX4 is a Holliday junction resolvase subunit that binds multiple DNA repair/recombination endonucleases. *Cell* **138**, 78–89.
34 Svendsen JM, Smogorzewska A, Sowa ME, O'Connell BC, Gygi SP, Elledge SJ & Harper JW (2009) Mammalian BTBD12/SLX4 assembles a Holliday junction resolvase and is required for DNA repair. *Cell* **138**, 63–77.
35 Koehler KE, Boulton CL, Collins HE, French RL, Herman KC, Lacefield SM, Madden LD, Schuetz CD & Hawley RS (1996) Spontaneous X chromosome MI and MII nondisjunction events in *Drosophila melanogaster* oocytes have different recombinational histories. *Nat Genet* **14**, 406–414.
36 Hassold T & Hunt P (2007) Rescuing distal crossovers. *Nat Genet* **39**, 1187–1188.
37 Lamb NE, Yu K, Shaffer J, Feingold E & Sherman SL (2005) Association between maternal age and meiotic recombination for trisomy 21. *Am J Hum Genet* **76**, 91–99.
38 Watanabe Y (2005) Shugoshin: guardian spirit at the centromere. *Curr Opin Cell Biol* **17**, 590–595.
39 Klein F, Mahr P, Galova M, Buonomo SB, Michaelis C, Nairz K & Nasmyth K (1999) A central role for cohesins in sister chromatid cohesion, formation of axial elements, and recombination during yeast meiosis. *Cell* **98**, 91–103.
40 Brar GA, Kiburz BM, Zhang Y, Kim JE, White F & Amon A (2006) Rec8 phosphorylation and recombination promote the step-wise loss of cohesins in meiosis. *Nature* **441**, 532–536.
41 Brar GA, Hochwagen A, Ee LS & Amon A (2009) The multiple roles of cohesin in meiotic chromosome morphogenesis and pairing. *Mol Biol Cell* **20**, 1030–1047.
42 Strom L, Karlsson C, Lindroos HB, Wedahl S, Katou Y, Shirahige K & Sjogren C (2007) Postreplicative formation of cohesion is required for repair and induced by a single DNA break. *Science* **317**, 242–245.
43 Hadjur S, Williams LM, Ryan NK, Cobb BS, Sexton T, Fraser P, Fisher AG & Merkenschlager M (2009) Cohesins form chromosomal cis-interactions at the developmentally regulated IFNG locus. *Nature* **460**, 410–413.
44 Petronczki M, Siomos MF & Nasmyth K (2003) Un menage a quatre: the molecular biology of chromosome segregation in meiosis. *Cell* **112**, 423–440.
45 Baudat F, Manova K, Yuen JP, Jasin M & Keeney S (2000) Chromosome synapsis defects and sexually dimorphic meiotic progression in mice lacking Spo11. *Mol Cell* **6**, 989–998.
46 Romanienko PJ & Camerini-Otero RD (2000) The mouse Spo11 gene is required for meiotic chromosome synapsis. *Mol Cell* **6**, 975–987.
47 Hassold T, Hall H & Hunt P (2007) The origin of human aneuploidy: where we have been, where we are going. *Hum Mol Genet* **16**(Spec No. 2), R203–R208.

48 Rockmill B, Voelkel-Meiman K & Roeder GS (2006) Centromere-proximal crossovers are associated with precocious separation of sister chromatids during meiosis in *Saccharomyces cerevisiae*. *Genetics* **174**, 1745–1754.
49 Ross LO, Treco D, Nicolas A, Szostak JW & Dawson D (1992) Meiotic recombination on artificial chromosomes in yeast. *Genetics* **131**, 541–550.
50 Sears DD, Hieter P & Simchen G (1994) An implanted recombination hot spot stimulates recombination and enhances sister chromatid cohesion of heterologous YACs during yeast meiosis. *Genetics* **138**, 1055–1065.
51 Boselli M, Rock J, Unal E, Levine SS & Amon A (2009) Effects of age on meiosis in budding yeast. *Dev Cell* **16**, 844–855.
52 Dang W, Steffen KK, Perry R, Dorsey JA, Johnson FB, Shilatifard A, Kaeberlein M, Kennedy BK & Berger SL (2009) Histone H4 lysine 16 acetylation regulates cellular lifespan. *Nature* **459**, 802–807.
53 Arnheim N & Calabrese P (2009) Understanding what determines the frequency and pattern of human germline mutations. *Nat Rev Genet* **10**, 478–488.
54 Chance PF & Lupski JR (1994) Inherited neuropathies: Charcot–Marie–Tooth disease and related disorders. *Baillieres Clin Neurol* **3**, 373–385.
55 Long FL, Duckett DP, Billam LJ, Williams DK & Crolla JA (1998) Triplication of 15q11–q13 with inv dup(15) in a female with developmental delay. *J Med Genet* **35**, 425–428.
56 Potocki L, Chen KS, Park SS, Osterholm DE, Withers MA, Kimonis V, Summers AM, Meschino WS, Anyane-Yeboa K, Kashork CD *et al*. (2000) Molecular mechanism for duplication 17p11.2 – the homologous recombination reciprocal of the Smith–Magenis microdeletion. *Nat Genet* **24**, 84–87.
57 Edelmann L, Pandita RK, Spiteri E, Funke B, Goldberg R, Palanisamy N, Chaganti RS, Magenis E, Shprintzen RJ & Morrow BE (1999) A common molecular basis for rearrangement disorders on chromosome 22q11. *Hum Mol Genet* **8**, 1157–1167.
58 Kirchhoff M, Bisgaard AM, Duno M, Hansen FJ & Schwartz M (2007) A 17q21.31 microduplication, reciprocal to the newly described 17q21.31 microdeletion, in a girl with severe psychomotor developmental delay and dysmorphic craniofacial features. *Eur J Med Genet* **50**, 256–263.
59 Turner DJ, Miretti M, Rajan D, Fiegler H, Carter NP, Blayney ML, Beck S & Hurles ME (2008) Germline rates of de novo meiotic deletions and duplications causing several genomic disorders. *Nat Genet* **40**, 90–95.
60 Strathern JN, Shafer BK & McGill CB (1995) DNA synthesis errors associated with double-strand-break repair. *Genetics* **140**, 965–972.
61 Laan R, Baarends WM, Wassenaar E, Roest HP, Hoeijmakers JH & Grootegoed JA (2005) Expression and possible functions of DNA lesion bypass proteins in spermatogenesis. *Int J Androl* **28**, 1–15.
62 Debrauwere H, Buard J, Tessier J, Aubert D, Vergnaud G & Nicolas A (1999) Meiotic instability of human minisatellite CEB1 in yeast requires DNA double-strand breaks. *Nat Genet* **23**, 367–371.
63 Christin-Maitre S (2008) The role of hormone replacement therapy in the management of premature ovarian failure. *Nat Clin Pract Endocrinol Metab* **4**, 60–61.
64 Corre T, Schuettler J, Bione S, Marozzi A, Persani L, Rossetti R, Torricelli F, Giotti I, Vogt P & Toniolo D (2009) A large-scale association study to assess the impact of known variants of the human INHA gene on premature ovarian failure. *Hum Reprod* **24**, 2023–2028.
65 Mandon-Pepin B, Touraine P, Kuttenn F, Derbois C, Rouxel A, Matsuda F, Nicolas A, Cotinot C & Fellous M (2008) Genetic investigation of four meiotic genes in women with premature ovarian failure. *Eur J Endocrinol* **158**, 107–115.
66 Hikiba J, Hirota K, Kagawa W, Ikawa S, Kinebuchi T, Sakane I, Takizawa Y, Yokoyama S, Mandon-Pepin B, Nicolas A *et al*. (2008) Structural and functional analyses of the DMC1-M200V polymorphism found in the human population. *Nucleic Acids Res* **36**, 4181–4190.
67 Bannister LA, Pezza RJ, Donaldson JR, de Rooij DG, Schimenti KJ, Camerini-Otero RD & Schimenti JC (2007) A dominant, recombination-defective allele of Dmc1 causing male-specific sterility. *PLoS Biol* **5**, e105.
68 Kauppi L, Jeffreys AJ & Keeney S (2004) Where the crossovers are: recombination distributions in mammals. *Nat Rev Genet* **5**, 413–424.
69 Mezard C, Baudat F, Debrauwere H, de Massy B, Smith K, Soustelle C, Varoutas PC, Vedel M & Nicolas A (1999) Mechanisms and control of meiotic recombination in the yeast *Saccharomyces cerevisiae*. *J Soc Biol* **193**, 23–27.
70 Lichten M & Goldman AS (1995) Meiotic recombination hotspots. *Annu Rev Genet* **29**, 423–444.
71 Myers S, Freeman C, Auton A, Donnelly P & McVean G (2008) A common sequence motif associated with recombination hot spots and genome instability in humans. *Nat Genet* **40**, 1124–1129.
72 Myers SR & McCarroll SA (2006) New insights into the biological basis of genomic disorders. *Nat Genet* **38**, 1363–1364.
73 Coop G, Wen X, Ober C, Pritchard JK & Przeworski M (2008) High-resolution mapping of crossovers reveals extensive variation in fine-scale recombination patterns among humans. *Science* **319**, 1395–1398.
74 Paigen K, Szatkiewicz JP, Sawyer K, Leahy N, Parvanov ED, Ng SH, Graber JH, Broman KW & Petkov PM (2008) The recombinational anatomy of a mouse chromosome. *PLoS Genet* **4**, e1000119.

75 Drouaud J, Camilleri C, Bourguignon PY, Canaguier A, Berard A, Vezon D, Giancola S, Brunel D, Colot V, Prum B *et al*. (2006) Variation in crossing-over rates across chromosome 4 of *Arabidopsis thaliana* reveals the presence of meiotic recombination 'hot spots'. *Genome Res* **16**, 106–114.
76 Hyppa RW, Cromie GA & Smith GR (2008) Indistinguishable landscapes of meiotic DNA breaks in rad50+ and rad50S strains of fission yeast revealed by a novel rad50+ recombination intermediate. *PLoS Genet* **4**, e1000267.
77 Baudat F & Nicolas A (1997) Clustering of meiotic double-strand breaks on yeast chromosome III. *Proc Natl Acad Sci USA* **94**, 5213–5218.
78 Blitzblau HG, Bell GW, Rodriguez J, Bell SP & Hochwagen A (2007) Mapping of meiotic single-stranded DNA reveals double-stranded-break hotspots near centromeres and telomeres. *Curr Biol* **17**, 2003–2012.
79 Borde V, Lin W, Novikov E, Petrini JH, Lichten M & Nicolas A (2004) Association of Mre11p with double-strand break sites during yeast meiosis. *Mol Cell* **13**, 389–401.
80 Buhler C, Borde V & Lichten M (2007) Mapping meiotic single-strand DNA reveals a new landscape of DNA double-strand breaks in *Saccharomyces cerevisiae*. *PLoS Biol* **5**, e324.
81 Gerton JL, DeRisi J, Shroff R, Lichten M, Brown PO & Petes TD (2000) Inaugural article: global mapping of meiotic recombination hotspots and coldspots in the yeast *Saccharomyces cerevisiae*. *Proc Natl Acad Sci USA* **97**, 11383–11390.
82 Chen SY, Tsubouchi T, Rockmill B, Sandler JS, Richards DR, Vader G, Hochwagen A, Roeder GS & Fung JC (2008) Global analysis of the meiotic crossover landscape. *Dev Cell* **15**, 401–415.
83 Mancera E, Bourgon R, Brozzi A, Huber W & Steinmetz LM (2008) High-resolution mapping of meiotic crossovers and non-crossovers in yeast. *Nature* **454**, 479–485.
84 Winzeler EA, Richards DR, Conway AR, Goldstein AL, Kalman S, McCullough MJ, McCusker JH, Stevens DA, Wodicka L, Lockhart DJ *et al*. (1998) Direct allelic variation scanning of the yeast genome. *Science* **281**, 1194–1197.
85 Guillon H, Baudat F, Grey C, Liskay RM & de Massy B (2005) Crossover and noncrossover pathways in mouse meiosis. *Mol Cell* **20**, 563–573.
86 Jeffreys AJ & May CA (2004) Intense and highly localized gene conversion activity in human meiotic crossover hot spots. *Nat Genet* **36**, 151–156.
87 Schwacha A & Kleckner N (1997) Interhomolog bias during meiotic recombination: meiotic functions promote a highly differentiated interhomolog-only pathway. *Cell* **90**, 1123–1135.
88 Robine N, Uematsu N, Amiot F, Gidrol X, Barillot E, Nicolas A & Borde V (2007) Genome-wide redistribution of meiotic double-strand breaks in *Saccharomyces cerevisiae*. *Mol Cell Biol* **27**, 1868–1880.
89 Muller HJ (1925) The regionally differential effect of X rays on crossing over in autosomes of Drosophila. *Genetics* **10**, 470–507.
90 de Boer E, Dietrich AJ, Hoog C, Stam P & Heyting C (2007) Meiotic interference among MLH1 foci requires neither an intact axial element structure nor full synapsis. *J Cell Sci* **120**, 731–736.
91 Martini E, Diaz RL, Hunter N & Keeney S (2006) Crossover homeostasis in yeast meiosis. *Cell* **126**, 285–295.
92 Oh SD, Lao JP, Taylor AF, Smith GR & Hunter N (2008) RecQ helicase, Sgs1, and XPF family endonuclease, Mus81–Mms4, resolve aberrant joint molecules during meiotic recombination. *Mol Cell* **31**, 324–336.
93 Maleki S, Neale MJ, Arora C, Henderson KA & Keeney S (2007) Interactions between Mei4, Rec114, and other proteins required for meiotic DNA double-strand break formation in *Saccharomyces cerevisiae*. *Chromosoma* **116**, 471–486.
94 Sasanuma H, Hirota K, Fukuda T, Kakusho N, Kugou K, Kawasaki Y, Shibata T, Masai H & Ohta K (2008) Cdc7-dependent phosphorylation of Mer2 facilitates initiation of yeast meiotic recombination. *Genes Dev* **22**, 398–410.
95 Young JA, Hyppa RW & Smith GR (2004) Conserved and nonconserved proteins for meiotic DNA breakage and repair in yeasts. *Genetics* **167**, 593–605.
96 Abdullah MF & Borts RH (2001) Meiotic recombination frequencies are affected by nutritional states in Saccharomycescerevisiae. *Proc Natl Acad Sci USA* **98**, 14524–14529.
97 Koehn DR, Haring SJ, Williams JM & Malone RE (2009) Tethering recombination initiation proteins in *Saccharomyces cerevisiae* promotes double strand break formation. *Genetics* **182**, 447–458.
98 Pecina A, Smith KN, Mezard C, Murakami H, Ohta K & Nicolas A (2002) Targeted stimulation of meiotic recombination. *Cell* **111**, 173–184.
99 Nicolas A, Treco D, Schultes NP & Szostak JW (1989) An initiation site for meiotic gene conversion in the yeast *Saccharomyces cerevisiae*. *Nature* **338**, 35–39.
100 de Massy B & Nicolas A (1993) The control in cis of the position and the amount of the ARG4 meiotic double-strand break of *Saccharomyces cerevisiae*. *EMBO J* **12**, 1459–1466.
101 Rocco V, de Massy B & Nicolas A (1992) The *Saccharomyces cerevisiae* ARG4 initiator of meiotic gene conversion and its associated double-strand DNA breaks can be inhibited by transcriptional interference. *Proc Natl Acad Sci USA* **89**, 12068–12072.

102 White MA, Dominska M & Petes TD (1993) Transcription factors are required for the meiotic recombination hotspot at the HIS4 locus in *Saccharomyces cerevisiae*. *Proc Natl Acad Sci USA* **90**, 6621–6625.
103 Hirota K, Steiner WW, Shibata T & Ohta K (2007) Multiple modes of chromatin configuration at natural meiotic recombination hot spots in fission yeast. *Eukaryot Cell* **6**, 2072–2080.
104 Schuchert P, Langsford M, Kaslin E & Kohli J (1991) A specific DNA sequence is required for high frequency of recombination in the ade6 gene of fission yeast. *EMBO J* **10**, 2157–2163.
105 Steiner WW, Steiner EM, Girvin AR & Plewik LE (2009) Novel nucleotide sequence motifs that produce hotspots of meiotic recombination in Schizosaccharomyces pombe. *Genetics* **182**, 459–469.
106 Ohta K, Shibata T & Nicolas A (1994) Changes in chromatin structure at recombination initiation sites during yeast meiosis. *EMBO J* **13**, 5754–5763.
107 Wu TC & Lichten M (1994) Meiosis-induced double-strand break sites determined by yeast chromatin structure. *Science* **263**, 515–518.
108 Murakami H, Borde V, Shibata T, Lichten M & Ohta K (2003) Correlation between premeiotic DNA replication and chromatin transition at yeast recombination initiation sites. *Nucleic Acids Res* **31**, 4085–4090.
109 Ohta H, Nicolas A, Furuse M, Nabetani A, Ogawa H & Shibata T (1998) Mutations in the MRE11, RAD50, XRS2, and MRE2 genes alter chromatin configuration at meiotic double-strand break sites in premeiotic and meiotic cells. *Proc Natl Acad Sci USA* **95**, 646–651.
110 Kirkpatrick DT, Wang YH, Dominska M, Griffith JD & Petes TD (1999) Control of meiotic recombination and gene expression in yeast by a simple repetitive DNA sequence that excludes nucleosomes. *Mol Cell Biol* **19**, 7661–7671.
111 Kouzarides T (2007) Chromatin modifications and their function. *Cell* **128**, 693–705.
111a Gaucher J, Reynoird N, Montellier E, Boussouar F, Rousseaux S & Khochbin S (2009) From meiosis to postmeiotic events: The secrets of histone disappearance. *FEBS J* **277**, doi:10.1111/j.1742-4658.2009.07504.x
112 Merker JD, Dominska M, Greenwell PW, Rinella E, Bouck DC, Shibata Y, Strahl BD, Mieczkowski P & Petes TD (2008) The histone methylase Set2p and the histone deacetylase Rpd3p repress meiotic recombination at the HIS4 meiotic recombination hotspot in *Saccharomyces cerevisiae*. *DNA Repair (Amst)* **7**, 1298–1308.
113 Carrozza MJ, Li B, Florens L, Suganuma T, Swanson SK, Lee KK, Shia WJ, Anderson S, Yates J, Washburn MP *et al*. (2005) Histone H3 methylation by Set2 directs deacetylation of coding regions by Rpd3S to suppress spurious intragenic transcription. *Cell* **123**, 581–592.
114 Mieczkowski PA, Dominska M, Buck MJ, Lieb JD & Petes TD (2007) Loss of a histone deacetylase dramatically alters the genomic distribution of Spo11p-catalyzed DNA breaks in *Saccharomyces cerevisiae*. *Proc Natl Acad Sci USA* **104**, 3955–3960.
115 Yamashita K, Shinohara M & Shinohara A (2004) Rad6–Bre1-mediated histone H2B ubiquitylation modulates the formation of double-strand breaks during meiosis. *Proc Natl Acad Sci USA* **101**, 11380–11385.
116 Borde V, Robine N, Lin W, Bonfils S, Geli V & Nicolas A (2009) Histone H3 lysine 4 trimethylation marks meiotic recombination initiation sites. *EMBO J* **28**, 99–111.
117 Sollier J, Lin W, Soustelle C, Suhre K, Nicolas A, Geli V & de La Roche Saint-Andre C (2004) Set1 is required for meiotic S-phase onset, double-strand break formation and middle gene expression. *EMBO J* **23**, 1957–1967.
118 Kniewel R & Keeney S (2009) Histone methylation sets the stage for meiotic DNA breaks. *EMBO J* **28**, 81–83.
119 Buard J, Barthes P, Grey C & de Massy B (2009) Distinct histone modifications define initiation and repair of meiotic recombination in the mouse. *EMBO J* **28**, 2616–2624.
120 Hayashi K, Yoshida K & Matsui Y (2005) A histone H3 methyltransferase controls epigenetic events required for meiotic prophase. *Nature* **438**, 374–378.
121 Ptak SE, Hinds DA, Koehler K, Nickel B, Patil N, Ballinger DG, Przeworski M, Frazer KA & Paabo S (2005) Fine-scale recombination patterns differ between chimpanzees and humans. *Nat Genet* **37**, 429–434.
122 Ahn SH, Henderson KA, Keeney S & Allis CD (2005) H2B (Ser10) phosphorylation is induced during apoptosis and meiosis in S. cerevisiae. *Cell Cycle* **4**, 780–783.
123 Burgess SM, Ajimura M & Kleckner N (1999) GCN5-dependent histone H3 acetylation and RPD3-dependent histone H4 deacetylation have distinct, opposing effects on IME2 transcription, during meiosis and during vegetative growth, in budding yeast. *Proc Natl Acad Sci USA* **96**, 6835–6840.
124 Krishnamoorthy T, Chen X, Govin J, Cheung WL, Dorsey J, Schindler K, Winter E, Allis CD, Guacci V, Khochbin S *et al*. (2006) Phosphorylation of histone H4 Ser1 regulates sporulation in yeast and is conserved in fly and mouse spermatogenesis. *Genes Dev* **20**, 2580–2592.
125 Mahadevaiah SK, Turner JM, Baudat F, Rogakou EP, de Boer P, Blanco-Rodriguez J, Jasin M, Keeney S, Bonner WM & Burgoyne PS (2001) Recombinational DNA double-strand breaks in mice precede synapsis. *Nat Genet* **27**, 271–276.
126 Recht J, Tsubota T, Tanny JC, Diaz RL, Berger JM, Zhang X, Garcia BA, Shabanowitz J, Burlingame AL,

Hunt DF *et al*. (2006) Histone chaperone Asf1 is required for histone H3 lysine 56 acetylation, a modification associated with S phase in mitosis and meiosis. *Proc Natl Acad Sci USA* **103**, 6988–6993.

127 Reddy KC & Villeneuve AM (2004) C. elegans HIM-17 links chromatin modification and competence for initiation of meiotic recombination. *Cell* **118**, 439–452.

128 Reynolds P, Koken MH, Hoeijmakers JH, Prakash S & Prakash L (1990) The rhp6+ gene of Schizosaccharomyces pombe: a structural and functional homolog of the RAD6 gene from the distantly related yeast *Saccharomyces cerevisiae*. *EMBO J* **9**, 1423–1430.

129 Roest HP, van Klaveren J, de Wit J, van Gurp CG, Koken MH, Vermey M, van Roijen JH, Hoogerbrugge JW, Vreeburg JT, Baarends WM *et al*. (1996) Inactivation of the HR6B ubiquitin-conjugating DNA repair enzyme in mice causes male sterility associated with chromatin modification. *Cell* **86**, 799–810.

130 Yamada T, Mizuno K, Hirota K, Kon N, Wahls WP, Hartsuiker E, Murofushi H, Shibata T & Ohta K (2004) Roles of histone acetylation and chromatin remodeling factor in a meiotic recombination hotspot. *EMBO J* **23**, 1792–1803.

Downloaded from genome.cshlp.org on October 26, 2021 - Published by Cold Spring Harbor Laboratory Press

RNA-DNA hybrid (R-loop) immunoprecipitation mapping: an analytical workflow to evaluate inherent biases

László Halász,[1,2,6] Zsolt Karányi,[1,3,6] Beáta Boros-Oláh,[1,2] Tímea Kuik-Rózsa,[1,2] Éva Sipos,[1,4] Éva Nagy,[1] Ágnes Mosolygó-L,[1,2] Anett Mázló,[5] Éva Rajnavölgyi,[5] Gábor Halmos,[4] and Lóránt Székvölgyi[1,2]

[1]*MTA-DE Momentum, Genome Architecture and Recombination Research Group, Research Centre for Molecular Medicine, University of Debrecen, 4032 Debrecen, Hungary;* [2]*Department of Biochemistry and Molecular Biology, University of Debrecen, 4032 Debrecen, Hungary;* [3]*Department of Internal Medicine, University of Debrecen, 4032 Debrecen, Hungary;* [4]*Department of Biopharmacy, University of Debrecen, 4032 Debrecen, Hungary;* [5]*Department of Immunology, University of Debrecen, 4032 Debrecen, Hungary*

The impact of R-loops on the physiology and pathology of chromosomes has been demonstrated extensively by chromatin biology research. The progress in this field has been driven by technological advancement of R-loop mapping methods that largely relied on a single approach, DNA-RNA immunoprecipitation (DRIP). Most of the DRIP protocols use the experimental design that was developed by a few laboratories, without paying attention to the potential caveats that might affect the outcome of RNA-DNA hybrid mapping. To assess the accuracy and utility of this technology, we pursued an analytical approach to estimate inherent biases and errors in the DRIP protocol. By performing DRIP-sequencing, qPCR, and receiver operator characteristic (ROC) analysis, we tested the effect of formaldehyde fixation, cell lysis temperature, mode of genome fragmentation, and removal of free RNA on the efficacy of RNA-DNA hybrid detection and implemented workflows that were able to distinguish complex and weak DRIP signals in a noisy background with high confidence. We also show that some of the workflows perform poorly and generate random answers. Furthermore, we found that the most commonly used genome fragmentation method (restriction enzyme digestion) led to the overrepresentation of lengthy DRIP fragments over coding ORFs, and this bias was enhanced at the first exons. Biased genome sampling severely compromised mapping resolution and prevented the assignment of precise biological function to a significant fraction of R-loops. The revised workflow presented herein is established and optimized using objective ROC analyses and provides reproducible and highly specific RNA-DNA hybrid detection.

[Supplemental material is available for this article.]

R-loops are three-stranded nucleic acid structures that are composed of an RNA-DNA hybrid and a displaced single-stranded DNA. Under physiological conditions, R-loops are prevalent along the chromosomes, constituting 5%–8% of the genome and impacting many cellular processes (Sanz et al. 2016; Wahba et al. 2016). For instance, R-loops (1) drive embryonic stem cell differentiation via modulating the chromosomal binding of chromatin-regulatory complexes (Chen et al. 2015), (2) ensure the optimal binding of transcriptional activators to the promoter of the human vimentin (*VIM*) gene (Boque-Sastre et al. 2015), (3) massively form on estrogen-responsive genes in human breast and other tissues upon estrogen-hormone stimulation (Stork et al. 2016), (4) induce heterochromatin formation in *Schizosaccharomyces pombe* (Nakama et al. 2012), and (5) inhibit the expression of an antisense noncoding RNA in *Arabidopsis thaliana*, associated with the flowering process (Sun et al. 2013). In a pathological context, perturbation or mutation of any of the following factors causes the chromosomal accumulation of RNA-DNA hybrids and consequent genomic instability: (1) mRNA splicing factors and RNA export factors (e.g., THO2, HPR1, MFT1, THP2, THOC1-7, SRSF1) (Huertas and Aguilera 2003; Li and Manley 2005; Domínguez-Sánchez et al. 2011; Gan et al. 2011); (2) RNA-DNA hybrid helicases (e.g., SETX/SEN1, AQR, PIF1) (Boulé and Zakian 2007; Mischo et al. 2011; Alzu et al. 2012; Sollier et al. 2014); (3) RNA-DNA ribonucleases (RNASEH1/RNH1, RNASEH2A-C/RNH201) (El Hage et al. 2010; Chon et al. 2013; Stuckey et al. 2015); (4) homologous recombination proteins (e.g., BRCA1, BRCA2, RTEL1, SRS2), (Bhatia et al. 2014; Hatchi et al. 2015); (5) Fanconi anemia proteins (FANCA, FANCB, FANCC) (García-Rubio et al. 2015; Schwab et al. 2015); and (6) topoisomerases (TOP1, TOP3B) (Wilson-Sali and Hsieh 2002; El Hage et al. 2010; Yang et al. 2014; Marinello et al. 2016).

The above examples clearly illustrate the massive progress in the field that has been driven by technological advancements of R-loop detection methods. These techniques involve, for instance, electrophoretic mobility shift assays (Yu et al. 2006), atomic force

[6]These authors are joint first authors and contributed equally to this work.
Corresponding author: lorantsz@med.unideb.hu
Article published online before print. Article, supplemental material, and publication date are at http://www.genome.org/cgi/doi/10.1101/gr.219394.116.

© 2017 Halász et al. This article is distributed exclusively by Cold Spring Harbor Laboratory Press for the first six months after the full-issue publication date (see http://genome.cshlp.org/site/misc/terms.xhtml). After six months, it is available under a Creative Commons License (Attribution-NonCommercial 4.0 International), as described at http://creativecommons.org/licenses/by-nc/4.0/.

Downloaded from genome.cshlp.org on October 26, 2021 - Published by Cold Spring Harbor Laboratory Press

microscopy (Brown et al. 2008), transmission electron microscopy (Pohjoismäki et al. 2010), fluorescent microscopy (Székvölgyi et al. 2007), fluorescence in situ hybridization (Nadel et al. 2015), native bisulfite modification (Yu et al. 2003), immunoprecipitation (Skourti-Stathaki et al. 2011; Ginno et al. 2012), and computational prediction (Jenjaroenpun et al. 2015). The increasing numbers of R-loop mapping data relied on a single approach, DNA-RNA immunoprecipitation (DRIP) and its variations (RDIP, DRIPc, S1-DRIP, DRIP-RNA, DIP, ChIP). The DRIP method applies the S9.6 anti-RNA-DNA hybrid antibody (Hu et al. 2006) to capture RNA-DNA hybrids in their native chromosomal context, followed by mapping the enriched DNA fragments on a selected number of loci or across the whole genome, using quantitative PCR, microarray hybridization, or deep sequencing.

Having surveyed the published RNA-DNA hybrid mapping studies (Supplemental Table S1; El Hage et al. 2010, 2014; Mischo et al. 2011; Skourti-Stathaki et al. 2011; Alzu et al. 2012; Ginno et al. 2012; Castellano-Pozo et al. 2013; Sun et al. 2013; Wahba and Koshland 2013; Bhatia et al. 2014; Chan et al. 2014; Groh et al. 2014; Herrera-Moyano et al. 2014; Loomis et al. 2014; Rigby et al. 2014; Salvi et al. 2014; Yang et al. 2014, 2016; Zhang et al. 2014a, 2014b, 2015; Boque-Sastre et al. 2015; Chen et al. 2015; García-Rubio et al. 2015; Hatchi et al. 2015; Jenjaroenpun et al. 2015; Lim et al. 2015; Nadel et al. 2015; Pefanis et al. 2015; Cloutier et al. 2016; Marinello et al. 2016; Ohle et al. 2016; Romanello et al. 2016; Sanz et al. 2016; Stork et al. 2016; Wahba et al. 2016; Zeller et al. 2016), we found that most DRIP protocols used the experimental design that was developed by a few laboratories (Supplemental Fig. S1). The original protocols are still being used without paying attention to their potential caveats: several critical points have remained exceedingly heterogeneous among the DRIP studies (Supplemental Table S1) that might account for at least some of the contradictory results (Ginno et al. 2012; Chan et al. 2014; El Hage et al. 2014; Nadel et al. 2015; Wahba et al. 2016). One can reveal technical heterogeneities (1) in terms of the studied model organisms and cell types, (2) in whether the cells were fixed by formaldehyde (HCHO) or not, (3) in whether the immunoprecipitation was chromatin-based or DNA-based (ChIP vs. DIP), (4) in the cell lysis temperature (65°C, 55°C, 37°C), (5) in the mode of DNA fragmentation (restriction enzyme digestion vs. sonication), (6) in total nucleic acid extraction (solid-phase purification vs. organic extraction, or salting out extraction), and (7) in the application of ribonuclease A digestion to eliminate free RNA from the nucleic acid prep. Obviously, each of these variables can introduce substantial bias that might obscure the overall outcome of the experiment, but their consequence, alone or in combination, has remained unexplored.

In the current study, we aimed to assess possible confounding effects related to key experimental variables of the DRIP procedure. Combining DRIP-qPCR, DRIP-sequencing, and receiver operator characteristic (ROC) calculation, we devised an unbiased and systematic analytical pipeline in human T lymphoblastoid cells regarding the most important DRIP variables so that we suggest a reproducible and specific RNA-DNA hybrid detection, underlain by the objective criteria of ROC analysis.

Results

Introducing DRIP classifiers to assess true and false R-loop associations

Based on the available workflows of published DRIP protocols and considering the main technical variables that might contribute to the observed heterogeneities, we designed forty DRIP experimental schemes (binary classifiers) so that we assess how they rank different test loci according to their known RNA-DNA hybrid status (Fig. 1). The classifiers ("DRIP experiments" or "dependent variables") were designed to systematically explore the main factors that might create experimental bias associated with the DRIP procedure.

Experiments 1–16 consider the effect of (1) formaldehyde (HCHO) fixation, (2) the method of nucleic acid isolation, (3) removal of free RNA, (4) the mode of nucleic acid fragmentation (Fig. 1A), and (5) cell lysis temperature (65°C as default vs. 37°C) (not shown in Fig. 1A, but referred to throughout the text as "37°C").

Step 1: Formaldehyde fixation

The basic assumption behind HCHO–cross-linking is to maximize the DRIP yield while preserving biologically meaningful RNA-DNA hybrid interactions. However, formaldehyde has some well-known adverse effects: (1) the DNA accompanies a conformational change upon cross-linking, involving local denaturation or "breathing" of the double helix (McGhee and von Hippel 1977). This might create ectopic R-loop sites or abolish physiological R-loop contacts. (2) HCHO-treatment can reduce antigen accessibility or mask epitopes recognized by the antibody used for the immunoprecipitation. This might prevent a fraction of R-loops from being detected. (3) HCHO-fixation elicits spurious localization of irrelevant proteins at highly expressed genes (Baranello et al. 2016) and induces massive poly(ADP)ribose polymer formation in live cells (Beneke et al. 2012). These examples warrant deeper investigation of the usage of HCHO-fixation in RNA-DNA hybrid mapping; therefore, we classified our DRIP samples as HCHO-treated and nontreated categories (Fig. 1A,B).

Step 2: Nucleic acid purification

Two common methods were compared: organic (phenol/chloroform) extraction versus solid-phase (silica membrane) purification of total nucleic acids (Fig. 1A,B).

Step 3: Ribonucleolytic treatment (RNase A, RNase H, and sodium hydroxide)

Most DRIP protocols do not treat the isolated nucleic acid with ribonucleases to remove free RNA; however, the S9.6 antibody can recognize RNA duplexes with an approximately fivefold reduced affinity compared to RNA-DNA hybrids (Phillips et al. 2013). At this point, four kinds of ribonucleoleolytic digestion were incorporated into our DRIP pipelines: (1) RNase H1 digestion that removes RNA-DNA hybrids (negative control #1); (2) alkaline hydrolysis by sodium hydroxide that degrades free RNA and RNA-DNA hybrids (negative control #2); (3) RNase A digestion at high (300 mM) NaCl concentration that removes free RNA; and (4) RNase A digestion at low (25 mM) NaCl concentration that removes free RNA and RNA-DNA hybrids.

RNase H1 treatment is an accepted negative control of the DRIP procedure since it degrades the RNA strand in the hybrids, preventing their recognition by the S9.6 antibody. Half of the nucleic acid prep was digested by RNase H1 before the DNA fragmentation step that let us estimate the bulk level of RNA-DNA hybrids (dot blot setting; Supplemental Fig. S2A). The other half was digested just before the S9.6 immunoprecipitation step that let us obtain crucial information about the specificity of the IP signal (see

Downloaded from genome.cshlp.org on October 26, 2021 - Published by Cold Spring Harbor Laboratory Press

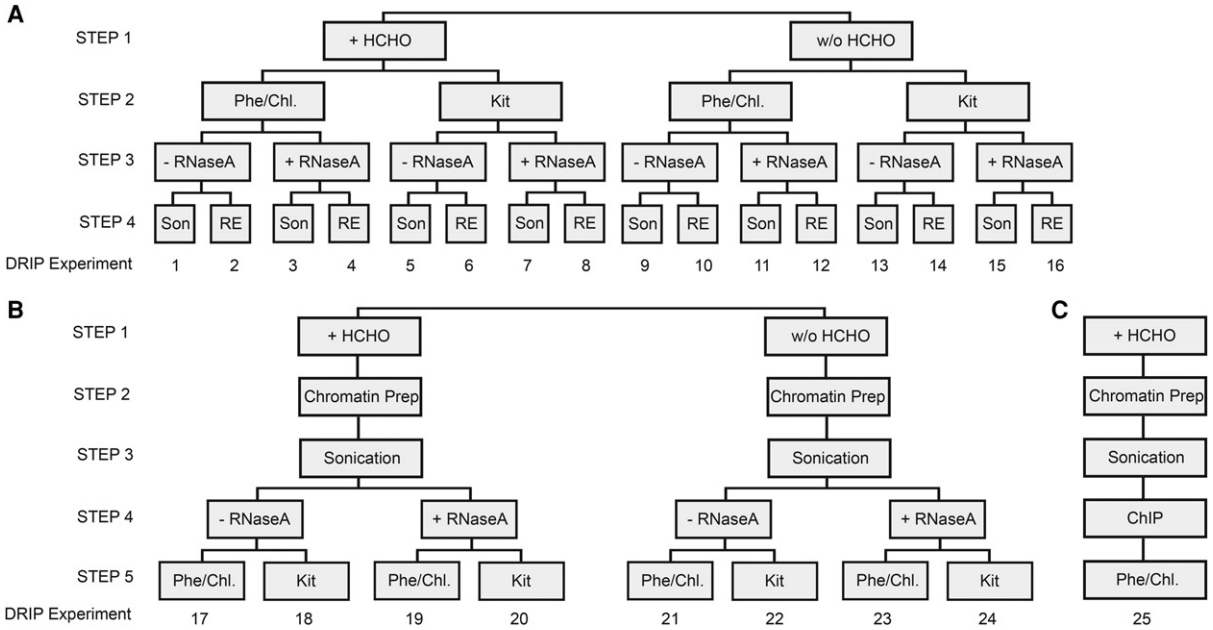


Figure 1. Experimental design: constructing DRIP schemes. (*A*) Experiments 1–16 explore the effect of formaldehyde-fixation (Step 1), nucleic acid isolation (Step 2), removal of free RNA (Step 3), and nucleic acid fragmentation (Step 4) on the outcome of RNA-DNA hybrid detection. Each experiment was performed at two parallel cell lysis temperatures (65°C and 37°C), respectively. The temperature variable is not depicted in the cartoon, but it is referred in the main text. (*B*) Experiments 17–24 test the impact of acoustic sharing performed on a chromatin prep rather than on naked nucleic acid, similarly to the ChIP protocol. Each experiment was performed at 65°C cell lysis temperature. (*C*) Workflow of a ChIP experiment (shown only for comparison with the DRIP pipeline). (HCHO) Formaldehyde fixation, (Phe/Chl) phenol-chloroform extraction, (Kit) silica membrane-based nucleic acid purification, (RNase A) Ribonuclease A digestion performed at high (300 mM) NaCl concentration, (Son) sonication, (RE) restriction enzyme cocktail digestion (HindIII, EcoRI, BsrGI, XbaI, and SspI). As a negative control, RNase H digestion was applied in all DRIP experiments (not indicated in the cartoon).

DRIP-qPCR). As expected, RNA-DNA hybrids were sensitive to RNase H1 digestion in vitro. Similarly to RNase H1, alkaline hydrolysis by 50 mM NaOH also efficiently eliminated the RNA-DNA hybrid signal (Supplemental Fig. S2A). Less is known about the salt-dependent RNase H-like activity of RNase A that is supposed to digest RNA-DNA hybrids as an efficient hybridase at low ionic strength (https://www.thermofisher.com/order/catalog/product/EN0531). As shown in Supplemental Figure S2B, the hybrids were indeed resistant to RNase A digestion at high ionic strength, but they became highly sensitive to RNase A as a function of decreasing monovalent concentration. The RNase H-like activity of RNase A at low salt condition was confirmed by an independent method (Supplemental Fig. S2C,D) applying fluorescent microscopic detection. Based on these experiences, RNase A digestion at high salt concentration (300 mM NaCl) was integrated into our DRIP protocol to test if removal of competing free RNA improves the specificity of the RNA-DNA hybrid signal. Also, RNase H1 digestion of the fragmented nucleic acid was kept as an obligatory negative control of the immunoprecipitation.

Step 4: Nucleic acid fragmentation

The choice of restriction enzymes defines the cleavage pattern of DNA that is critical to achieve optimal fragment length distribution and mapping resolution. Based on the original DRIP protocol (Ginno et al. 2012), we combined five enzymes (HindIII, EcoRI, BsrGI, XbaI, and SspI) for in silico digestion, resulting in a median restriction fragment length of 314 bp (Supplemental Fig. S3A). In contrast to the theoretical fragment size distribution, we observed a broad DNA size range in a real digestion reaction (between 100–10,000 bp) (Supplemental Fig. S4A). As a control, we repeated the restriction enzyme cleavage in varying reaction conditions without detecting any improvement in the digestion efficacy (Supplemental Fig. S3B). When a budding yeast genomic DNA was digested in a parallel experiment, we managed to obtain the expected (in silico) fragment size distribution (Supplemental Fig. S3C). These observations necessitate the proper control of DNA fragment length distribution in DRIP samples that derive from restriction enzyme-fragmented nucleic acid.

As opposed to restriction enzyme digestion, sonication creates random DNA fragments with a typical size of 150–500 bp that dictate the spatial resolution of the DRIP assay (Supplemental Fig. S4B). However, excessive sonication can introduce strand breaks in the DNA or simply shake off a subset of R-loops from the chromosomes, potentially compromising their detectability by qPCR. Because of the above, the mode of DNA fragmentation (restriction enzymes and sonication) was introduced as an important parameter in our DRIP pipeline (Fig. 1A).

Fragmenting chromatin rather than purified genomic DNA (experiments 17–24)

In comparison to the original DRIP protocol, classical chromatin immunoprecipitation (ChIP) involves the capture of RNA-DNA hybrids by immunoprecipitation from cross-linked and sonicated chromatin (rather than naked DNA) followed by phenol/chloroform purification (Fig. 1C). Since sonication, performed on purified genomic DNA, led to loss of ~80% of the DRIP signal in yeast (Wahba et al. 2016), we tested if acoustic shearing performed on a chromatin prep rather than on naked nucleic acid (Supplemental Fig. S4C) could improve the signal-to-noise ratio of the DRIP measurement (Fig. 1B).

Downloaded from genome.cshlp.org on October 26, 2021 - Published by Cold Spring Harbor Laboratory Press

Varying the cell lysis temperature

Published DRIP protocols apply various cell lysis temperatures, ranging from 37°C to 65°C and lasting from a couple of hours to overnight. To test the effect of temperature on the specificity of RNA-DNA hybrid detection, we lysed the samples at 65°C for 7 h or at 37°C overnight. Experiments 1–16 were processed in parallel at both temperatures, while exp. 17–24 were omitted from the temperature analysis since cross-link reversal typically occurs at 65°C.

Taken together, the above experimental variables resulted in 40 ($16 \times 2 + 8$) autonomous DRIP classifiers (schemes) for which RNA-DNA hybrid enrichment scores were determined at several test loci. This allowed us to assess whether the S9.6 signal represented true or false R-loop associations within the applied condition.

Making a reference R-loop set for benchmarking the DRIP classifiers

To derive the parameters of the DRIP classifiers, known positive and negative examples (genomic sites) could be chosen from the scientific literature based on their known R-loop profiles; however, the heterogeneity of the available DRIP-qPCR and DRIP-seq data sets (see Introduction) prompted us to establish our independent R-loop training set. We performed DNA-RNA hybrid mapping (DRIP-seq) in two closely related human cell types (Jurkat T cell leukemia cell line and naive $CD4^+$ T lymphocytes) and identified 88,830 and 99,337 R-loop enriched regions, respectively (Fig. 2A). A high-confidence R-loop peak set was generated from the identified binding sites, and their chromosomal distribution was characterized. The peaks were significantly enriched at gene promoters and repetitive elements (Fig. 2B), consistent with previously published DRIP-seq results (Ginno et al. 2012; Nadel et al. 2015). R-loop sites were underrepresented at protein coding exons, similarly to earlier DRIP experiments performed with sonicated nucleic acid; however, restriction enzyme-fragmented DRIP samples were positively biased toward exons. Sonicated and restriction enzyme-digested samples were strikingly different in their R-loop length distributions (narrow: 179–2369 bp vs. wide: 178–22,479 bp) (Fig. 2C), and the identified R-loop binding sites significantly overlapped within each group but sharply stood apart between the two groups (Fig. 2D). We attribute these differences to the extensive variation of R-loop lengths and heterogeneities of the studied cell types. Biological implications of having too wide peak sizes will be discussed later. With the observed variances in mind, our consensus R-loop set was regarded as an amenable reference to benchmark the DRIP classifiers.

Measuring RNA-DNA hybrid enrichment over the DRIP classifiers

Positive and negative test regions were selected from the identified R-loop set (Supplemental Fig. S5) and were systematically probed for RNA-DNA hybrid enrichment across the DRIP classifiers (Supplemental Fig. S6). Five test regions were frequently used as positive and negative controls in various published DRIP studies (*SNRPN, ZNF554, MYADM, FMR1, APOE*) (Ginno et al. 2012; Bhatia et al. 2014; Groh et al. 2014; Herrera-Moyano et al. 2014; Loomis et al. 2014; Yang et al. 2014; Boque-Sastre et al. 2015; García-Rubio et al. 2015; Marinello et al. 2016), while the remaining sites were picked at random from the consensus R-loop set (*PRR5L, LOC440704, NOP58, VIM, ING3*). The reference DRIP-seq signal (benchmarking the classifiers) is shown over selected test regions along with DRIP-seq patterns taken from published studies (Supplemental Fig. S5). DRIP-qPCR yields were measured in control and RNase H-treated samples for 40 ($16 \times 2 + 8$) DRIP classifiers, at 10 test regions, in five independent experiments. The resulting 4000 ($40 \times 2 \times 10 \times 5$) DRIP enrichment scores were then readily used as an input parameter of receiver operator characteristics calculation.

Determining the sensitivity and specificity of RNA-DNA hybrid detection: ROC analysis

We quantitated the relative trade-offs between true positive hits and experimental errors (false R-loop associations) by performing ROC analysis (Robin et al. 2011) on the DRIP-qPCR screen characterizing the classifiers (Supplemental Figs. S6–S10). The sensitivity, specificity, and the area under the curve (AUC) values were extracted from the ROC plots (Supplemental Table S2) and used as an objective measure of the robustness of the 40 experiments. High (>0.7) AUC values were obtained for 10 DRIP classifiers (exp. 5, 6, 13, 15, 17, 18, 19, 21, and 24), implying that those experiments could predict the presence or absence of an RNA-DNA hybrid with high efficacy (Fig. 3A). AUC values close to 0.5 were obtained in four experiments (exp. 2, 10, 11, and 16), implying that the classifiers gave random answers without any predictive power as to the presence of an R-loop. Based on these considerations, the top four DRIP classifiers were: exp. 5, 13, 17, and 19 (Fig. 3B,C), with a sensitivity of 68.5%–75% and specificity of 68%–79%. Similar (or even higher) ROC parameters were obtained in a repeated experiment using a B lymphoblastoid cell line (Supplemental Fig. S8), demonstrating the reliability of the tested DRIP protocols in other cell types.

Pairwise comparison of the main experimental variables (Fig. 3D) revealed no significant difference between (1) formaldehyde-fixed vs. unfixed samples, (2) phenol-chloroform extracted vs. silica membrane-purified nucleic acid samples, and (3) DNA-fragmented (exp. 1–16) vs. chromatin-fragmented DRIP samples (exp. 17–24). Cell lysis temperature (65°C vs. 37°C) did not change the specificity and sensitivity of the DRIP assay (Supplemental Figs. S9, S10). A statistically significant difference was obtained for RNase A-treated vs. untreated samples ($P = 0.03$), suggesting that addition of RNase A does not improve the efficacy of RNA-DNA hybrid detection (Step 3, Fig. 3D). We explain the adverse effect of RNase A by its reported DNA binding activity (Benore-Parsons and Ayoub 1997; Dona and Houseley 2014) that selectively eliminates a vast amount (micrograms) of melted DNA regions upon nucleic acid purification (Dona and Houseley 2014). We confirmed the strong DNA binding of RNase A as migration defects on DNA gels, when a plasmid DNA was incubated with the enzyme (Supplemental Fig. S11). The observed electrophoretic mobility shift was prevalent on supercoiled, nicked-circular, and linearized DNA templates.

Finally, by comparing sonicated and restriction enzyme fragmented DRIP samples (Step 4, Fig. 3D), we found a statistically significant difference ($P = 0.0002$) in the ROC parameters, suggesting that sonication is more efficient in discriminating true positive signals from false positives, at least within the tested conditions.

Good DRIP practice: impact on the annotation and basic biological function of R-loops

Suboptimal DRIP conditions might prevent the assignment of precise biological function to a significant fraction of R-loops.

Downloaded from genome.cshlp.org on October 26, 2021 - Published by Cold Spring Harbor Laboratory Press

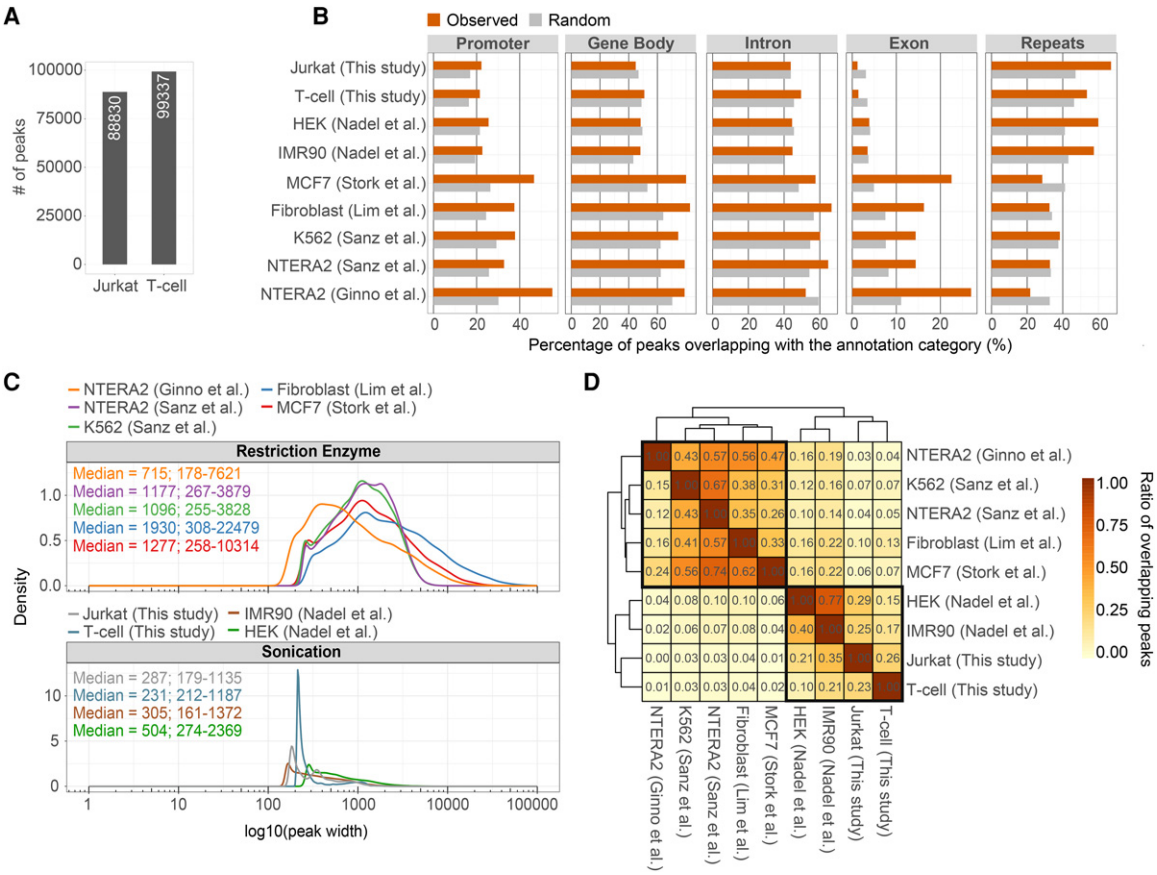


Figure 2. Summary of available human DRIP-seq experiments. (*A*) Bar chart showing the number of identified R-loop peaks in human Jurkat cells and naive T cells (this study). (*B*) Annotation of R-loop binding sites over functional genomic elements. DRIP-seq peaks were determined in Jurkat cells and naive T cells, and in other published cell types (NTERA2, K562, Fibroblast, MCF7, IMR90, HEK293T). The upper four rows represent DRIP experiments fragmenting the nucleic acid by sonication, while the lower five rows highlight restriction enzyme-digested DRIP samples. The difference between the two groups is especially noticeable over exons (associated to 14%–27% and 1%–3.5% of R-loops, respectively) and repeat elements (SINEs, LINEs, LTRs, simple and low complexity repeats) that involve 22%–38% and 54%–67% of the R-loop peaks, respectively. At other annotation categories (gene body, introns, and promoters), the difference was not significant between the two groups. (*C*) Density plots showing the distribution of R-loop peak sizes, classified by fragmentation method (restriction enzyme vs. sonication). Median peak length and 2.5%–97.5% quantiles are indicated. Peak length distributions differ significantly between the two fragmentation methods. (*D*) Heat map showing the overlap of R-loop binding sites between independent DRIP-seq experiments. Values and cell colors represent pairwise and unique overlap ratios between each peak set. The difference between the two nucleic acid fragmentation methods is clearly apparent, as peak sets from the same fragmentation process better resemble each other (highlighted in black).

Although the average DNA fragment size resulting from restriction enzyme digestion fits the requirements of the DRIP assay, we found that the frequency of cutting sites was significantly higher within intergenic regions, producing lengthy restriction fragments over protein coding ORFs (Fig. 4). Biased genome sampling, related to the nonrandom distribution of restriction enzyme recognition sequences, was even more pronounced over exons (Fig. 4C), especially over the first exons (Fig. 4D). In 82% of first exons, there were only 0–1 suitable restriction sites, compared to intergenic regions (59%). We estimated the digestion efficiency of restriction enzyme cutting sites as ~50% over intergenic regions (based on the proportion of zero reads over restriction enzyme cutting sequences, representing cleaved sites), which was significantly reduced over gene coding regions (Fig. 4E,F). Consequently, genic regions void of suitable restriction sites appear as long DRIP fragments that potentially compromise mapping resolution. The *MYC*, *BCL6*, and *VIM* genes are shown as representative examples for large, restriction fragment-sized DRIP peaks (Fig. 5). Precise genomic position of R-loops could be resolved by sonication.

Discussion

The increasing recognition of RNA-DNA hybrid structures in the physiology and pathology of chromosomes has prompted us to develop an analytical approach to estimate the inherent biases and errors of existing DRIP protocols and to assess the power of the technology. The determined ROC parameters (AUC, sensitivity, specificity, threshold) served as an objective measure for the efficacy of predicting the presence or absence of RNA-DNA hybrids. In the tested experimental conditions, we managed to find and verify DRIP workflows that were able to distinguish complex or weak DRIP-qPCR signals from a noisy background with high confidence across a number of genomic regions (exp. 5, 13, 17, and 19). On the contrary, some DRIP workflows performed unreliably and generated random answers (exp. 2, 10, 11, and 16). Under our experimental conditions, we highlight these groups as "preferred" and "not preferred." By testing the main parameters of the DRIP experimental scheme—involving formaldehyde fixation, cell lysis temperature, nucleic acid isolation, free RNA removal, and DNA

Downloaded from genome.cshlp.org on October 26, 2021 - Published by Cold Spring Harbor Laboratory Press

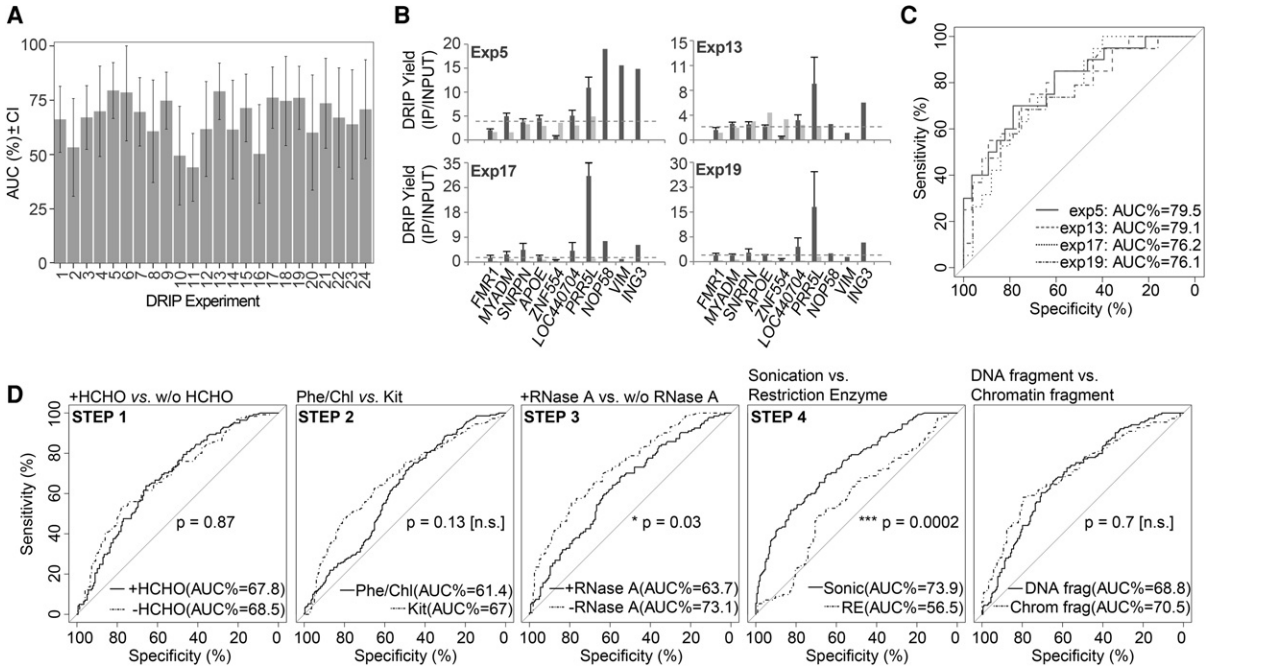


Figure 3. Good DRIP practice. (*A*) Bar charts showing the distribution of AUC (area under the curve) values of ROC plots for 24 DRIP classifiers. Error bars represent the confidence interval of AUCs. High (>0.7) AUC values were obtained for 10 DRIP classifiers (exp. 5, 6, 13, 15, 17, 18, 19, 21, and 24). Low (~0.5) AUC values were obtained in four DRIP experiments (exp. 2, 10, 11, and 16). We highlight these groups as "preferred" and "not preferred," respectively. (*B*,*C*) The top four DRIP experiments ranked by AUCs (exp. 5, 13, 17, and 19). (*B*) DRIP-qPCR enrichment scores are displayed over the test regions. Horizontal dotted lines represent the cutoff value (calculated from the ROC curves) separating the true R-loop signal from background. (*C*) ROC curves of the top four experiments. (*D*) Paired-ROC plots, comparing the main variables (steps) of the DRIP experiments. The level of statistical significance was 0.05.

fragmentation—we found that fragmenting the nucleic acid by sonication and omitting RNase A digestion could improve the precision and specificity of RNA-DNA hybrid detection (Fig. 3D). At this point, we emphasize the lack of correlation between the DRIP scores (IP/input ratios) and AUC values, as these quantities are not related to each other. The former highlights the yield of immunoprecipitation, while the latter is a quantitative measure of true and false R-loop associations. For instance, the worst and best DRIP schemes (exp. 2 and exp. 5) had a qPCR yield of 10%–95% and 1%–18% over the studied regions, respectively (Supplemental Figs. S6, S8). Consequently, high DRIP enrichment is not necessarily accompanied by increased accuracy, and vice versa.

We also showed that genome fragmentation by restriction enzymes led to the overrepresentation of long DRIP fragments over ORFs, which was especially enhanced over the first exons of protein coding genes (Figs. 4–5). Biased genome sampling severely compromised mapping resolution and, as a consequence, the assignment of clear biological function to a fraction of R-loops. For instance, correct estimation of evolutionary conservation between R-loop binding sites, relying on sequence homologies of exons that are associated with R-loops (Sanz et al. 2016), becomes uncertain.

Based on the above experiences, we suggest the following refinements of DRIP workflows to obtain accurate estimates of RNA-DNA hybrid occupancies: (1) Omission of HCHO-fixation and RNase A treatment, isolation of nucleic acid by silica membrane (kit) purification, nucleic acid fragmentation by sonication, followed by immunoprecipitation with the S9.6 antibody (see Methods). (2) If formaldehyde-fixation is applied, we recommend preparing soluble chromatin and fragmenting the prep by sonication (similarly to the ChIP protocol), followed by organic extraction and immunoprecipitation with the S9.6 antibody. (3) If restriction enzyme fragmentation needs to be applied (e.g., in some cases, sonication might be too harsh to capture transient or very weak RNA-DNA hybrid interactions), we advise the careful control of DNA fragment size distribution before immunoprecipitation.

An important premise is that our recommendations apply to the experimental conditions investigated by this study. Generalization should be avoided since altering critical parameters in the experiment (e.g., incorporating S1 nuclease [S1-DRIP] [Wahba et al. 2016] or lambda exonuclease digestion [DRIP-exo] [Ohle et al. 2016], or changing the model organism) might significantly affect the outcome of RNA-DNA hybrid detection.

In conclusion, the DRIP method remains a gold-standard for identifying bona fide R-loop binding sites across individual chromosomes, but a continued effort is needed to find alternatives and test complementary protocols. We hope that this aim has been achieved, at least in part, by this study, and will help recognize real R-loop binding events and enable a better interpretation of DRIP-seq mapping data.

Methods

Detection of RNA-DNA hybrids by DNA-RNA immunoprecipitation

DRIP classifiers 1–16

Cross-linking (Step 1)

Cross-linking of Jurkat cells (experiments 1–8) was done with 1% paraformaldehyde (UP) for 10 min, then quenched with

Downloaded from genome.cshlp.org on October 26, 2021 - Published by Cold Spring Harbor Laboratory Press

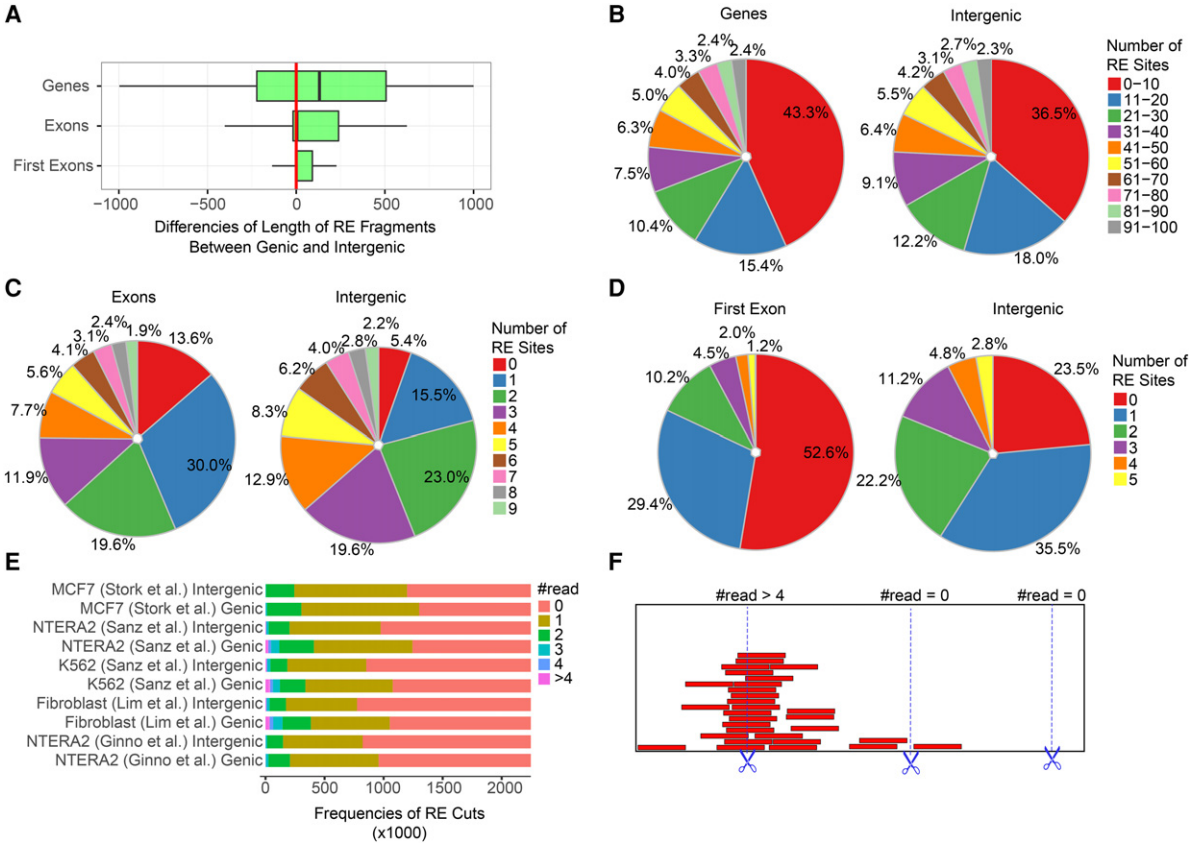


Figure 4. Analysis of restriction sites over genic and intergenic regions. (*A*) Restriction fragment lengths over genic regions (gene bodies, exons, first exons) are significantly larger compared to intergenic regions. The plot shows the difference of genic (observed) and intergenic (expected) fragment sizes in base pairs. The following enzymes were applied in combination: HindIII, EcoRI, BsrGI, XbaI, and SspI. (*B–D*) The number of restriction sites over genic regions is significantly lower compared to intergenic regions. Colors indicate the proportion of cutting sites in each category. Red and blue slices, marking the rarest restriction site frequencies, are prevalent over genic elements in each pie chart. (*E*) Cutting efficiency of restriction enzymes applied in the indicated DRIP-seq experiments. Zero read: the restriction site was cut. Greater equal than one read: the restriction site was uncut in a fraction of cells. There were uncut reads (sites) over half of the theoretical restriction sites. The proportion of uncut reads was even higher within gene coding regions compared to intergenic regions. See the model of cutting efficiency in panel *F*.

2.5 M glycine (pH 6, final concentration: 500 mM) for 5 min at room temperature. Cross-linking was omitted from experiments 9–16.

Cell lysis

Cells were lysed in 1 mL lysis buffer composed of 500 µL 2× lysis buffer (1% SDS, 20 mM Tris-HCl pH 7.5, 40 mM EDTA pH 8, 100 mM NaCl, ddH_2O) plus 500 µL TE buffer (100 mM Tris-HCl pH 8, 10 mM EDTA pH 8) per 5 million cells. Cell lysis was performed at two different temperatures: either at 65°C for 7 h, or at 37°C overnight, as indicated in the text.

Phenol–chloroform extraction of total nucleic acid (Step 2)

In experiments 1–4 and 9–12, total nucleic acid was prepared by phenol-chloroform extraction. Before the phenol-chloroform extraction step, the nucleic acid preps were treated with 10 µL of Proteinase K (20 mg/mL; Thermo Fisher Scientific) at 65°C for 7 h, or at 37°C overnight, to remove the proteins. The extracted DNA was precipitated with 1/10 volume 3 M Na-acetate (pH 5.2) plus 1 volume of isopropanol. The DNA pellet was dissolved in 200 µL of 10 mM Tris-HCl pH 8.

Silica membrane–based (kit) extraction of total nucleic acid (Step 2)

In experiments 5–8 and 13–16, total nucleic acid was isolated by the NucleoSpin Tissue kit (Macherey-Nagel) according to the manufacturer's protocol, except the cell lysis step was performed either at 65°C for 7 h (according to the kit protocol), or at 37°C overnight, where indicated in the text. Nucleic acids were eluted in 500 µL of elution buffer (5 mM Tris-HCl pH 8.5).

Removal of free RNA by RNase A treatment (Step 3)

In experiments 3–4, 7–8, 11–12, and 15–16, the DNA purification step was directly followed by the RNase A digestion of free ribonucleic acids. The purified DNA preps (from Step 2) were supplemented with 18 µL of 5 M NaCl and 2 µL of RNase A (10 mg/mL; UD-GenoMed Ltd.) in a buffer containing 10 mM Tris-HCl (pH 8) and 300 mM NaCl (V = 300 µL) at 37°C for 1 h. RNase A-treated samples were repurified either by phenol-chloroform extraction (experiments 4, 12) or by the NucleoSpin Gel and PCR Clean-up kit (Macherey-Nagel) (experiments 8, 16). Phenol-chloroform–extracted DNA was dissolved in 100 µL of 5 mM Tris-HCl pH 8.5. The DNA purified with the kit was eluted in 5 mM Tris-HCl pH 8.5.

Downloaded from genome.cshlp.org on October 26, 2021 - Published by Cold Spring Harbor Laboratory Press

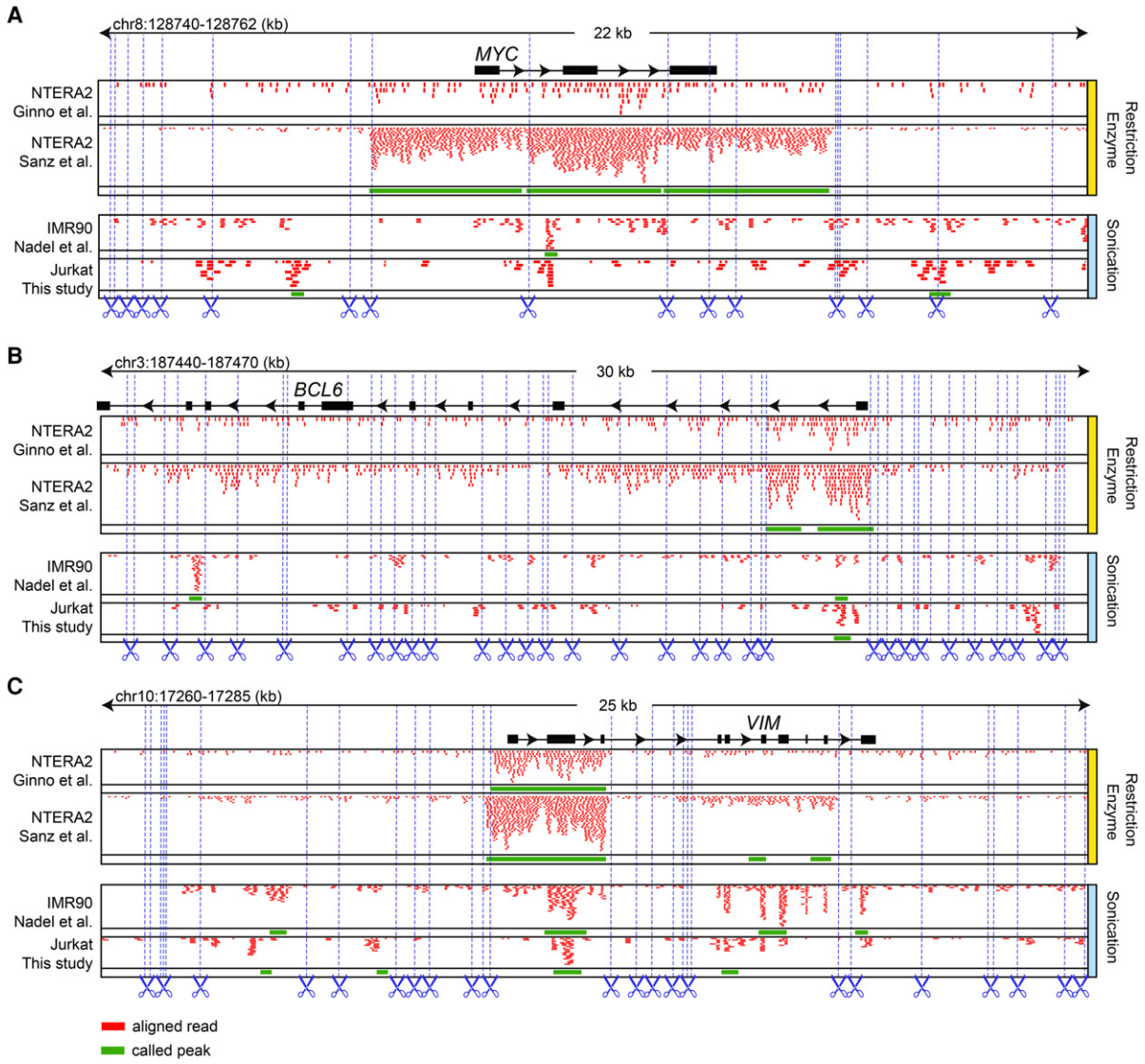


Figure 5. Large restriction fragments over gene bodies cause uncertainty in the precise localization of R-loops, potentially impeding their functional annotation. (*A*–*C*) Genome browser tracks showing three representative examples (*MYC*, *BCL6*, and *VIM*). *Upper* two tracks: restriction fragment-sized R-loops are prevalent over the 5′ end of genes, vastly exceeding the gene borders in the case of *MYC*. *Lower* two tracks: the precise genomic position of R-loops was resolved in the sonicated group of samples. Green boxes represent R-loop enriched regions predicted by the peak callers. Blue dashed lines represent cutting sites for restriction enzymes (HindIII, EcoRI, BsrGI, XbaI, and SspI).

Nucleic acid fragmentation by sonication (Step 4)

In experiments 1, 3, 5, 7, 9, 11, 13, 15, the purified nucleic acid preps were sonicated in a buffer of 10 mM Tris-HCl pH 8.5 supplemented with 300 mM NaCl (V = 300 µL) for 2 × 5 min (30 sec ON, 30 sec OFF, LOW; Bioruptor, Diagenode) to yield an average DNA fragment size of ~300 bp.

Nucleic acid fragmentation by restriction enzyme digestion (Step 4)

In exp. 2, 4, 6, 8, 10, 12, 14, and 16, purified DNA samples (~25 µg each) were fragmented using a restriction enzyme cocktail of 1 µL HindIII (20 U/µL), 1 µL EcoRI (20 U/µL), 2 µL BsrGI (10 U/µL), 1 µL XbaI (20 U/µL), and 4 µL SspI (5 U/µL) in NEB Buffer 2 (NEB) (V = 300 µL) at 37°C for 4 h.

The fragmented DNA samples were repurified either by phenol-chloroform extraction (experiments 1–4, 9–12) or by the NucleoSpin Gel and PCR Clean-up kit (Macherey-Nagel) (experiments 5–8, 13–16). The DNA was dissolved in 100 µL of 5 mM Tris-HCl pH 8.5.

Two percent (V/V%) of the DNA preps were kept as input DNA for the DRIP-qPCR measurement. Half of samples were treated with 8 µL of RNase H (5000 U/mL; NEB) in a total volume of 80 µL, at 37°C overnight.

DRIP classifiers 17–24

Cross-linking (Step 1)

Cross-linking of Jurkat cells (experiments 17–20) was done with 1% paraformaldehyde (UP) for 10 min, then quenched with 2.5 M glycine (pH 6, final concentration: 500 mM) for 5 min at room temperature. Cross-linking was omitted from experiments 21–24.

Downloaded from genome.cshlp.org on October 26, 2021 - Published by Cold Spring Harbor Laboratory Press

Chromatin preparation (Step 2): cell lysis

Cells were lysed in 750 µL of ChIP lysis buffer (50 mM HEPES-KOH at pH 7.5, 140 mM NaCl, 1 mM EDTA at pH 8, 1% Triton X-100, 0.1% Na-Deoxycholate, 1% SDS) per 10 million cells and homogenized using Fast Prep-24 5G (MP Biomedicals, speed: 6 m/sec; time: 40 sec; 2 cycles; pause time: 120 sec; A lysing matrix).

Chromatin fragmentation by sonication (Step 3)

Three hundred microliters of chromatin preps were sonicated for 2 × 5 min (30 sec ON, 30 sec OFF, LOW, Bioruptor) to yield an average DNA fragment size of ~300 bp.

Removal of free RNA by RNase A treatment (Step 4)

In experiments 19, 20, 23, and 24, the sonication step was directly followed by the RNase A digestion of free ribonucleic acids. The fragmented chromatin was supplemented with 270 µL of 5 M NaCl (300 mM) and 10 µL of RNase A (10 mg/mL; UD-GenoMed Ltd.) in 4500 µL of TE buffer (10 mM Tris-HCl pH 8, 10 mM EDTA pH 8) at 37°C for 1 h.

Before Step 5, the chromatin preps were treated with 30 µL of Proteinase K (20 mg/mL; Thermo Fisher Scientific) at 65°C overnight to remove the proteins and reverse the cross-links.

Phenol–chloroform extraction of total nucleic acid (Step 5)

In experiments 17, 19, 21, and 23, total nucleic acid was prepared by phenol-chloroform extraction. The extracted DNA was precipitated with 1/10 volume 3 M Na-acetate (pH 5.2) plus 1 volume of isopropanol. The DNA pellet was dissolved in 100 µL of 5 mM Tris-HCl pH 8.5.

Silica membrane–based (kit) extraction of total nucleic acid (Step 5)

In experiments 18, 20, 22, and 24, total nucleic acids were isolated by the NucleoSpin Gel and PCR Clean-up kit (Macherey-Nagel) according to the manufacturer's protocol. Nucleic acids were eluted in 100 µL of elution buffer (5 mM Tris-HCl pH 8.5).

Two percent (V/V%) of the DNA preps were kept as input DNA for the DRIP-qPCR measurement. Half of the samples were treated with 8 µL of RNase H (5000 U/mL; NEB) in a total volume of 80 µL at 37°C overnight.

RNA–DNA hybrid immunoprecipitation with the S9.6 antibody

Dynabeads Protein A magnetic beads (Thermo Fisher Scientific) were pre-blocked with PBS/EDTA containing 0.5% BSA. To immobilize the S9.6 antibody, 50 µL pre-blocked Dynabeads Protein A were incubated with 10 µg of S9.6 antibody in IP buffer (50 mM Hepes/KOH at pH 7.5; 0.14 M NaCl; 5 mM EDTA; 1% Triton X-100; 0.1% Na-Deoxycholate, ddH_2O) at 4°C for 4 h with rotation. Six micrograms of digested genomic DNA were added to the mixture and gently rotated at 4°C overnight. Beads were recovered and washed successively with 1 mL lysis buffer (low salt, 50 mM Hepes/KOH pH 7.5, 0.14 M NaCl, 5 mM EDTA pH 8, 1% Triton X-100, 0.1% Na-Deoxycholate), 1 mL lysis buffer (high salt, 50 mM Hepes/KOH pH 7.5, 0.5 M NaCl, 5 mM EDTA pH 8, 1% Triton X-100, 0.1% Na-Deoxycholate), 1 mL wash buffer (10 mM Tris-HCl pH 8, 0.25 M LiCl, 0.5% NP-40, 0.5% Na-Deoxycholate, 1 mM EDTA pH 8), and 1 mL TE (100 mM Tris-HCl pH 8, 10 mM EDTA pH 8) at 4°C, two times. Elution was performed in 100 µL of elution buffer (50 mM Tris-HCl pH 8, 10 mM EDTA, 1% SDS) for 15 min at 65°C. After purification with the NucleoSpin Gel and PCR Clean-up kit (Macherey-Nagel), nucleic acids were eluted in 55 µL of elution buffer (5 mM Tris-HCl pH 8.5). The recovered DNA was then analyzed by quantitative real-time PCR (qPCR). qPCR was performed with a LightCycler 480 SYBR Green I Master (Roche) and analyzed on a QuantStudio 12K Flex Real-Time PCR System (Thermo Fisher Scientific). Primer sequences are listed in Supplemental Table S3. qPCR results were analyzed using the comparative C_T method. The RNA-DNA hybrid enrichment was calculated based on the IP/Input ratio.

DRIP-sequencing

DRIP-sequencing was performed in human Jurkat cells and naive T $CD4^+$ lymphocytes. A full description of the DRIP-seq experiment and bioinformatics analysis can be found in the Supplemental Material.

Receiver operating characteristic analysis

ROC curves were obtained for each DRIP variable (DRIP experiments) by ranking the studied genomic loci having known RNA-DNA hybrid states (based on the training set) according to their DRIP-qPCR profile, starting from the lowest to the highest estimated DRIP scores and then calculating sensitivity and specificity. The ROC curves plotted the sensitivity or true-positive rate (TPR) against the false-positive rate (FPR) or 1-specificity, estimated as follows: TPR = P(positive DRIP-qPCR result|R-loop present), FPR = P(positive DRIP-qPCR result|R-loop absent), where P means conditional probability. The AUC values were then calculated from the observed DRIP-qPCR (IP/input) yields using the pROC algorithm.

Data access

DRIP Sequencing data from this study have been submitted to the NCBI Sequence Read Archive (SRA; https://www.ncbi.nlm.nih.gov/sra) under accession number SRP095885.

Acknowledgments

L.S. received funding from the Hungarian Academy of Sciences (Lendület programme, Magyar Tudományos Akadémia, LP2015-9/2015), from the International Center for Genetic Engineering and Biotechnology (CRP-ICGEB), Italy (CRP/HUN13-01), from the European Union, Seventh Framework Programme (FP7/Marie Curie Actions/CIG_#292259), from IMéRA/Inserm and the Aix-Marseille University (France), from NKFIH_ERC_HU_#117670, and from H2020/NKFIH_GINOP-2.3.2-15-2016-00024 and -00043 (G.H.). We thank Ibolya Fürtös for the excellent technical assistance and for the Genomic Medicine and Bioinformatics Core Facility (University of Debrecen) for the NGS service. We thank Dr. György Fenyőfalvi for critical discussions on the topic and for his idea about the RNase H-like activity of RNase A. We thank Dr. Gábor Szabó for providing us with the S9.6 antibody.

Author contributions: B.B.O., T.K.R., É.S., É.N., and Á.M.L. conceived and performed the experiments; A.M., É.R., G.H., and Z.K. provided reagents, expertise, and feedback; L.H., Z.K., and L.S. analyzed the data; and L.S. wrote the manuscript and secured funding.

References

Alzu A, Bermejo R, Begnis M, Lucca C, Piccini D, Carotenuto W, Saponaro M, Brambati A, Cocito A, Foiani M, et al. 2012. Senataxin associates with replication forks to protect fork integrity across RNA-polymerase-II-transcribed genes. *Cell* **151:** 835–846.

Baranello L, Kouzine F, Sanford S, Levens D. 2016. ChIP bias as a function of cross-linking time. *Chromosom Res* **24:** 175–181.

Beneke S, Meyer K, Holtz A, Hüttner K, Bürkle A. 2012. Chromatin composition is changed by poly(ADP-ribosyl)ation during chromatin immunoprecipitation. *PLoS One* **7:** e32914.

Downloaded from genome.cshlp.org on October 26, 2021 - Published by Cold Spring Harbor Laboratory Press

Benore-Parsons M, Ayoub MA. 1997. Presence of RNase A causes aberrant DNA band shifts. *Biotechniques* **23:** 128–131.
Bhatia V, Barroso SI, García-Rubio ML, Tumini E, Herrera-Moyano E, Aguilera A. 2014. BRCA2 prevents R-loop accumulation and associates with TREX-2 mRNA export factor PCID2. *Nature* **511:** 362–365.
Boque-Sastre R, Soler M, Oliveira-Mateos C, Portela A, Moutinho C, Sayols S, Villanueva A, Esteller M, Guil S. 2015. Head-to-head antisense transcription and R-loop formation promotes transcriptional activation. *Proc Natl Acad Sci* **112:** 5785–5790.
Boulé J-B, Zakian VA. 2007. The yeast Pif1p DNA helicase preferentially unwinds RNA DNA substrates. *Nucleic Acids Res* **35:** 5809–5818.
Brown TA, Tkachuk AN, Clayton DA. 2008. Native R-loops persist throughout the mouse mitochondrial DNA genome. *J Biol Chem* **283:** 36743–36751.
Castellano-Pozo M, Santos-Pereira JM, Rondón AG, Barroso S, Andújar E, Pérez-Alegre M, García-Muse T, Aguilera A. 2013. R loops are linked to histone H3 S10 phosphorylation and chromatin condensation. *Mol Cell* **52:** 1–8.
Chan YA, Aristizabal MJ, Lu PYT, Luo Z, Hamza A, Kobor MS, Stirling PC, Hieter P. 2014. Genome-wide profiling of yeast DNA:RNA hybrid prone sites with DRIP-chip. *PLoS Genet* **10:** e1004288.
Chen PB, Chen HV, Acharya D, Rando OJ, Fazzio TG. 2015. R loops regulate promoter-proximal chromatin architecture and cellular differentiation. *Nat Struct Mol Biol* **22:** 999–1007.
Chon H, Sparks JL, Rychlik M, Nowotny M, Burgers PM, Crouch RJ, Cerritelli SM. 2013. RNase H2 roles in genome integrity revealed by unlinking its activities. *Nucleic Acids Res* **41:** 3130–3143.
Cloutier SC, Wang S, Ma WK, Al Husini N, Dhoondia Z, Ansari A, Pascuzzi PE, Tran EJ. 2016. Regulated formation of lncRNA-DNA hybrids enables faster transcriptional induction and environmental adaptation. *Mol Cell* **61:** 393–404.
Domínguez-Sánchez MS, Barroso S, Gómez-González B, Luna R, Aguilera A. 2011. Genome instability and transcription elongation impairment in human cells depleted of THO/TREX. *PLoS Genet* **7:** e1002386.
Dona F, Houseley J. 2014. Unexpected DNA loss mediated by the DNA binding activity of ribonuclease A. *PLoS One* **9:** e115008.
El Hage A, French SL, Beyer AL, Tollervey D. 2010. Loss of Topoisomerase I leads to R-loop-mediated transcriptional blocks during ribosomal RNA synthesis. *Genes Dev* **24:** 1546–1558.
El Hage A, Webb S, Kerr A, Tollervey D. 2014. Genome-wide distribution of RNA-DNA hybrids identifies RNase H targets in tRNA genes, retrotransposons and mitochondria. *PLoS Genet* **10:** e1004716.
Gan W, Guan Z, Liu J, Gui T, Shen K, Manley JL, Li X. 2011. R-loop-mediated genomic instability is caused by impairment of replication fork progression. *Genes Dev* **25:** 2041–2056.
García-Rubio ML, Pérez-Calero C, Barroso SI, Tumini E, Herrera-Moyano E, Rosado IV, Aguilera A. 2015. The Fanconi anemia pathway protects genome integrity from R-loops. *PLoS Genet* **11:** e1005674.
Ginno PA, Lott PL, Christensen HC, Korf I, Chédin F. 2012. R-loop formation is a distinctive characteristic of unmethylated human CpG island promoters. *Mol Cell* **45:** 814–825.
Groh M, Lufino MMP, Wade-Martins R, Gromak N. 2014. R-loops associated with triplet repeat expansions promote gene silencing in Friedreich ataxia and fragile X syndrome. *PLoS Genet* **10:** e1004318.
Hatchi E, Skourti-Stathaki K, Ventz S, Pinello L, Yen A, Kamieniarz-Gdula K, Dimitrov S, Pathania S, McKinney KM, Eaton ML, et al. 2015. BRCA1 recruitment to transcriptional pause sites is required for R-loop-driven DNA damage repair. *Mol Cell* **57:** 636–647.
Herrera-Moyano E, Mergui X, García-Rubio ML, Barroso S, Aguilera A. 2014. The yeast and human FACT chromatin-reorganizing complexes solve R-loop-mediated transcription-replication conflicts. *Genes Dev* **1:** 735–748.
Hu Z, Zhang A, Storz G, Gottesman S, Leppla SH. 2006. An antibody-based microarray assay for small RNA detection. *Nucleic Acids Res* **34:** e52.
Huertas P, Aguilera A. 2003. Cotranscriptionally formed DNA:RNA hybrids mediate transcription elongation impairment and transcription-associated recombination. *Mol Cell* **12:** 711–721.
Jenjaroenpun P, Wongsurawat T, Yenamandra SP, Kuznetsov VA. 2015. QmRLFS-finder: amodel, web server and stand-alone tool for prediction and analysis of R-loop forming sequences. *Nucleic Acids Res* **43:** W527–W534.
Li X, Manley JL. 2005. Inactivation of the SR protein splicing factor ASF/SF2 results in genomic instability. *Cell* **122:** 365–378.
Lim YW, Sanz LA, Xu X, Hartono SR, Chédin F. 2015. Genome-wide DNA hypomethylation and RNA:DNA hybrid accumulation in Aicardi-Goutières syndrome. *eLife* **4:** e08007.
Loomis EW, Sanz LA, Chédin F, Hagerman PJ. 2014. Transcription-associated R-loop formation across the human FMR1 CGG-repeat region. *PLoS Genet* **10:** e1004294.
Marinello J, Bertoncini S, Aloisi I, Cristini A, Tagliazucchi GM, Forcato M, Sordet O, Capranico G. 2016. Dynamic effects of topoisomerase i inhibition on R-loops and short transcripts at active promoters. *PLoS One* **11:** e0147053.
McGhee JD, von Hippel PH. 1977. Formaldehyde as a probe of DNA structure. 4. Mechanism of the initial reaction of formaldehyde with DNA. *Biochemistry* **16:** 3276–3293.
Mischo HE, Gómez-González B, Grzechnik P, Rondón AG, Wei W, Steinmetz L, Aguilera A, Proudfoot NJ. 2011. Yeast Sen1 helicase protects the genome from transcription-associated instability. *Mol Cell* **41:** 21–32.
Nadel J, Athanasiadou R, Lemetre C, Wijetunga NA, Broin PÓ, Sato H, Zhang Z, Jeddeloh J, Montagna C, Golden A, et al. 2015. RNA:DNA hybrids in the human genome have distinctive nucleotide characteristics, chromatin composition, and transcriptional relationships. *Epigenetics Chromatin* **8:** 46.
Nakama M, Kawakami K, Kajitani T, Urano T, Murakami Y. 2012. DNA-RNA hybrid formation mediates RNAi-directed heterochromatin formation. *Genes Cells* **17:** 218–233.
Ohle C, Tesorero R, Schermann G, Dobrev N, Sinning I, Fischer T. 2016. Transient RNA-DNA hybrids are required for efficient double-strand break repair. *Cell* **167:** 1001–1013.
Pefanis E, Wang J, Rothschild G, Lim J, Kazadi D, Sun J, Federation A, Chao J, Elliott O, Liu Z-P, et al. 2015. RNA exosome-regulated long non-coding RNA transcription controls super-enhancer activity. *Cell* **161:** 774–789.
Phillips DD, Garboczi DN, Singh K, Hu Z, Leppla SH, Leysath CE. 2013. The sub-nanomolar binding of DNA-RNA hybrids by the single-chain Fv fragment of antibody S9.6. *J Mol Recognit* **26:** 376–381.
Pohjoismäki JLO, Holmes JB, Wood SR, Yang M-Y, Yasukawa T, Reyes A, Bailey LJ, Cluett TJ, Goffart S, Willcox S, et al. 2010. Mammalian mitochondrial DNA replication intermediates are essentially duplex but contain extensive tracts of RNA/DNA hybrid. *J Mol Biol* **397:** 1144–1155.
Rigby RE, Webb LM, Mackenzie KJ, Li Y, Leitch A, Reijns MAM, Lundie RJ, Revuelta A, Davidson DJ, Diebold S, et al. 2014. RNA:DNA hybrids are a novel molecular pattern sensed by TLR9. *EMBO J* **33:** 542–558.
Robin X, Turck N, Hainard A, Tiberti N, Lisacek F, Sanchez J-C, Müller M. 2011. pROC: an open-source package for R and S+ to analyze and compare ROC curves. *BMC Bioinformatics* **12:** 77.
Romanello M, Schiavone D, Frey A, Sale JE. 2016. Histone H3.3 promotes IgV gene diversification by enhancing formation of AID-accessible single-stranded DNA. *EMBO J* **35:** 1452–1464.
Salvi JS, Chan JNY, Szafranski K, Liu TT, Wu JD, Olsen JB, Khanam N, Poon BPK, Emili A, Mekhail K. 2014. Roles for Pbp1 and caloric restriction in genome and lifespan maintenance via suppression of RNA-DNA hybrids. *Dev Cell* **30:** 177–191.
Sanz LA, Hartono SR, Lim YW, Ginno PA, Sanz LA, Hartono SR, Lim YW, Steyaert S, Rajpurkar A, Ginno PA, et al. 2016. Prevalent, dynamic, and conserved R-loop structures associate with specific epigenomic signatures in mammals. *Mol Cell* **63:** 167–178.
Schwab RA, Nieminuszczy J, Shah F, Langton J, Lopez Martinez D, Liang CC, Cohn MA, Gibbons RJ, Deans AJ, Niedzwiedz W. 2015. The Fanconi anemia pathway maintains genome stability by coordinating replication and transcription. *Mol Cell* **60:** 351–361.
Skourti-Stathaki K, Proudfoot NJ, Gromak N. 2011. Human senataxin resolves RNA/DNA hybrids formed at transcriptional pause sites to promote Xrn2-dependent termination. *Mol Cell* **42:** 794–805.
Sollier J, Stork CT, García-Rubio ML, Paulsen RD, Aguilera A, Cimprich KA. 2014. Transcription-coupled nucleotide excision repair factors promote R-loop-induced genome instability. *Mol Cell* **56:** 777–785.
Stork CT, Bocek M, Crossley MP, Sollier J, Sanz LA, Chédin F, Swigut T, Cimprich KA. 2016. Co-transcriptional R-loops are the main cause of estrogen-induced DNA damage. *eLife* **5:** e17548.
Stuckey R, García-Rodríguez N, Aguilera A, Wellinger RE. 2015. Role for RNA:DNA hybrids in origin-independent replication priming in a eukaryotic system. *Proc Natl Acad Sci* **112:** 5779–5784.
Sun Q, Csorba T, Skourti-Stathaki K, Proudfoot NJ, Dean C. 2013. R-loop stabilization represses antisense transcription at the *Arabidopsis* FLC locus. *Science* **340:** 619–621.
Székvölgyi L, Rákosy Z, Bálint BL, Kókai E, Imre L, Vereb G, Bacsó Z, Goda K, Varga S, Balázs M, et al. 2007. Ribonucleoprotein-masked nicks at 50-kbp intervals in the eukaryotic genomic DNA. *Proc Natl Acad Sci* **104:** 14964–14969.
Wahba L, Koshland D. 2013. The Rs of biology: R-loops and the regulation of regulators. *Mol Cell* **50:** 611–612.
Wahba L, Costantino L, Tan FJ, Zimmer A, Koshland D. 2016. S1-DRIP-seq identifies high expression and polyA tracts as major contributors to R-loop formation. *Genes Dev* **30:** 1327–1338.
Wilson-Sali T, Hsieh T-S. 2002. Preferential cleavage of plasmid-based R-loops and D-loops by *Drosophila* topoisomerase IIIβ. *Proc Natl Acad Sci* **99:** 7974–7979.
Yang Y, McBride KM, Hensley S, Lu Y, Chedin F, Bedford MT. 2014. Arginine methylation facilitates the recruitment of TOP3B to chromatin to prevent R loop accumulation. *Mol Cell* **53:** 484–497.

Downloaded from genome.cshlp.org on October 26, 2021 - Published by Cold Spring Harbor Laboratory Press

Yang Y, La H, Tang K, Miki D, Yang L, Wang B, Duan C-G, Nie W, Wang X, Wang S, et al. 2016. SAC3B, a central component of the mRNA export complex TREX-2, is required for prevention of epigenetic gene silencing in *Arabidopsis*. *Nucleic Acids Res* **45:** 181–197.

Yu K, Chedin F, Hsieh C-L, Wilson TE, Lieber MR. 2003. R-loops at immunoglobulin class switch regions in the chromosomes of stimulated B cells. *Nat Immunol* **4:** 442–451.

Yu K, Roy D, Huang F-T, Lieber MR. 2006. Detection and structural analysis of R-loops. *Methods Enzymol* **409:** 316–329.

Zeller P, Padeken J, van Schendel R, Kalck V, Tijsterman M, Gasser SM. 2016. Histone H3K9 methylation is dispensable for *Caenorhabditis elegans* development but suppresses RNA:DNA hybrid-associated repeat instability. *Nat Genet* **48:** 1385–1395.

Zhang ZZ, Pannunzio NR, Han L, Hsieh C-L, Yu K, Lieber MR. 2014a. The strength of an Ig switch region is determined by its ability to drive R loop formation and its number of WGCW sites. *Cell Rep* **8:** 557–569.

Zhang ZZ, Pannunzio NR, Hsieh C-L, Yu K, Lieber MR. 2014b. The role of G-density in switch region repeats for immunoglobulin class switch recombination. *Nucleic Acids Res* **42:** 13186–13193.

Zhang ZZ, Pannunzio NR, Hsieh C-L, Yu K, Lieber MR. 2015. Complexities due to single-stranded RNA during antibody detection of genomic RNA: DNA hybrids. *BMC Res Notes* **8:** 127.

Received December 8, 2016; accepted in revised form March 23, 2017.

***Genome Research* 27:** 1063–1073 (2017)

Corrigendum: RNA-DNA hybrid (R-loop) immunoprecipitation mapping: an analytical workflow to evaluate inherent biases

László Halász, Zsolt Karányi, Beáta Boros-Oláh, Tímea Kuik-Rózsa, Éva Sipos, Éva Nagy, Ágnes Mosolygó-L, Anett Mázló, Éva Rajnavölgyi, Gábor Halmos, and Lóránt Székvölgyi

The authors would like to correct erroneous text relating to the *Step-by-step protocol of the best-performing DRIP experiment (exp. 5)* section in the Supplemental Material. The corrected text is as follows and has been updated in the Revised Supplemental Material online:

"Cells were lysed in the lysis buffer provided by the NucleoSpin Tissue kit (Macherey-Nagel) at 65°C for 7 h (according to the kit protocol), or at 37°C overnight (where indicated in the main text)."

doi: 10.1101/gr.246652.118

Downloaded from genome.cshlp.org on October 26, 2021 - Published by Cold Spring Harbor Laboratory Press

dc_1977_21

RNA-DNA hybrid (R-loop) immunoprecipitation mapping: an analytical workflow to evaluate inherent biases

László Halász, Zsolt Karányi, Beáta Boros-Oláh, et al.

Genome Res. 2017 27: 1063-1073 originally published online March 24, 2017
Access the most recent version at doi:10.1101/gr.219394.116

Supplemental Material
http://genome.cshlp.org/content/suppl/2017/04/24/gr.219394.116.DC1
http://genome.cshlp.org/content/suppl/2019/01/02/gr.219394.116.DC2

Related Content
Corrigendum: RNA-DNA hybrid (R-loop) immunoprecipitation mapping: an analytical workflow to evaluate inherent biases
László Halász, Zsolt Karányi, Beáta Boros-Oláh, et al.
Genome Res. January , 2019 29: 157

References
This article cites 61 articles, 11 of which can be accessed free at:
http://genome.cshlp.org/content/27/6/1063.full.html#ref-list-1

Articles cited in:
http://genome.cshlp.org/content/27/6/1063.full.html#related-urls

Creative Commons License
This article is distributed exclusively by Cold Spring Harbor Laboratory Press for the first six months after the full-issue publication date (see http://genome.cshlp.org/site/misc/terms.xhtml). After six months, it is available under a Creative Commons License (Attribution-NonCommercial 4.0 International), as described at http://creativecommons.org/licenses/by-nc/4.0/.

Email Alerting Service
Receive free email alerts when new articles cite this article - sign up in the box at the top right corner of the article or **click here.**

To subscribe to *Genome Research* go to:
https://genome.cshlp.org/subscriptions

© 2017 Halász et al.; Published by Cold Spring Harbor Laboratory Press

Molecular Therapy
Methods & Clinical Development
dc_1977_21
Commentary

In Silico Restriction Enzyme Digests to Minimize Mapping Bias in Genomic Sequencing

Jason Roszik,[1,2] György Fenyőfalvi,[3] László Halász,[3,4] Zsolt Karányi,[3] and Lóránt Székvölgyi[3,4]

http://dx.doi.org/10.1016/j.omtm.2017.06.003

A commonly used genome fragmentation method in next generation sequencing, restriction endonuclease (RE) digestion, may severely compromise genomic mapping resolution and prevent the functional annotation of certain chromosomal regions unless REs are applied in correct combinations to sample all genomic regions with an equal probability.

Genome fragmentation by REs is routinely used in multiple genomic mapping technologies, including RNA-DNA hybrid (R-loop) immunoprecipitation sequencing (DRIP-seq),[1,2] chromosome conformation capture (4C/5C, Hi-C),[3] reduced-representation bisulfite sequencing (RRBS),[4] and restriction site associated marker (RAD) genotyping.[5] The performance of these approaches depends on (1) the length distribution of the restriction fragments (determining the spatial resolution of the assay) and (2) the randomness of RE digestion (ensuring that all genomic regions are sampled with an equal probability).[6] Therefore, selecting the proper combination of REs for genome fragmentation is of crucial importance to obtain representative next-generation sequencing (NGS) libraries and to assign clear biological functions to the mapped regions.

Using the DRIP-seq technique, we have recently shown that this technology contains inherent biases related to RE digestion that might prevent functional annotation of a significant fraction of R-loops.[2] R-loops, nucleic acid structures that are composed of an RNA-DNA hybrid and a single-stranded DNA, are involved in multiple cellular processes and may also mediate genomic instability in a pathological context. The DRIP method uses an anti-RNA-DNA hybrid antibody to capture RNA-DNA hybrids associated with RE-fragmented DNA or chromatin, followed by fragment mapping to the genome. The main reason for the over-representation of lengthy DRIP fragments may be that the distribution of restriction enzyme cutting sites is not random in the human genome.[7] This bias is especially enhanced over the first exons. The over-representation of first exons in RE-fragmented samples may also be an issue in other species and sequencing methods. For instance, the mouse genome also contains long intronic sequences that may cause similar biases. Similar to the DRIP method, suboptimal RE fragment size distribution and first exon bias might affect the outcome and interpretation of other frequently used genomic technologies (e.g., all C-based methods [4C, 5C, and Hi-C]), potentially introducing false-positive spatial contacts that fall proximal to open reading frames (ORFs), especially to first exons. Finally, the estimation of the evolutionary conservation of R-loop binding sites between species that reflect the sequence homology/divergence of exonic DRIP fragments,[8] but precisely located R-loop binding sites, is potentially also problematic.

Superimposed on the RE bias, multiple other genome characteristics can affect the efficacy of RE digestion. DNA methylation is present in higher organisms, and the majority of REs do not cut at methylated cytosines. Furthermore, most REs do not cut DNA-RNA hybrids that are prevalent over the chromosomes (constituting 5%–8% of the eukaryote genome). Restriction enzyme accessibility is also limited by the chromatin (nucleosome) structure that inherently prefers the cutting of linker DNA sequences.

The randomness and uniformity of restriction fragment length distributions can be tested for any combination of REs using in silico restriction endonuclease digests (Figure 1), and RE cocktails with theoretically justified cutting parameters can be selected for use in experiments. We recommend using the DECIPHER R package, which is available in Bioconductor.[9] To predict the expected DNA fragments, the "digestDNA" function can be used to perform in silico restriction digestion of given DNA sequences. Issues related to CpG methylation can be experimentally addressed by methylation-insensitive REs that cleave methylated DNA. RNase H1 digestion of nucleic acid preps can also be applied to remove RNA from DNA-RNA hybrids. Furthermore, short treatment of live cells with chromatin decompaction agents (e.g., HDAC inhibitors) may provide increased RE accessibility in experiments involving in situ RE fragmentation (e.g., Hi-C). Collectively, the above recommendations can help identify RE cocktails and experimental conditions that result in proper DNA fragment size distributions and optimal resolution in genomic sequencing technologies.

[1]Department of Melanoma Medical Oncology, The University of Texas MD Anderson Cancer Center, 1515 Holcombe Blvd., Houston, TX 77030, USA; [2]Department of Genomic Medicine, The University of Texas MD Anderson Cancer Center, 1515 Holcombe Blvd., Houston, TX 77030, USA; [3]MTA-DE Momentum, Genome Architecture and Recombination Research Group, Research Centre for Molecular Medicine, University of Debrecen, Nagyerdei krt 98, Debrecen 4032, Hungary; [4]Department of Biochemistry and Molecular Biology, University of Debrecen, Egyetem sq. 1, Debrecen 4032, Hungary

Correspondence: Lóránt Székvölgyi, PhD, Department of Biochemistry and Molecular Biology, University of Debrecen, Nagyerdei krt. 98, Debrecen 4032, Hungary.
E-mail: lorantsz@med.unideb.hu

This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

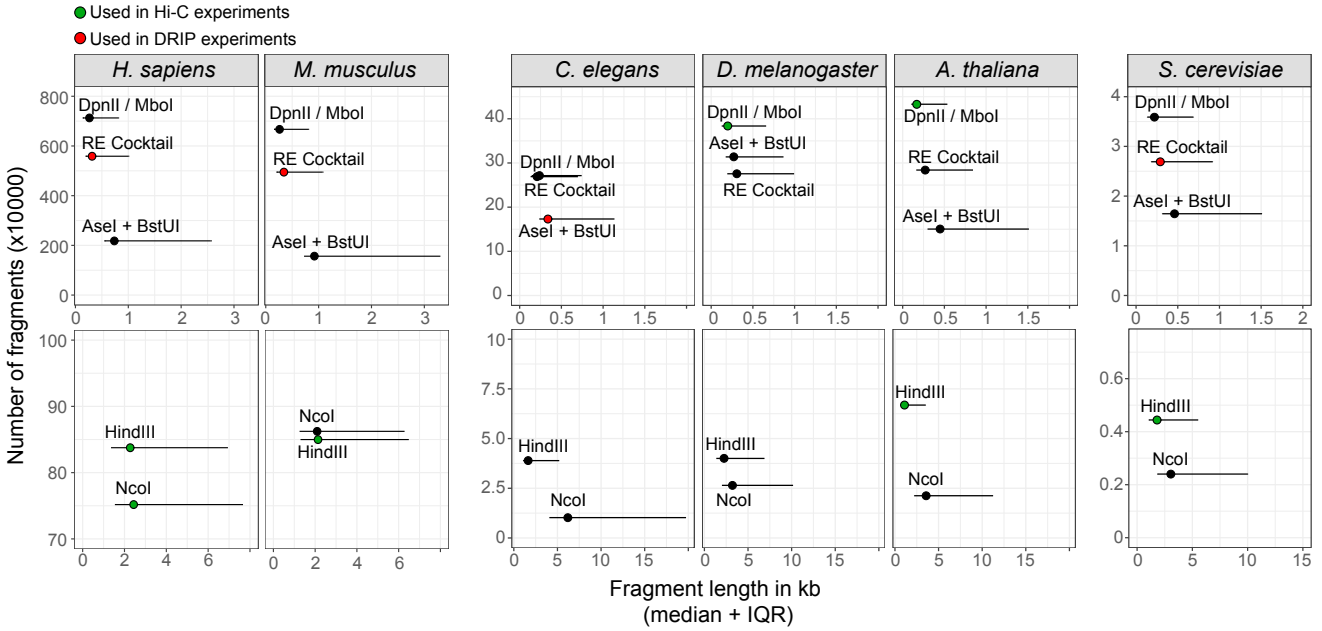


Figure 1. Genome Fragmentation by In Silico Restriction Enzyme Digestion in Species That Were Analyzed by DRIP-seq or HI-C

The absolute number of restriction fragments is shown in terms of the average fragment lengths (mean + interquartile range [IQR]) obtained by the indicated restriction enzymes applied alone or in combination. RE cocktail denotes the HindIII, EcoRI, BsrGI, XbaI, and SspI enzymes.

ACKNOWLEDGMENTS

This work was supported by the Hungarian Academy of Sciences (Lendület programme, LP2015-9/2015) and grants from the National Research, Development and Innovation Office (NKFIH-ERC-HU-117670 and GINOP-2.3.2-15-2016-00024).

REFERENCES

1. Ginno, P.A., Lott, P.L., Christensen, H.C., Korf, I., and Chédin, F. (2012). R-loop formation is a distinctive characteristic of unmethylated human CpG island promoters. Mol. Cell *45*, 814–825.

2. Halász, L., Karányi, Z., Boros-Oláh, B., Kuik-Rózsa, T., Sipos, É., Nagy, É., Mosolygó-L, Á., Mázló, A., Rajnavölgyi, É., Halmos, G., and Székvölgyi, L. (2017). RNA-DNA hybrid (R-loop) immunoprecipitation mapping: an analytical workflow to evaluate inherent biases. Genome Res. *27*, 1063–1073.

3. Lieberman-Aiden, E., van Berkum, N.L., Williams, L., Imakaev, M., Ragoczy, T., Telling, A., Amit, I., Lajoie, B.R., Sabo, P.J., Dorschner, M.O., et al. (2009). Comprehensive mapping of long-range interactions reveals folding principles of the human genome. Science *326*, 289–293.

4. Meissner, A., Gnirke, A., Bell, G.W., Ramsahoye, B., Lander, E.S., and Jaenisch, R. (2005). Reduced representation bisulfite sequencing for comparative high-resolution DNA methylation analysis. Nucleic Acids Res. *33*, 5868–5877.

5. Miller, M.R., Dunham, J.P., Amores, A., Cresko, W.A., and Johnson, E.A. (2007). Rapid and cost-effective polymorphism identification and genotyping using restriction site associated DNA (RAD) markers. Genome Res. *17*, 240–248.

6. Bystrykh, L.V. (2013). A combinatorial approach to the restriction of a mouse genome. BMC Res. Notes *6*, 284.

7. Hartono, S.R., Korf, I.F., and Chédin, F. (2015). GC skew is a conserved property of unmethylated CpG island promoters across vertebrates. Nucleic Acids Res. *43*, 9729–9741.

8. Sanz, L.A., Hartono, S.R., Lim, Y.W., Steyaert, S., Rajpurkar, A., Ginno, P.A., Xu, X., and Chédin, F. (2016). Prevalent, Dynamic, and Conserved R-Loop Structures Associate with Specific Epigenomic Signatures in Mammals. Mol. Cell *63*, 167–178.

9. Wright, E.S. (2016). Using DECIPHER v2.0 to analyze big biological sequence data in R. R J. *8*, 352–359.

dc_1977_21

DNA Repair 84 (2019) 102642

Contents lists available at ScienceDirect

DNA Repair

journal homepage: www.elsevier.com/locate/dnarepair

Drugging the R-loop interactome: RNA-DNA hybrid binding proteins as targets for cancer therapy★,★★

Beáta Boros-Oláh[a,1], Nikoletta Dobos[b,1], Lilla Hornyák[a], Zoltán Szabó[c], Zsolt Karányi[a,d], Gábor Halmos[b], Jason Roszik[e,*], Lóránt Székvölgyi[a,*]

[a] MTA-DE Momentum, Genome Architecture and Recombination Research Group, Department of Biochemistry and Molecular Biology, Faculty of Medicine, University of Debrecen, Debrecen, Hungary
[b] Department of Biopharmacy, Faculty of Pharmacy, University of Debrecen, Debrecen, Hungary
[c] Department of Emergency Medicine, Faculty of Medicine, University of Debrecen, Debrecen, Hungary
[d] Department of Internal Medicine, Faculty of Medicine, University of Debrecen, Debrecen, Hungary
[e] Departments of Melanoma Medical Oncology and Genomic Medicine, The University of Texas MD Anderson Cancer Center, 1515 Holcombe Blvd., Houston, TX, 77030, USA

ARTICLE INFO

Keywords:
R-loop
RNA-DNA hybrid
Cancer diagnostics and therapy
Chemotherapy
Clinical trial
Survival association
Drug sensitivity

ABSTRACT

Unravelling the origin of genetic alterations from point mutations to chromosomal rearrangements was greatly enhanced by the discovery of RNA-DNA hybrids (R-loops) that behave as hotspots of genomic instability in a variety of organisms. Current models suggest that uncontrolled R-loops are a hazard to genome integrity, therefore, identifying proteins that are involved in recognising and signalling R-loop structures are of key importance. Herein we analysed key RNA-DNA hybrid binding proteins in humans taking advantage of large-scale gene expression, survival rate, and drug-sensitivity data from cancer genomics databases. We show that expression of RNA-DNA hybrid binding proteins in various cancer types is associated with survival and may have contrasting outcomes in responding to therapeutic treatments. Based on the revealed pharmacogenomic landscape of human RNA-DNA hybrid binding proteins, we propose that R-loops and R-loop binding proteins are potentially relevant new epigenetic markers and therapeutic targets in multiple cancers.

1. Introduction

R-loops are special, three-stranded nucleic acid structures, composed of an RNA-DNA hybrid and a non-template, single-stranded DNA. R-loops have been implicated in a number of human diseases including repeat-expansion disorders, neurological syndromes, and cancer [1–3]. The molecular symptoms of cancer resemble the genome instability phenotype of human cell lines that accumulate R-loops and undergo replication/transcriptional stress-induced DNA damage [4]. Oncogenic mutation of $HRAS^{V12}$, for instance, has been shown to increase the protein levels of endogenous RNaseH1 (the enzyme that specifically degrades RNA-DNA hybrids) [5], which further supports a mechanistic link between R-loop formation and tumorigenesis.

It is well-established that single-agent therapy seldom leads to cure in oncology, therefore identifying novel targets for combination therapies is necessary [6–8]. Unfortunately, drug-targeting is not feasible in the case of tumour suppressor gene inactivating (loss-of-function) mutations (constituting most driver mutations in cancer [6], therefore, finding alternative targets could help prevent the development of drug resistance to targeted therapies. Since R-loops are universal by-products of transcription [9–11], the R-loop tolerance of cancer cells could be exploited to sensitize certain tumours to chemotherapeutic treatments. Reactivating the endogenous response of cancers to R-loop structures and R-loop-mediated genetic damage (R-loop intolerance) may make tumours more vulnerable to DNA damaging agents and cell death, which might potentially lead to the development of new combination therapies involving cytotoxic regimens and R-loop-targeting drugs. Various ligands are able target DNA-RNA hybrids, including i) small molecule intercalators that directly react with RNA-DNA hybrids, e.g. ethidium bromide, the aminoglycosides

★ This Special Issue is edited by Guest Editor.
★★ This article is part of the special issue “RNases H”.
* Corresponding author.
E-mail addresses: JRoszik@mdanderson.org (J. Roszik), lorantsz@med.unideb.hu (L. Székvölgyi).
[1] These authors contributed equally to this study.

https://doi.org/10.1016/j.dnarep.2019.102642
Received 2 March 2019; Received in revised form 16 May 2019; Accepted 2 July 2019
Available online 06 July 2019
1568-7864/ © 2019 The Authors. Published by Elsevier B.V. This is an open access article under the CC BY license (http://creativecommons.org/licenses/BY/4.0/).

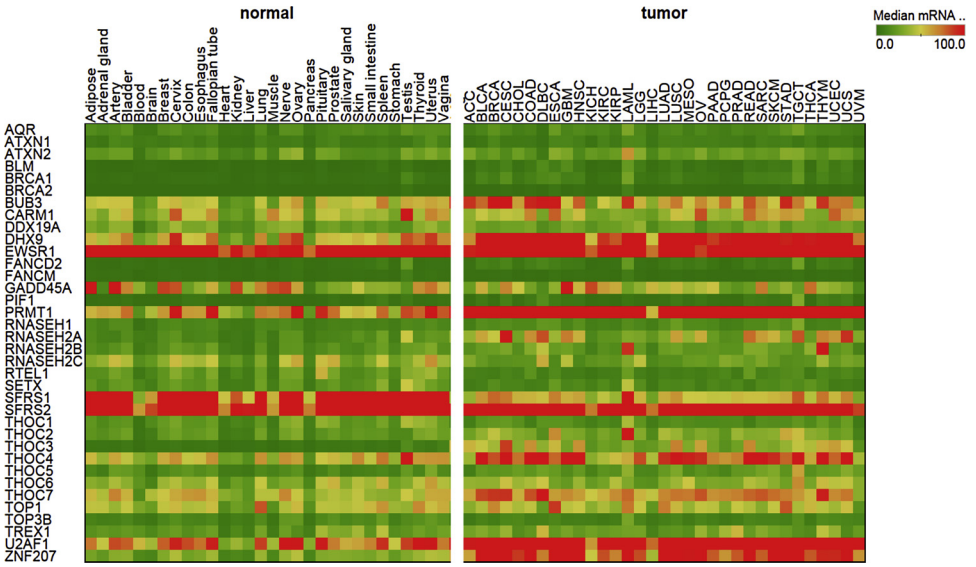


Fig. 1. Median mRNA expression levels of R-loop genes in healthy tissues and in primary tumours, extracted from the Cancer Genome Atlas database. The intensity of green and red colours is proportional to expression levels.

neomycin and paramomycin, actinomycin D [12,13], ii) G quadruplex (G4) ligands since G4s form in a similar genomic context to R-loops [14,15], e.g. Quarfloxin [16] (CX-3543; targeting rDNA G4s; phase II clinical trial completed for treatment of low grade neuroendocrine carcinoma - NCT00780663), BSU1051 [15] (targeting human telomeric G4s), TMPyP4 [17] (targeting G4s in the c-Myc promoter), pyridostatin [18] (PDS; targeting telomeric G4s and proto-oncogene tyrosine-protein kinase Src), Braco-19, perylenes and naphthalene diimides, bisquinolinium derivatives (Phen-DC(3/6) [19], iii) drugs targeting RNA-DNA hybrid-binding proteins, which have been identified recently by high-throughput screening [20,21], e.g. TDRL-505 analogues targeting RPA [22], or YK-4-279 targeting EWS-FLI1 [23].

There is evidence that R-loops are targetable by anticancer drugs to revert pathological phenotype [24,25]. For instance, in synovial and Ewing sarcoma (SS/ES), the use of clinical ATR inhibitors (ATRi) led to the accumulation of R-loops and increased sensitivity to chemotherapy [26,27]. PARP inhibitors augmented the antitumor activity of ATRi of SS cells, suggesting that combination therapies using ATRi are promising new approaches to treat sarcomas [26]. The anti-tumour drugs trabectedin and lurbinectedin have been shown to induce replicative stress and cell death in an R-loop dependent manner [28]. Using a cell line system where R-loops were stabilized by depleting THOC1 or BRCA2, cells with increased R-loop levels were more sensitive to trabectedin treatment. Consistently, cancers that accumulate R-loops in the absence of BRCA1/2 or Fanconi anaemia proteins show higher sensitivity to trabectedin therapy [29,30].

The EWS-FLI fusion protein has been shown to increase R-loop formation and inactivates BRCA1 that makes Ewing sarcoma cell lines hypersensitive to genotoxic drugs such as etoposide, camptothecin, and PARP inhibitors [31]. Since mutations in EWSR1 and its homologues are associated with several therapeutically challenging cancers, clastogenic agents that augment R-loop mediated stress could be administered as potentially effective co-therapeutic treatments in various tumours that are associated with EWSR1 mutations.

Specific G4 ligands (PDS, Braco-19, and FG) induce R-loop-mediated DNA damage and cell death in human cancer cells [18], establishing a link between the toxic effects of G4 ligands and R-loop formation. CX-5461 is another G4 ligand showing specific toxicity against BRCA-deficient cancer cells and patient-derived xenografts [16]. Repair of DNA damage induced by CX-5461 required BRCA and NHEJ repair pathways. CX-5461 is now in an advanced phase I clinical trial for patients with BRCA1/2 deficient tumours (NCT02719977).

EZM2302 is a selective inhibitor of the histone arginine methylase CARM1, an enzyme that affects R-loop homeostasis by recruiting TOP3B to RNA-DNA hybrid structures [32]. EZM2302 exhibits anti-proliferative effects both in vitro and in vivo [33]. GSK3368715 is a specific inhibitor of the PRMT1 histone arginine methyltransferase that has also been linked to R-loop metabolism [32] and being developed for the treatment of diffuse large B cell lymphoma and solid tumours (phase I clinical trial; NCT03666988).

Several other compounds have also been described to increase R-loop levels, including topoisomerase 1 inhibitors [34] (promoting R-loops by modulating the superhelicity of DNA), spliceosome [35–38] inhibitors (promoting R-loops through the retention of intronic sequences), and reactive aldehydes [39]. Furthermore, the RNA-DNA degrading enzyme RNase H2 has been recognised as a putative anticancer drug target [40]. With the recently identified RNase H2 inhibitors [41], RNaseH2 might serve as an effective cancer target to stabilize RNA-DNA hybrids. Finally, a series of anticancer sulphonamides have been shown to induce proteasomal degradation of the U2AF-related splicing factor coactivator of activating protein-1 in human cancer cell lines [37,42], designating targeted protein degradation by E3 ubiquitin ligases as a potent drug target for selective inactivation of splicing factors to increase R-loop levels for cancer therapy.

In the current study we aimed to investigate whether key R-loop-binding proteins are associated with cancer survival and drug sensitivity. We performed a systematic pharmacogenomic analysis to identify these associations that may suggest that R-loop formation processes in cancer cells could be exploited as biomarkers, therapeutic targets, as well as be used to sensitize certain tumours to chemotherapeutic treatments. Our results offer new avenues for epigenetic therapies that are based on modifying R-loop levels in tumours.

2. Materials and methods

2.1. Tumour types included in the analysis

We included in our analysis 33 primary cancer types available from the Cancer Genome Atlas (TCGA) [43]. These were: Adrenocortical carcinoma (ACC), Bladder urothelial carcinoma (BLCA), Breast invasive carcinoma (BRCA), Cervical and endocervical cancers (CESC), Cholangiocarcinoma (CHOL), Colon adenocarcinoma (COAD), Lymphoid Neoplasm Diffuse Large B-cell Lymphoma (DLBC), Esophageal

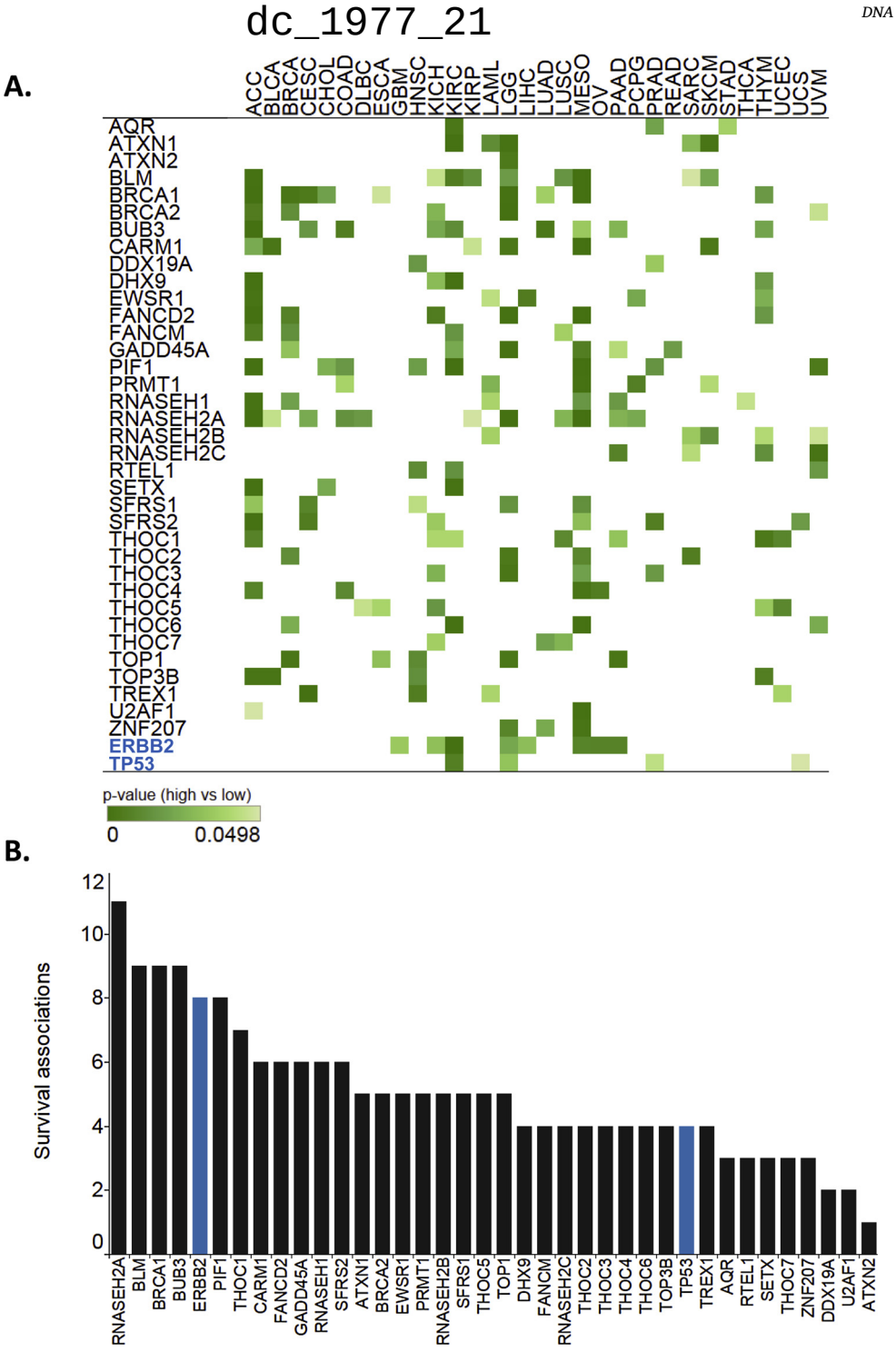


Fig. 2. Association of R-loop gene expression and survival rate in various cancer types. **(A)** Association heatmap showing the correlation of R-loop gene expression and overall survival (OS) in TCGA tumours. Statistical significance ($p < 0.05$) is highlighted in green. The number of significant associations is 179. **(B)** Histogram showing the distribution of survival associations across the studied R-loop genes. The ERBB2 (HER2) oncogene and the TP53 tumour suppressor gene are known positive controls regarding cancer association (highlighted in blue), having no established R-loop function.

carcinoma (ESCA), Glioblastoma multiforme (GBM), Head and Neck squamous cell carcinoma (HNSC), Kidney Chromophobe (KICH), Kidney renal clear cell carcinoma (KIRC), Kidney renal papillary cell carcinoma (KIRP), Acute Myeloid Leukaemia (LAML), Brain Lower Grade Glioma (LGG), Liver hepatocellular carcinoma (LIHC), Lung adenocarcinoma (LUAD), Lung squamous cell carcinoma (LUSC), Mesothelioma (MESO), Ovarian serous cystadenocarcinoma (OV), Pancreatic adenocarcinoma (PAAD), Pheochromocytoma and Paraganglioma (PCPG), Prostate adenocarcinoma (PRAD), Rectum adenocarcinoma (READ), Sarcoma (SARC), Skin Cutaneous Melanoma (SKCM), Stomach adenocarcinoma (STAD), Testicular Germ Cell Tumours (TGCT), Thyroid carcinoma (THCA), Thymoma (THYM), Uterine Corpus Endometrial Carcinoma (UCEC), Uterine Carcinosarcoma (UCS), and Uveal Melanoma (UVM).

2.2. Gene expression and survival data analysis

Gene expression data (RNA-sequencing) from the tumour samples

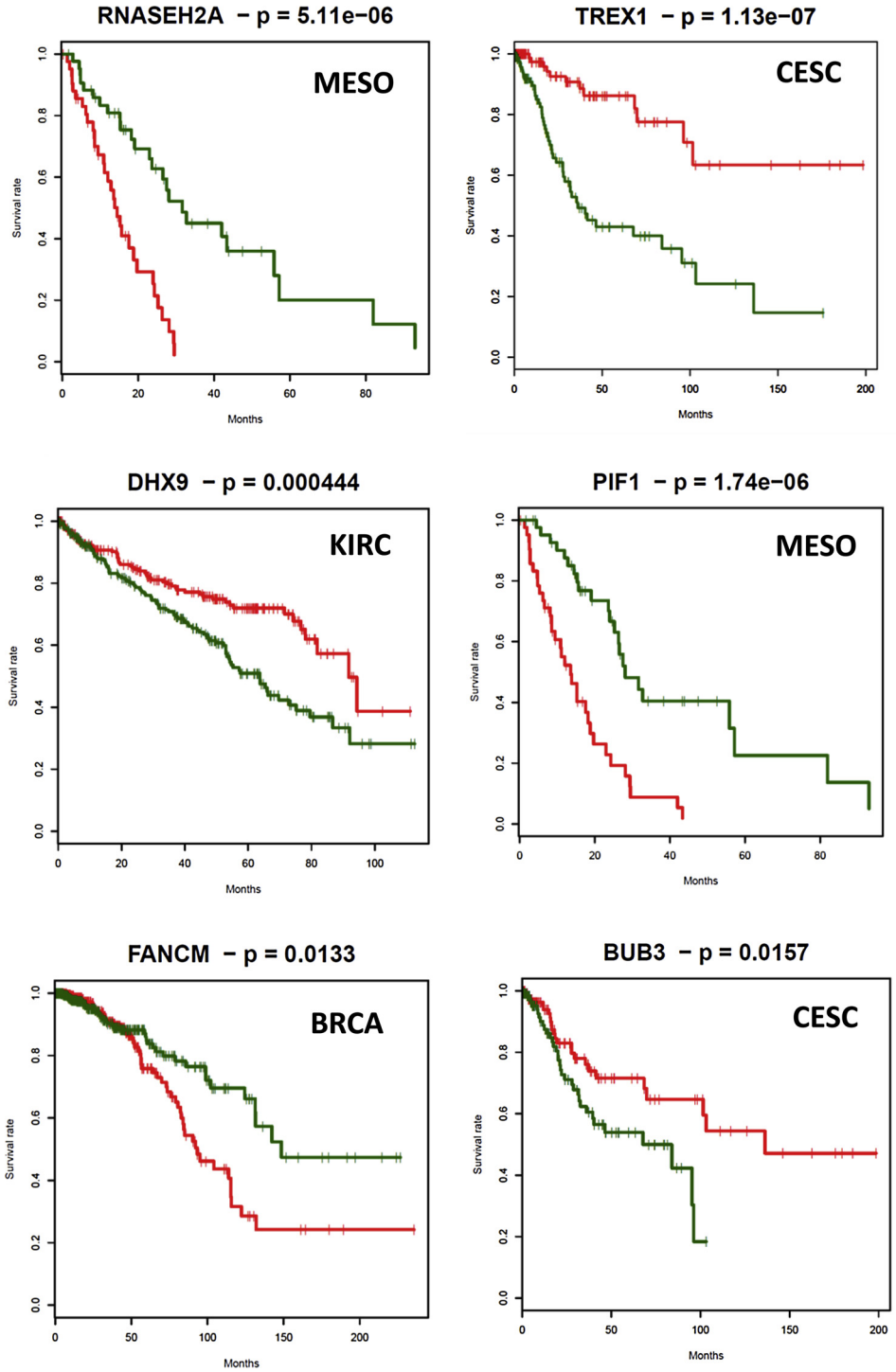


Fig. 3. Representative survival plots showing the overall survival (OS) of high expressor (red) and low expressor (green) groups in various cancer types. R-loop genes, cancer IDs, and p-values are indicated. The p-values from Kaplan-Meier analyses of all R-loop genes can be found in Table S1_3.

and overall survival time data were downloaded from TCGA project [43,44]. Healthy tissue gene expressions were obtained from the website of the Genotype-Tissue Expression (GTEx) project [45]. Survival analysis was performed and Kaplan-Meier plots were made using the "survival" package of the R software. The level of significance was $p < 0.05$. All p-values and adjusted p-values - corrected for multiple testing using the Benjamini-Hochberg method - are included in Table S1.

2.3. Cell line analyses

IC50 values of cancer cell lines upon treatment with 276 anticancer drugs were obtained from the Genomics of Drug Sensitivity in Cancer (GDSC) database as described and analysed previously [46]. We also downloaded additional cell line and drug information from the web page of the GDSC project [47]. Drug sensitivity and gene expression associations of cancer cell lines were presented as the Spearman correlation of IC50 values and mRNA expression scores of the R-loop genes analysed. We considered both negative and positive correlations with p-

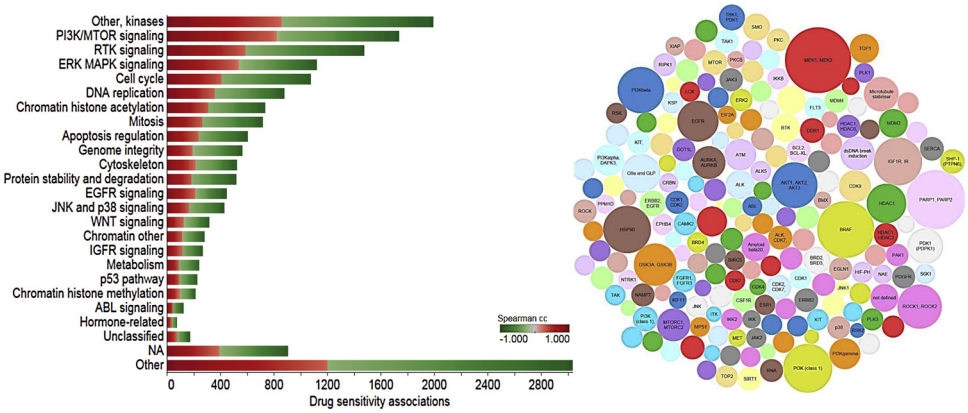


Fig. 4. Drug sensitivity associations of R-loop genes grouped by molecular pathways and drug targets. *Left:* Distribution of drug interactions over molecular pathways. Green and red colours indicate higher or lower drug sensitivity, respectively, associated with the mRNA expression levels of R-loop genes. The level of significance is $p < 0.05$. The number of significant associations is 22,414. *Right:* Statistical representation of drug targets showing significant drug sensitivity interactions with the R-loop genes analysed. Circle sizes are proportional to drug target frequencies.

values < 0.05. All p-values and adjusted p-values using Benjamini-Hochberg correction are included in Table S1.

2.4. RNA-DNA hybrid binding proteins

RNA-DNA hybrid binding proteins of various categories having clear R-loop function were selected for detailed analyses: AQR [48,49], ATXN1/2 [50–52], BLM [53], BRCA1/2 [54,55], BUB3 [36], BUGZ (ZNF207) [36], CARM1 [32], DDX19A [56], DHX9 [21,35], EWSR [31], FANCD2 [57,58], FANCM [58,59], GADD45A [60], PIF1 [61,62], PRMT1 [32], RNASEH1 [63,64], RNASEH2 [65,66], RTEL1 [17], SETX [48,67,68], SRSF1 [69], SRSF2 [37], THO/TREX [70], TOP1 [71–74], TOP3B [32], U2AF1 [37].

3. Results and discussion

To examine the relationship between R-loop genes and cancer, we first determined how many of the recently identified RNA-DNA hybrid binding proteins [20] are represented in the registry of cancer genes [75] representing the largest repository of 2372 genes whose somatic mutations have cancer driver roles. Of the 448 R-loop genes examined, 92 were cancer genes (**Table S1_1**) showing statistically significant enrichment ($p = 0.005792$; two prop z test). Next, we selected a core set of R-loop genes for a detailed analysis of tumorigenesis (n = 36), representing prominent molecular pathways with clear R-loop association that are also implicated in the formation of cancer (e.g. DNA topology; RNA-DNA hybrid ribonucleases and helicases; RNA processing, splicing and export; DNA damage; chromatin modifications; **Table S1_2**). The mRNA expression of these R-loop genes was extracted from the Cancer Genome Atlas (TCGA) database, allowing us to identify differences in the gene expression signatures of healthy tissues and primary tumours (Fig. 1). For instance, expression of BUB3, DHX9, PRMT1, THOC4, THOC7, U2AF1 and ZNF207 (BUGZ) was increased in several primary tumours compared to normal tissues, while SRSF1 (ASF/SF2) was downregulated in most cancers (except for acute myeloid leukaemia; LAML). Next, we asked whether R-loop gene expression levels correlate with survival rate of cancer patients. Taking advantage of gene expression (RNA-seq) and overall survival (OS) data from the TCGA, we generated 12,862 Kaplan-Meier survival curves in our analysis and identified numerous cases showing significant survival association with R-loop gene expression levels (179 at $p < 0.05$; 33 at FDR < 0.05; Fig. 2**, Supplementary** Fig. S1, and **Table S1_3**). In 70% of cases (123 at $p < 0.05$), low expression of R-loop genes was found to be associated with prolonged survival of cancer patients. Low expression of RNASEH2A, THOC6, PRMT1, and PIF1, for instance, significantly lengthened survival time in mesothelioma (MESO), while low FANCM mRNA level was advantageous for breast cancer survival (BRCA; (representative survival plots of low- and high-expressor groups are shown in Fig. 3). In 30% of cases (52 at $p < 0.05$), high expression of R-loop genes was detected and associated with better survival; for example, TREX1 and BUB3 were beneficial for cervical squamous cell carcinoma and endocervical adenocarcinoma survival (CESC; Fig. 3 and **Table S1_3**). In the case of 10 R-loop genes, long-term survival was observed exclusively in the low-expressor group of patients, irrespective of cancer type (ATXN2 (1), BRCA2 (5), CARM1 (6), DDX19A (2), RNASEH1 (6), THOC2 (4), THOC3 (4), TOP1 (5), U2AF1 (2), ZNF207 (3). For the remaining 26 R-loop genes, high or low expressing groups varied by cancer type (**Supplementary** Fig. S2). Most survival associations were observed for adrenocortical carcinoma (19 genes) and mesothelioma (18 genes) (Fig. 2B) correlated with low expression of R-loop genes (**Table S1_3**). According to the number of survival associations (Fig. 2B), RNASEH2A, BLM, BRCA1, BUB3, and PIF1 appears to be generally important for survival in multiple cancers, while the effects of ATXN2, DDX19A and U2AF1 are limited to specific types of cancer.

To show that our approach can identify real relationships, we included an oncogene (ERBB2 - HER2) and a tumour suppressor (TP53) in our analysis that are known positive controls for cancer association, with no established R-loop function (Fig. 2, highlighted in blue). R-loop gene expressions were associated with survival in a similar number of cancers (or more in a few cases) as ERBB2 and TP53, which demonstrates the relevance of our observations on RNA-DNA hybrid binding proteins.

In order to identify R-loop genes that might serve as potential drug targets for therapeutic intervention, we sought to find synergistic interactions between the mRNA expression status of R-loop genes in a large collection of cancer cell lines and sensitivity to chemotherapeutics approved by the US Food and Drug Administration (FDA). Pharmacogenomic (RNA-seq and drug sensitivity) data for 267 anticancer compounds across 1065 cancer cell lines were extracted from the GDSC database [47], covering a wide range of biological pathways including protein kinase signalling, cytoskeleton, DNA replication, DNA repair, and cell cycle control. The total number of significant interactions showed considerable variations in the number of drug-specific associations (22,414 at $p < 0.05$; 508 at FDR < 0.05; range: 70–230 interactions per drug; **Supplementary** Fig. S3 and **Table S1_4**) and in the molecular pathways/drug targets involved (Fig. 4). Highly represented pathways included the Ser/Thr protein kinase pathway and PI3K/RTK/MAPK signalling (Fig. 4**, left panel**), while the most common (top10) drug targets were MEK1/2, BRAF, PARP1/2, HSP90, AKT1/3, GSKs, PI3Ks, IGF1R, ROCK1/3, and EGFR (Fig. 4**, right panel**). We found significant drug sensitivity associations in 80% of the studied R-loop genes (29 at $p < 0.05$; except for BRCA2, BUGZ, DDX19, RNASEH1, RTEL1, THOC3, and THOC4; **Table S1_4**), of which CARM1, EWSR1, DHX9, and THOC1 showed the highest number of drug interactions (Fig. 5; **Table S1_4**). This highlights the importance of considering R-loop gene mRNA expression levels as these may affect drug response. However, we observed significant variability in the number of drug interactions between various cancer cells lines. For

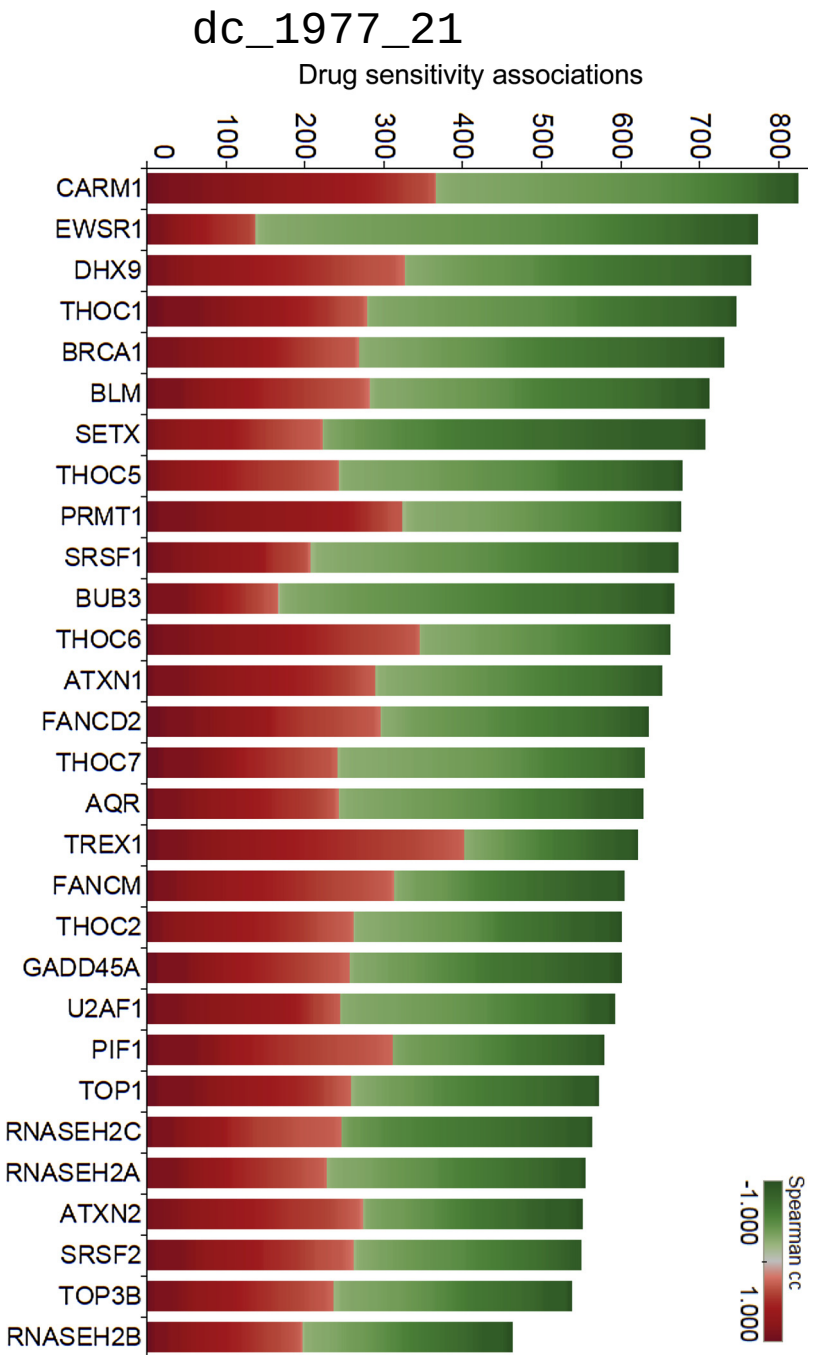


Fig. 5. Drug sensitivity associations of R-loop genes in various cancer cell lines. Green and red colours show significant negative/positive correlations between the expression of R-loop genes and drug sensitivities, indicating higher drug sensitivity or resistance of tumour cells against cancer therapeutic drugs. The level of significance is $p < 0.05$. The number of significant associations is 22,414.

example, lung small cell carcinoma and ovarian cancer cells were sensitive to most of the compounds tested, but B-cell leukaemia, Hodgkin's lymphoma, head and neck cancer, and Ewing sarcoma cells exhibited significantly lower drug efficacy (**Supplementary** Fig. S4, **green bars**). These differences indicate that RNA-DNA hybrid binding proteins may have contrasting outcomes in responding to various chemotherapeutic treatments depending on genomic context/cancer type. For example, in many cancer cell lines, high expression of TREX was associated with high IC_{50} values (half maximal inhibitory concentration) for most drug treatments indicating worse efficacy (i.e. higher drug concentration achieving half-maximal response) when TREX is overexpressed. This is entirely consistent with a recent study demonstrating that reducing the level of TREX1 leads to improved sensitivity of glioma and melanoma cells to the anticancer drugs topotecan, nimustine, and fotemustine [76]. Chondrosarcoma, lymphoblastic T cell leukaemia, biliary tract cancer, breast cancer, and pancreatic cancer cell lines showed exceptionally high IC_{50} values (resistance) to drug treatments in the case of increased BUB3 expression. In contrast, significant negative correlations were observed between BUB3, EWSR1, and SETX expression levels and IC_{50} values in most (but not all) cancer cell lines, i.e., overexpression of these R-loop genes typically made cancer cells more sensitive to drug treatment. These results collectively indicate that targeting TREX, BUB3, SETX, and EWSR to reduce or increase their expression levels in the above tumours may help increase the efficacy of cancer chemotherapeutics.

Since cancer cell lines - deriving from natural tumours - recapitulate the genomic context and tissue type of primary cancers [47,77], we narrowed down the identified drug interactions to the fraction of R-loop genes that showed significant survival associations in the matching primary cancer (drug and survival data in non-matching cell types were

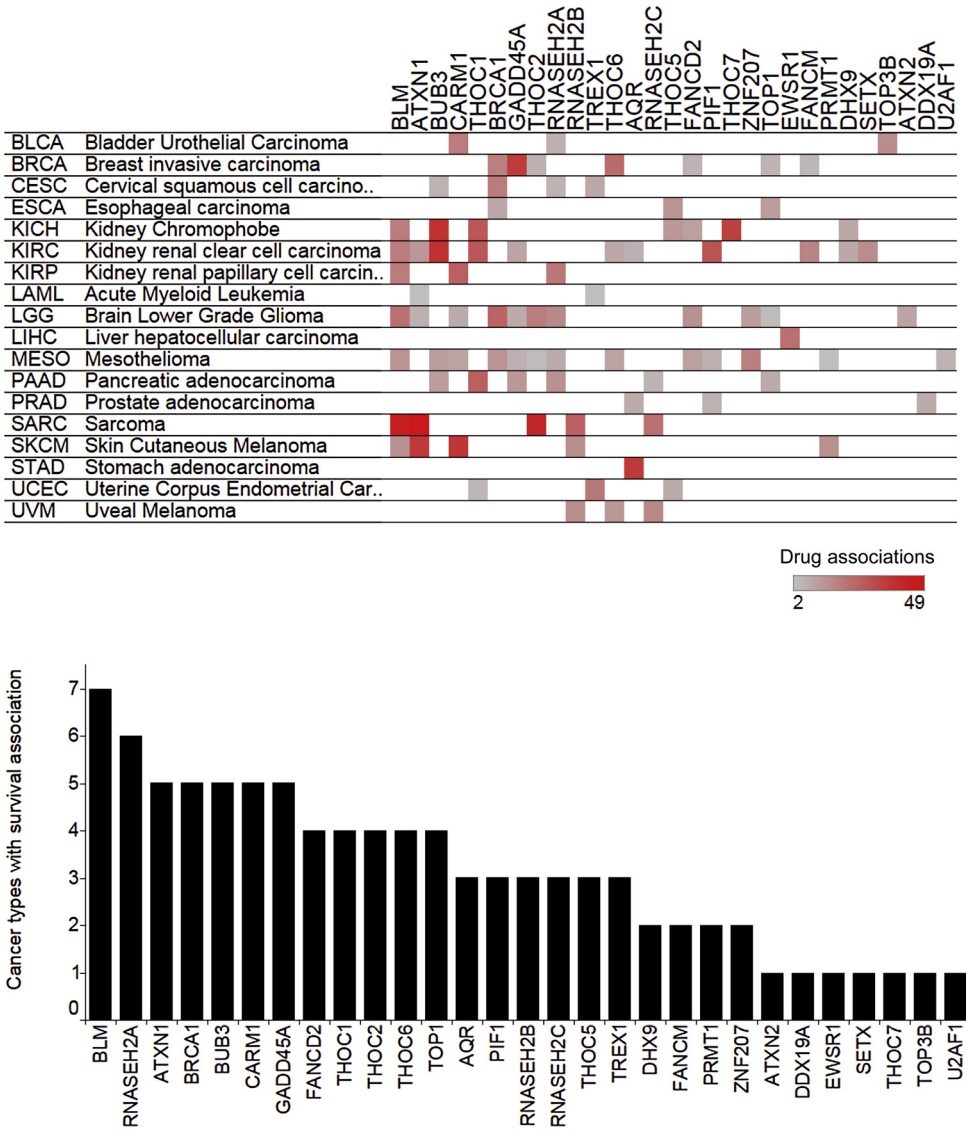


Fig. 6. Survival associations of R-loop genes in primary cancers that show significant drug sensitivity associations in the matching cancer cell line. The colour of survival association map (upper panel) highlights the number of drug interactions. Histogram (lower panel) shows the distribution of survival associations across the studied R-loop genes with significant drug interactions. The level of significance is $p < 0.05$. The number of significant associations is 1630.

excluded from further analysis; Fig. 6). Collectively, we identified 1,630 significant survival & drug associations (at $p < 0.05$) related to the expression of 29 R-loop genes (**Table S1_4**; seven R-loop genes were omitted because there was no survival or drug association, or the primary tumour could not be parsed with the corresponding tumour cell line). We observed the following trends:

i) most survival interactions (top10) were related to BLM, RNASEH2A, ATXN1, BRCA1, BUB3, CARM1, GADD45A, FANCD2, THOC1, and THOC2 associated with various cancers (Fig. 6).
ii) high expression of THOC2 in sarcoma cell lines was associated with a high IC_{50} value to CX-5461 (G4 ligand/RNA polymerase I inhibitor), suggesting that CX-5461 may be less effective for the treatment of sarcomas showing high THOC2 expression. On the other hand, high expression of PIF1, FANCD2 and BRCA1 was associated with low IC_{50} (sensitivity) to CX-5461 and low survival in mesothelioma (**Table S1_4**). Somewhat differently, low IC_{50} to CX-5461 was associated with high expression of ATXN1 and TREX1 and increased survival in melanoma and endometrial cancer, respectively (**Table S1_4**).
iii) BRCA1 expression in oesophageal carcinoma cell lines were associated with high IC_{50} (resistance) to 5-fluorouracil (5-FU), while low BRCA1 level was associated with better survival in oesophageal carcinoma. It follows that 5-FU treatment may be efficient for the therapy of BRCA1(-) oesophageal carcinomas. Similarly, BRCA1 expression in glioma cell lines was associated with reduced efficacy to (52)-7-oxozeaenol, GDC0941, refametimib, and selumetinib, while low BRCA1 level was beneficial for the survival of brain lower grade glioma. Moreover, mesothelioma cell lines are resistant to doxorubicin, or OSU-03012 or thapsigargin when expressing BRCA1. These drugs are potentially effective chemotherapeutics for BRCA(-) gliomas and mesotheliomas. Importantly, 5-FU is being tested in a clinical trial related to oesophageal carcinoma (NCT00052910), selumetinib is being investigated in low-grade glioma (NCT01089101), and doxorubicin is evaluated in patients with mesothelioma (NCT00634205).
iv) Most drug/gene associations were observed for RDEA119, selumetinib, and olaparib with most cell lines showing resistance to RDEA119 and selumetinib, and sensitivity to olaparib associated showing high R-loop gene expression levels (**Supplementary** Fig. S3). For instance, Selumetinib treatment was ineffective in stomach adenocarcinoma cell lines overexpressing AQR (NCT02448290 -

Phase 2), in sarcoma cell lines overexpressing ATXN1 or BLM (NCT03155620 - Phase 2, NCT01752569 - Phase 2), in glioma cell lines overexpressing BRCA1 or RNASEH2A, in kidney cancer cell lines overexpressing FANCM or RNASEH2A, and in endometrial carcinoma cell lines overexpressing TREX1. In terms of patient survival, low expression of AQR, ATXN1, BLM, BRCA1 and RNASEH2A was associated with longer survival in the matching primary cancers, except for FANCM and TREX1 where higher mRNA level was beneficial for kidney renal clear cell carcinoma and endometrial carcinoma patients, respectively. Olaparib treatment was effective in prostate adenocarcinoma cells showing high DDX19A expression, and in the breast cancer cells overexpressing GADD45A, but mesothelioma cell lines with high ZNF207 (BUGZ) expression turned out to be resistant to the treatment. Olaparib is in clinical trial for patients with mesothelioma (NCT03531840 - Phase 2), prostate adenocarcinoma (NCT01682772 - Phase 2) and breast cancer (NCT02000622 - Phase 3).

v) Several drugs that are now in clinical studies to treat breast cancer show low IC_{50} values in breast cancer cell lines with high BRCA1 gene expression. For example, rucaparib (NCT02505048 - Phase 2, BRCA1,2 mutant patients), dabrafenib (NCT02465060 - Phase 2, BRAF V600E/R/K/D mutant patients), vismodegib (NCT02465060 - Phase 2 - SMO or PTCH1 mutant patients; NCT02694224 - Phase 2), BMS-754807 (NCT01225172 - Phase 2), and ruxolitimib (NCT01562873 - Phase 2, NCT02928978 - Phase 2).

vi) Oesophageal carcinoma cell lines with high BRCA1 gene expression were associated with low IC_{50} values to motesamib, cisplatin and trametinib treatments, which are now being tested in various clinical trials (NCT00101907 - Phase 1, NCT00655876 - Phase 3, NCT02465060 - Phase 2).

vii) In glioma cell lines, low CARM1 expression was associated with better efficacy to SN38 treatment (type I topoisomerase inhibitor, active metabolite of irinotecan), while glioma patients with low CARM1 levels showed longer survival. It is possible that CARM1(-) cancers are more susceptible to SN38 drug treatment leading to inhibition of Top3B and Top1, which could in turn increase R-loop levels [32] and induce cell death.

The above pharmacogenomic associations of human RNA-DNA hybrid binding proteins support the role of R-loop genes in tumorigenesis and in determining the efficiency of tumour therapies. We note that, however, our analytical approach represents a hypothesis generating exercise performed on multi-experiment observations to identify all possible associations of R-loop genes and human malignancies, which should be experimentally validated in further high-content analysis projects. Our data suggest that modulating the expression levels of R-loop genes may affect clinical responses to anticancer drug treatments, and these expression changes could be used to define patient groups who are most likely to benefit from a therapy. Based on the revealed pharmacogenomic interactions, we propose that R-loops and R-loop binding proteins are potentially relevant new epigenetic markers and therapeutic targets in multiple cancers. Further exploration of the recognised associations is expected to improve drug effectiveness and identify potential combination therapeutics.

Funding

L. Székvölgyi received grants from the Hungarian Academy of Sciences (Lendület program, grant LP2015-9/2015), and the National Research, Development, and Innovation Office, Hungary (grant GINOP-2.3.2-15-2016-00024). This work was also supported by the GINOP-2.3.2-15-2016-00043 and the Higher Education Institutional Excellence Programme of the Ministry of Human Capacities in Hungary, within the framework of the Biotechnology Thematic Programme of the University of Debrecen 20428-3/2018/FEKUTSTRAT (to G. Halmos), and GINOP-2.3.2-15-2016-00062 (to Z. Szabó).

Appendix A. Supplementary data

Supplementary material related to this article can be found, in the online version, at doi:https://doi.org/10.1016/j.dnarep.2019.102642.

References

[1] M.P. Crossley, M. Bocek, K.A. Cimprich, R-loops as cellular regulators and genomic threats, Mol. Cell 73 (2019) 398–411, https://doi.org/10.1016/j.molcel.2019.01.024.
[2] P. Richard, J.L. Manley, R loops and links to human disease, J. Mol. Biol. (2016), https://doi.org/10.1016/j.jmb.2016.08.031.
[3] K. Skourti-Stathaki, N.J. Proudfoot, A double-edged sword: R loops as threats to genome integrity and powerful regulators of gene expression, Genes Dev. 28 (2014) 1384–1396, https://doi.org/10.1101/gad.242990.114.Freely.
[4] A. Aguilera, T. García-Muse, R loops: from transcription byproducts to threats to genome stability, Mol. Cell 46 (2012) 115–124, https://doi.org/10.1016/j.molcel.2012.04.009.
[5] P. Kotsantis, L.M. Silva, S. Irmscher, R.M. Jones, L. Folkes, N. Gromak, E. Petermann, Increased global transcription activity as a mechanism of replication stress in cancer, Nat. Commun. 7 (2016) 13087, https://doi.org/10.1038/ncomms13087.
[6] B. Vogelstein, N. Papadopoulos, V.E. Velculescu, S. Zhou, L.A. Diaz Jr., K.W. Kinzler, Cancer Genome Landscapes, Science 339 (80-) (2013) 1546–1558, https://doi.org/10.1126/science.1235122.
[7] R.M. Webster, Combination therapies in oncology, Nat. Rev. Drug Discov. 15 (2016) 81–82, https://doi.org/10.1038/nrd.2016.3.
[8] M. Gianni, Y. Qin, G. Wenes, B. Bandstra, A.P. Conley, V. Subbiah, R. Leibowitz-Amit, S. Ekmekcioglu, E.A. Grimm, J. Roszik, High-throughput architecture for discovering combination Cancer therapeutics, JCO Clin Cancer Inform. 2 (2018) 1–12, https://doi.org/10.1200/CCI.17.00054.
[9] L.A. Sanz, S.R. Hartono, Y.W. Lim, P.A. Ginno, L.A. Sanz, S.R. Hartono, Y.W. Lim, S. Steyaert, A. Rajpurkar, P.A. Ginno, X. Xu, F. Chédin, Prevalent, Dynamic, and Conserved R-Loop Structures Associate with Specific Epigenomic Signatures in Mammals, Mol. Cell 63 (2016) 167–178, https://doi.org/10.1016/j.molcel.2016.05.032.
[10] L. Halász, Z. Karányi, B. Boros-oláh, T. Kuik-rózsa, É. Sipos, É. Nagy, Á. Mosolygó-l, A. Mázló, É. Rajnavölgyi, G. Halmos, L. Székvölgyi, RNA-DNA hybrid (R-loop) immunoprecipitation mapping: an analytical workflow to evaluate inherent biases, Genome Res. 27 (2017) 1063–1073, https://doi.org/10.1101/gr.219394.
[11] S. Hamperl, K.A. Cimprich, The contribution of co-transcriptional RNA: DNA hybrid structures to DNA damage and genome instability, DNA Repair (Amst). 19 (2014) 84–94, https://doi.org/10.1016/j.dnarep.2014.03.023.
[12] N.N. Shaw, D.P. Arya, Recognition of the unique structure of DNA:RNA hybrids, Biochimie. 90 (2008) 1026–1039, https://doi.org/10.1016/j.biochi.2008.04.011.
[13] F. Takusagawa, K.T. Takusagawa, R.G. Carlson, R.F. Weaver, Selectivity of F8-Actinomycin D for RNA:DNA hybrids and its anti-leukemia activity, Bioorg. Med. Chem. Lett. 5 (1997) 1197–1207, https://doi.org/10.1016/S0968-0896(97)00062-X.
[14] S.Q. Mao, A.T. Ghanbarian, J. Spiegel, S. Martínez Cuesta, D. Beraldi, M. Di Antonio, G. Marsico, R. Hänsel-Hertsch, D. Tannahill, S. Balasubramanian, DNA G-quadruplex structures mold the DNA methylome, Nat. Struct. Mol. Biol. 25 (2018) 951–957, https://doi.org/10.1038/s41594-018-0131-8.
[15] A.J. Angelbello, J.L. Chen, J.L. Childs-Disney, P. Zhang, Z.F. Wang, M.D. Disney, Using genome sequence to enable the design of medicines and chemical probes, Chem. Rev. 118 (2018) 1599–1663, https://doi.org/10.1021/acs.chemrev.7b00504.
[16] H. Xu, M. Di Antonio, S. McKinney, V. Mathew, B. Ho, N.J. O'Neil, N. Dos Santos, J. Silvester, V. Wei, J. Garcia, F. Kabeer, D. Lai, P. Soriano, J. Banáth, D.S. Chiu, D. Yap, D.D. Le, F.B. Ye, A. Zhang, K. Thu, J. Soong, S.C. Lin, A.H.C. Tsai, T. Osako, T. Algara, D.N. Saunders, J. Wong, J. Xian, M.B. Bally, J.D. Brenton, G.W. Brown, S.P. Shah, D. Cescon, T.W. Mak, C. Caldas, P.C. Stirling, P. Hieter, S. Balasubramanian, S. Aparicio, CX-5461 is a DNA G-quadruplex stabilizer with selective lethality in BRCA1/2 deficient tumours, Nat. Commun. 8 (2017), https://doi.org/10.1038/ncomms14432.
[17] J.B. Vannier, V. Pavicic-Kaltenbrunner, M.I.R. Petalcorin, H. Ding, S.J. Boulton, RTEL1 dismantles T loops and counteracts telomeric G4-DNA to maintain telomere integrity, Cell. 149 (2012) 795–806, https://doi.org/10.1016/j.cell.2012.03.030.
[18] A. De Magis, S.G. Manzo, M. Russo, J. Marinello, O. Sordet, R. Morigi Capranico, DNA damage and genome instability by G-quadruplex ligands are mediated by R-loops in human cancer cells, Proc. Natl. Acad. Sci. 116 (2018) 816–825, https://doi.org/10.1073/pnas.1810409116.
[19] E. Ruggiero, S.N. Richter, G-quadruplexes and G-quadruplex ligands: targets and tools in antiviral therapy, Nucleic Acids Res. 46 (2018) 3270–3283, https://doi.org/10.1093/nar/gky187.
[20] I.X. Wang, C. Grunseich, J. Fox, J. Burdick, Z. Zhu, N. Ravazian, M. Hafner, V.G. Cheung, Human proteins that interact with RNA / DNA hybrids, Genome Res. (2018) 1–10, https://doi.org/10.1101/gr.237362.118.
[21] A. Cristini, M. Groh, M.S. Kristiansen, N. Gromak, RNA/DNA hybrid interactome identifies DXH9 as a molecular player in transcriptional termination and R-Loop-Associated DNA damage, Cell Rep. 23 (2018) 1891–1905, https://doi.org/10.1016/j.celrep.2018.04.025.
[22] S.C. Shuck, J.J. Turchi, Targeted inhibition of replication protein A reveals cytotoxic activity, synergy with chemotherapeutic DNA-damaging agents, and insight

into cellular function, Cancer Res. 70 (2010) 3789–3798, https://doi.org/10.1158/0008-5472.CAN-09-3422.
[23] T.Z. Minas, J. Han, T. Javaheri, S.-H. Hong, M. Schlederer, Y. Saygideğer-Kont, H. Çelik, K.M. Mueller, I. Temel, M. Özdemirli, H. Kovar, H.V. Erkizan, J. Toretsky, L. Kenner, R. Moriggl, A. Üren, YK-4-279 effectively antagonizes EWS-FLI1 induced leukemia in a transgenic mouse model, Oncotarget. 6 (2015) 37678–37694, https://doi.org/10.18632/oncotarget.5520.
[24] D. Colak, N. Zaninovic, M.S. Cohen, Z. Rosenwaks, W. Yang, Promoter-bound trinucleotide repeat mRNA drives epigenetic silencing in fragile X syndrome, Science (80) (2014) 1002–1005, https://doi.org/10.1126/science.1245831.
[25] W.T. Powell, R.L. Coulson, M.L. Gonzales, F.K. Crary, S.S. Wong, S. Adams, R. a Ach, P. Tsang, N.A. Yamada, D.H. Yasui, F. Chédin, J.M. LaSalle, R-loop formation at Snord116 mediates topotecan inhibition of Ube3a-antisense and allele-specific chromatin decondensation, Proc. Natl. Acad. Sci. U. S. A. 110 (2013) 13938–13943, https://doi.org/10.1073/pnas.1305426110.
[26] S.E. Jones, E.D.G. Fleuren, J. Frankum, A. Konde, C.T. Williamson, D.B. Krastev, H.N. Pemberton, J. Campbell, A. Gulati, R. Elliott, M. Menon, J.L. Selfe, R. Brough, S.J. Pettitt, W. Niedzwiedz, W.T.A. Van Der Graaf, J. Shipley, A. Ashworth, C.J. Lord, ATR is a therapeutic target in synovial sarcoma, Cancer Res. 77 (2017) 7014–7026, https://doi.org/10.1158/0008-5472.CAN-17-2056.
[27] M. Nieto-Soler, I. Morgado-Palacin, V. Lafarga, E. Lecona, M. Murga, E. Callen, D. Azorin, J. Alonso, A.J. Lopez-Contreras, A. Nussenzweig, O. Fernandez-Capetillo, Efficacy of ATR inhibitors as single agents in Ewing sarcoma, Oncotarget. 7 (2016) 58759–58767, https://doi.org/10.18632/oncotarget.11643.
[28] E. Tumini, E. Herrera-Moyano, M. San Martín-Alonso, S. Barroso, C.M. Galmarini, A. Aguilera, The antitumor drugs trabectedin and lurbinectedin induce transcription-dependent replication stress and genome instability, Mol. Cancer Res. (2018), https://doi.org/10.1158/1541-7786.MCR-18-0575 molcanres.0575.2018.
[29] J.A. Casado, P. Rio, E. Marco, V. Garcia-Hernandez, A. Domingo, L. Perez, J.C. Tercero, J.J. Vaquero, B. Albella, F. Gago, J.A. Bueren, Relevance of the Fanconi anemia pathway in the response of human cells to trabectedin, Mol. Cancer Ther. 7 (2008) 1309–1318, https://doi.org/10.1158/1535-7163.MCT-07-2432.
[30] B.J. Monk, D. Lorusso, A. Italiano, S.B. Kaye, M. Aracil, A. Tanović, M. D'Incalci, Trabectedin as a chemotherapy option for patients with BRCA deficiency, Cancer treat. Rev. 50 (2016) 175–182, https://doi.org/10.1016/j.ctrv.2016.09.009.
[31] A. Gorthi, J.C. Romero, E. Loranc, L. Cao, L.A. Lawrence, E. Goodale, A.B. Iniguez, X. Bernard, V.P. Masamsetti, S. Roston, E.R. Lawlor, J.A. Toretsky, K. Stegmaier, S.L. Lessnick, Y. Chen, A.J.R. Bishop, EWS-FLI1 increases transcription to cause R-Loops and block BRCA1 repair in Ewing sarcoma, Nature. 555 (2018) 387–391, https://doi.org/10.1038/nature25748.
[32] Y. Yang, K.M. McBride, S. Hensley, Y. Lu, F. Chedin, M.T. Bedford, Arginine methylation facilitates the recruitment of TOP3B to chromatin to prevent r loop accumulation, Mol. Cell 53 (2014) 484–497, https://doi.org/10.1016/j.molcel.2014.01.011.
[33] A.E. Drew, O. Moradei, S.L. Jacques, N. Rioux, A.P. Boriack-Sjodin, C. Allain, M.P. Scott, L. Jin, A. Raimondi, J.L. Handler, H.M. Ott, R.G. Kruger, M.T. McCabe, C. Sneeringer, T. Riera, G. Shapiro, N.J. Waters, L.H. Mitchell, K.W. Duncan, M.P. Moyer, R.A. Copeland, J. Smith, R. Chesworth, S.A. Ribich, Identification of a CARM1 inhibitor with potent in vitro and in vivo activity in preclinical models of multiple myeloma, Sci. Rep. 7 (2017) 17993, https://doi.org/10.1038/s41598-017-18446-z.
[34] J. Marinello, S. Bertoncini, I. Aloisi, A. Cristini, G.M. Tagliazucchi, M. Forcato, O. Sordet, G. Capranico, Dynamic effects of topoisomerase i inhibition on R-loops and short transcripts at active promoters, PLoS One 11 (2016) 1–18, https://doi.org/10.1371/journal.pone.0147053.
[35] P. Chakraborty, J.T.J. Huang, K. Hiom, DHX9 helicase promotes R-loop formation in cells with impaired RNA splicing, Nat. Commun. 9 (2018) 4346, https://doi.org/10.1038/s41467-018-06677-1.
[36] Y. Wan, X. Zheng, H. Chen, Y. Guo, H. Jiang, X. He, X. Zhu, Y. Zheng, Splicing function of mitotic regulators links R-loop-mediated DNA damage to tumor cell killing, J. Cell Biol. 209 (2015) 235–246, https://doi.org/10.1083/jcb.201409073.
[37] L. Chen, J.Y. Chen, Y.J. Huang, Y. Gu, J. Qiu, H. Qian, C. Shao, X. Zhang, J. Hu, H. Li, S. He, Y. Zhou, O. Abdel-Wahab, D.E. Zhang, X.D. Fu, The augmented R-Loop is a unifying mechanism for myelodysplastic syndromes induced by high-risk splicing factor mutations, Mol. Cell 69 (2018) 412–425, https://doi.org/10.1016/j.molcel.2017.12.029 e6.
[38] S. Sorrells, S. Nik, M. Casey, R.C. Cameron, H. Truong, C. Toruno, M. Gulfo, A. Lowe, C. Jette, R.A. Stewart, T.V. Bowman, Spliceosomal components protect embryonic neurons from R-loop-mediated DNA damage and apoptosis, Dis. Model. Mech. 11 (2018), https://doi.org/10.1242/dmm.031583 pii: dmm031583.
[39] S.L.W. Tan, S. Chadha, Y. Liu, E. Gabasova, D. Perera, K. Ahmed, S. Constantinou, X. Renaudin, M. Lee, R. Aebersold, A.R. Venkitaraman, A class of environmental and endogenous toxins induces BRCA2 haploinsufficiency and genome instability, Cell. 169 (2017) 1105–1118, https://doi.org/10.1016/j.cell.2017.05.010 e15.
[40] J.M. Flanagan, J.M. Funes, S. Henderson, L. Wild, N. Carey, C. Boshoff, Genomics screen in transformed stem cells reveals RNASEH2A, PPAP2C, and ADARB1 as putative anticancer drug targets, Mol. Cancer Ther. 8 (2009) 249–260, https://doi.org/10.1158/1535-7163.MCT-08-0636.
[41] R. White, B. Saxty, J. Large, C.A. Kettleborough, A.P. Jackson, Identification of small-molecule inhibitors of the ribonuclease H2 enzyme, J. Biomol. Screen. 18 (2013) 610–620, https://doi.org/10.1177/1087057113476550.
[42] T. Uehara, Y. Minoshima, K. Sagane, N.H. Sugi, K.O. Mitsuhashi, N. Yamamoto, H. Kamiyama, K. Takahashi, Y. Kotake, M. Uesugi, A. Yokoi, A. Inoue, T. Yoshida, M. Mabuchi, A. Tanaka, T. Owa, Selective degradation of splicing factor CAPERα by anticancer sulfonamides, Nat. Chem. Biol. 13 (2017) 675–680, https://doi.org/10.1038/nchembio.2363.
[43] The Cancer genome atlas research network, J.N. Weinstein, Eric A. Collisson, Gordon B. Mills, Kenna R.Mills Shaw, Chris Sander, K. Ellrott, Joshua M. Stuart, B.A. Ozenberger, I. Shmulevich, The cancer genome atlas pan-cancer analysis project, Nat. Genet. 45 (2013) 1113–1120, https://doi.org/10.1038/ng.2764.
[44] R.L. Grossman, A. Heath, M. Murphy, Toward data science as a service, Comput. Sci. Eng. 18 (2016) 10–20, https://doi.org/10.1109/MCSE.2016.92.
[45] B.E. Stranger, L.E. Brigham, R. Hasz, M. Hunter, C. Johns, M. Johnson, G. Kopen, W.F. Leinweber, J.T. Lonsdale, A. McDonald, B. Mestichelli, K. Myer, B. Roe, M. Salvatore, S. Shad, J.A. Thomas, G. Walters, M. Washington, J. Wheeler, J. Bridge, B.A. Foster, B.M. Gillard, E. Karasik, R. Kumar, M. Miklos, M.T. Moser, S.D. Jewell, R.G. Montroy, D.C. Rohrer, D.R. Valley, D.A. Davis, D.C. Mash, S.E. Gould, P. Guan, S. Koester, A.R. Little, C. Martin, H.M. Moore, A. Rao, J.P. Struewing, S. Volpi, K.D. Hansen, P.F. Hickey, L.F. Rizzardi, L. Hou, Y. Liu, B. Molinie, Y. Park, N. Rinaldi, L. Wang, N. Wittenberghe, M. Claussnitzer, E.T. Gelfand, Q. Li, S. Linder, R. Zhang, K.S. Smith, E.K. Tsang, L.S. Chen, K. Demanelis, J.A. Doherty, J. Farzana, M.G. Kibriya, L. Jiang, S. Lin, M. Wang, R. Jian, X. Li, J. Chan, D. Bates, M. Diegel, J. Halow, E. Haugen, A. Johnson, R. Kaul, K. Lee, M.T. Maurano, J. Nelson, F.J. Neri, R. Sandstrom, M.S. Fernando, C. Linke, M. Oliva, A. Skol, F. Wu, J.M. Akey, A.P. Feinberg, J.L. Billy, B.L. Pierce, J.A. Stamatoyannopoulos, H. Tang, K.G. Ardlie, M. Kellis, M.P. Snyder, S.B. Montgomery, Enhancing GTEx by bridging the gaps between genotype, gene expression, and disease, Nat. Genet. 49 (2017) 1664–1670, https://doi.org/10.1038/ng.3969.
[46] Y. Qin, A.P. Conley, E.A. Grimm, J. Roszik, A tool for discovering drug sensitivity and gene expression associations in cancer cells, PLoS One 12 (2017) 1–6, https://doi.org/10.1371/journal.pone.0176763.
[47] W. Yang, J. Soares, P. Greninger, E.J. Edelman, H. Lightfoot, S. Forbes, N. Bindal, D. Beare, J.A. Smith, I.R. Thompson, S. Ramaswamy, P.A. Futreal, D.A. Haber, M.R. Stratton, C. Benes, U. McDermott, M.J. Garnett, Genomics of Drug Sensitivity in Cancer (GDSC): a resource for therapeutic biomarker discovery in cancer cells, Nucleic Acids Res. 41 (2013) 955–961, https://doi.org/10.1093/nar/gks1111.
[48] J. Sollier, C.T. Stork, M.L. García-Rubio, R.D. Paulsen, A. Aguilera, K.A. Cimprich, Transcription-coupled nucleotide excision repair factors promote R-Loop-Induced genome instability, Mol. Cell 56 (2014) 777–785, https://doi.org/10.1016/j.molcel.2014.10.020.
[49] R. Sakasai, M. Isono, M. Wakasugi, M. Hashimoto, Y. Sunatani, T. Matsui, A. Shibata, T. Matsunaga, K. Iwabuchi, Aquarius is required for proper CtIP expression and homologous recombination repair, Sci. Rep. 7 (2017) 13808, https://doi.org/10.1038/s41598-017-13695-4.
[50] L.A. Ostrowski, A.C. Hall, K. Mekhail, Ataxin-2: from RNA control to human health and disease, Genes (Basel). 8 (2017) 2–21, https://doi.org/10.3390/genes8060157.
[51] J.S. Salvi, J.N.Y. Chan, K. Szafranski, T.T. Liu, J.D. Wu, J.B. Olsen, N. Khanam, B.P.K. Poon, A. Emili, K. Mekhail, Roles for Pbp1 and caloric restriction in genome and lifespan maintenance via suppression of RNA-DNA hybrids, Dev. Cell 30 (2014) 177–191, https://doi.org/10.1016/j.devcel.2014.05.013.
[52] K.J. Abraham, J.N.Y. Chan, J.S. Salvi, B. Ho, A. Hall, E. Vidya, R. Guo, S.A. Killackey, N. Liu, J.E. Lee, G.W. Brown, K. Mekhail, Intersection of calorie restriction and magnesium in the suppression of genome-destabilizing RNA–DNA hybrids, Nucleic Acids Res. 44 (2016), https://doi.org/10.1093/nar/gkw752 gkw752.
[53] E. Yun, C. Chang, C.A. Novoa, M.J. Aristizabal, Y. Coulombe, R. Segovia, R. Chaturvedi, Y. Shen, C. Keong, A.S. Tam, S.J.M. Jones, J.Y. Masson, M.S. Kobor, P.C. Stirling, RECQ-like helicases Sgs1 and BLM regulate R-loop-associated genome instability, J. Cell Biol. 216 (2017) 3991–4005, https://doi.org/10.1083/jcb.201703168.
[54] E. Hatchi, K. Skourti-Stathaki, S. Ventz, L. Pinello, A. Yen, K. Kamieniarz-Gdula, S. Dimitrov, S. Pathania, K.M. McKinney, M.L. Eaton, M. Kellis, S.J. Hill, G. Parmigiani, N.J. Proudfoot, D.M. Livingston, BRCA1 recruitment to transcriptional pause sites is required for R-loop-driven DNA damage repair, Mol. Cell 57 (2015) 636–647, https://doi.org/10.1016/j.molcel.2015.01.011.
[55] V. Bhatia, S.I. Barroso, M.L. García-Rubio, E. Tumini, E. Herrera-Moyano, A. Aguilera, BRCA2 prevents R-loop accumulation and associates with TREX-2 mRNA export factor PCID2, Nature. 511 (2014) 362–365, https://doi.org/10.1038/nature13374.
[56] D. Hodroj, B. Recolin, K. Serhal, S. Martinez, N. Tsanov, R. Abou Merhi, D. Maiorano, An ATR-dependent function for the Ddx19 RNA helicase in nuclear R-loop metabolism, EMBO J. 36 (2017) 1182–1198, https://doi.org/10.15252/embj.201695131.
[57] M.L. García-Rubio, C. Pérez-Calero, S.I. Barroso, E. Tumini, E. Herrera-Moyano, I.V. Rosado, A. Aguilera, The fanconi Anemia pathway protects genome integrity from R-loops, PLoS Genet. 11 (2015) 1–17, https://doi.org/10.1371/journal.pgen.1005674.
[58] Z. Liang, F. Liang, Y. Teng, X. Chen, J. Liu, S. Longerich, T. Rao, A.M. Green, N.B. Collins, Y. Xiong, L. Lan, P. Sung, G.M. Kupfer, Binding of FANCI-FANCD2 complex to RNA and R-Loops stimulates robust FANCD2 monoubiquitination, Cell Rep. 26 (2019) 564–572, https://doi.org/10.1016/J.CELREP.2018.12.084 e5.
[59] R.A. Schwab, J. Nieminuszczy, F. Shah, J. Langton, D. Lopez Martinez, C.C. Liang, M.A. Cohn, R.J. Gibbons, A.J. Deans, W. Niedzwiedz, The fanconi Anemia pathway maintains genome stability by coordinating replication and transcription, Mol. Cell 60 (2015) 351–361, https://doi.org/10.1016/j.molcel.2015.09.012.
[60] K. Arab, E. Karaulanov, M. Musheev, P. Trnka, A. Schäfer, I. Grummt, C. Niehrs, GADD45A binds R-loops and recruits TET1 to CpG island promoters, Nat. Genet. (2019), https://doi.org/10.1038/s41588-018-0306-6.
[61] P.L.T. Tran, T.J. Pohl, C.-F. Chen, A. Chan, S. Pott, V.A. Zakian, PIF1 family DNA helicases suppress R-loop mediated genome instability at tRNA genes, Nat. Commun. 8 (2017) 15025, https://doi.org/10.1038/ncomms15025.

[62] J.-B. Boulé, V. a Zakian, The yeast Pif1p DNA helicase preferentially unwinds RNA DNA substrates, Nucleic Acids Res. 35 (2007) 5809–5818, https://doi.org/10.1093/nar/gkm613.
[63] a Arudchandran, S. Cerritelli, S. Narimatsu, M. Itaya, D.Y. Shin, Y. Shimada, R.J. Crouch, The absence of ribonuclease H1 or H2 alters the sensitivity of Saccharomyces cerevisiae to hydroxyurea, caffeine and ethyl methanesulphonate: implications for roles of RNases H in DNA replication and repair, Genes Cells 5 (2000) 789–802.
[64] R.W. Maul, H. Chon, K. Sakhuja, S.M. Cerritelli, L.A. Gugliotti, P.J. Gearhart, R.J. Crouch, R-loop depletion by over-expressed RNase H1 in mouse B cells increases activation-induced deaminase access to the transcribed strand without altering frequency of isotype switching, J. Mol. Biol. (2017) 1–9, https://doi.org/10.1016/j.jmb.2016.12.020.
[65] R. Uehara, S.M. Cerritelli, N. Hasin, K. Sakhuja, M. London, J. Iranzo, H. Chon, A. Grinberg, R.J. Crouch, Two RNase H2 mutants with differential rNMP processing activity reveal a threshold of ribonucleotide tolerance for embryonic development, Cell Rep. 25 (2018) 1135–1145, https://doi.org/10.1016/j.celrep.2018.10.019 e5.
[66] K.J. Mackenzie, P. Carroll, L. Lettice, A. Revuelta, E. Abbondati, R.E. Rigby, B. Rabe, F. Kilanowski, G. Grimes, A. Fluteau, P.S. Devenney, R.E. Hill, M.A.M. Reijns, A.P. Jackson, Ribonuclease H2 mutations induce a cGAS / STING-dependent innate immune response, EMBO J. (2016) 1–14, https://doi.org/10.15252/embj.201593339.
[67] O.J. Becherel, A.J. Yeo, A. Stellati, E.Y.H. Heng, J. Luff, A.M. Suraweera, R. Woods, J. Fleming, D. Carrie, K. McKinney, X. Xu, C. Deng, M.F. Lavin, Senataxin plays an essential role with DNA damage response proteins in meiotic recombination and gene silencing, PLoS Genet. 9 (2013), https://doi.org/10.1371/journal.pgen.1003435.
[68] M. Groh, L.O. Albulescu, A. Cristini, N. Gromak, Senataxin: genome guardian at the interface of transcription and neurodegeneration, J. Mol. Biol. 429 (2017) 3181–3195, https://doi.org/10.1016/j.jmb.2016.10.021.
[69] X. Li, J.L. Manley, Inactivation of the SR protein splicing factor ASF/SF2 results in genomic instability, Cell. 122 (2005) 365–378, https://doi.org/10.1016/j.cell.2005.06.008.
[70] B. Gómez-González, M. García-Rubio, R. Bermejo, H. Gaillard, K. Shirahige, A. Marín, M. Foiani, A. Aguilera, Genome-wide function of THO/TREX in active genes prevents R-loop-dependent replication obstacles, EMBO J. 30 (2011) 3106–3119, https://doi.org/10.1038/emboj.2011.206.
[71] W. Shen, H. Sun, C.L. De Hoyos, J.K. Bailey, X.-H. Liang, S.T. Crooke, Dynamic nucleoplasmic and nucleolar localization of mammalian RNase H1 in response to RNAP I transcriptional R-loops, Nucleic Acids Res. (2017) 1–21, https://doi.org/10.1093/nar/gkx710.
[72] A. El Hage, S. Webb, A. Kerr, D. Tollervey, Genome-wide distribution of RNA-DNA hybrids identifies RNase H targets in tRNA genes, retrotransposons and mitochondria, PLoS Genet. 10 (2014), https://doi.org/10.1371/journal.pgen.1004716.
[73] É. Hegedüs, E. Kókai, P. Nánási, L. Imre, L. Halász, R. Jossé, Z. Antunovics, M.R. Webb, A. El Hage, Y. Pommier, L. Székvölgyi, V. Dombrádi, G. Szabó, Endogenous single-strand DNA breaks at RNA polymerase II promoters in Saccharomyces cerevisiae, Nucleic Acids Res. 46 (2018) 10649–10668, https://doi.org/10.1093/nar/gky743.
[74] L. Székvölgyi, Z. Rákosy, B.L. Bálint, E. Kókai, L. Imre, G. Vereb, Z. Bacsó, K. Goda, S. Varga, M. Balázs, V. Dombrádi, L. Nagy, G. Szabó, Ribonucleoprotein-masked nicks at 50-kbp intervals in the eukaryotic genomic DNA, Proc. Natl. Acad. Sci. U. S. A. 104 (2007) 14964–14969, https://doi.org/10.1073/pnas.0702269104.
[75] S.K. Venkata, D. Repana, J. Nulsen, L. Dressler, M. Bortolomeazzi, A. Tourna, A. Yakovleva, T. Palmieri, F.D. Ciccarelli, The Network of Cancer genes (NCG): a comprehensive catalogue of known and candidate cancer genes from cancer sequencing screens, Genome Biol. 20 (2019) 1–12, https://doi.org/10.1186/s13059-018-1612-0.
[76] M.T. Tomicic, D. Aasland, T. Nikolova, B. Kaina, M. Christmann, Human three prime exonuclease TREX1 is induced by genotoxic stress and involved in protection of glioma and melanoma cells to anticancer drugs, Biochim. Biophys. Acta - Mol. Cell Res. 1833 (2013) 1832–1843, https://doi.org/10.1016/j.bbamcr.2013.03.029.
[77] J. Barretina, G. Caponigro, N. Stransky, K. Venkatesan, A.A. Margolin, S. Kim, C.J. Wilson, J. Lehár, G.V. Kryukov, D. Sonkin, A. Reddy, M. Liu, L. Murray, M.F. Berger, J.E. Monahan, P. Morais, J. Meltzer, A. Korejwa, J. Jané-Valbuena, F.A. Mapa, J. Thibault, E. Bric-Furlong, P. Raman, A. Shipway, I.H. Engels, J. Cheng, G.K. Yu, J. Yu, P. Aspesi, M. De Silva, K. Jagtap, M.D. Jones, L. Wang, C. Hatton, E. Palescandolo, S. Gupta, S. Mahan, C. Sougnez, R.C. Onofrio, T. Liefeld, L. MacConaill, W. Winckler, M. Reich, N. Li, J.P. Mesirov, S.B. Gabriel, G. Getz, K. Ardlie, V. Chan, V.E. Myer, B.L. Weber, J. Porter, M. Warmuth, P. Finan, J.L. Harris, M. Meyerson, T.R. Golub, M.P. Morrissey, W.R. Sellers, R. Schlegel, L.A. Garraway, The Cancer Cell Line Encyclopedia enables predictive modelling of anticancer drug sensitivity, Nature 483 (2012) 603–607, https://doi.org/10.1038/nature11003.

nature communications dc_1977_21

Article

https://doi.org/10.1038/s41467-022-32709-y

NODULIN HOMEOBOX is required for heterochromatin homeostasis in *Arabidopsis*

Received: 3 October 2021

Accepted: 11 August 2022

Published online: 27 August 2022

Check for updates

Zsolt Karányi[1,2,8], Ágnes Mosolygó-L[1,8], Orsolya Feró[1], Adrienn Horváth[1], Beáta Boros-Oláh[1,3,4], Éva Nagy[1], Szabolcs Hetey[1], Imre Holb[5], Henrik Mihály Szaker[6,7], Márton Miskei[1], Tibor Csorba[6] ✉ & Lóránt Székvölgyi[1,3] ✉

Arabidopsis NODULIN HOMEOBOX (NDX) is a nuclear protein described as a regulator of specific euchromatic genes within transcriptionally active chromosome arms. Here we show that NDX is primarily a heterochromatin regulator that functions in pericentromeric regions to control siRNA production and non-CG methylation. Most NDX binding sites coincide with pericentromeric het-siRNA loci that mediate transposon silencing, and are antagonistic with R-loop structures that are prevalent in euchromatic chromosomal arms. Inactivation of NDX leads to differential siRNA accumulation and DNA methylation, of which CHH/CHG hypomethylation colocalizes with NDX binding sites. Hi-C analysis shows significant chromatin structural changes in the *ndx* mutant, with decreased intrachromosomal interactions at pericentromeres where NDX is enriched in wild-type plants, and increased interchromosomal contacts between KNOT-forming regions, similar to those observed in DNA methylation mutants. We conclude that NDX is a key regulator of heterochromatin that is functionally coupled to het-siRNA loci and non-CG DNA methylation pathways.

Functionally, the eukaryotic genome is divided into transcriptionally active euchromatin and silent heterochromatin (so-called 'open' and 'closed' chromatin). In *Arabidopsis*, euchromatin is located along chromosome arms containing most transcriptionally active genes, whose expression depends on tissue/cell type, developmental phase, and environmental conditions. Heterochromatin is depleted in protein coding genes and typically associates with telomeres, (peri)centromeres, transposable elements, silenced rDNA, and small heterochromatic islands interspersed along euchromatic arms. Key functions of heterochromatin include: (i) protection of the genome against unscheduled transposon activities that would lead to genomic instability, (ii) regulation of cell differentiation and cell-type identity by hindering transcription factor-mediated cellular reprogramming, (iii) controlling mitotic cell division by ensuring kinetochore assembly at pericentromeres and sister chromatin cohesion.

Several molecular pathways interact to regulate heterochromatin status, including DNA methylation, histone modifications, and RNA silencing[1,2]. DNA methylation in plants occurs in three sequence contexts (CG, CHG, CHH) driven by different genetic pathways[3,4]. MET1 is responsible for CG methylation, the chromomethylases CMT3 and, to lesser extent, CMT2 catalyze CHG methylation, while DRM1/2 and CMT2 perform CHH methylation. De novo methylation of CHG/CHHs

[1]MTA-DE Momentum, Genome Architecture and Recombination Research Group, Department of Biochemistry and Molecular Biology, Faculty of Medicine, University of Debrecen, Debrecen H-4032, Hungary. [2]Department of Internal Medicine, Faculty of Medicine, University of Debrecen, H-4032 Debrecen, Hungary. [3]Department of Pharmaceutical Technology, Faculty of Pharmacy, University of Debrecen, H-4032 Debrecen, Hungary. [4]Doctoral School of Molecular Cell and Immune Biology, University of Debrecen, H-4032 Debrecen, Hungary. [5]Faculty of Agronomy, University of Debrecen, Böszörményi út 138, 4032 Debrecen, Hungary. [6]MATE University, Genetics and Biotechnology Institute, Gödöllő Pest H-2100, Hungary. [7]Institute of Plant Biology, Biological Research Centre, Szeged H-6726, Hungary. [8]These authors contributed equally: Zsolt Karányi, Ágnes Mosolygó-L.
✉e-mail: Csorba.Tibor.Levente@uni-mate.hu; lorantsz@med.unideb.hu

dc_1977_21

is established by the RNA-dependent DNA methylation pathway (RdDM) involving the generation of specific 24nt heterochromatic small interfering RNAs (het-siRNA)[1,2]. In the upstream phase of RdDM, precursor transcripts are produced by RNA Polymerase IV (Pol IV) that mature into functional het-siRNAs by RNA-dependent RNA polymerase 2 (RDR2) and DICER-Like 3 (DCL3) endoribonuclease[5,6]. In the downstream phase, het-siRNAs are incorporated into the central component of the RNA-Induced Transcriptional Silencing Complex (RITSC), ARGONAUTE4 (AGO4) or its homologs (AGO6, AGO9). RITSC is then recruited to target loci by scaffold transcripts produced by another plant-specific RNA polymerase, Pol V, which ultimately tethers the DRM2 methylase to perform de novo cytosine methylation[5,7].

In *Arabidopsis*, CHG and CHH methylation is mainly associated with heterochromatin and is functionally linked to histone H3 lysine 9 methylation. CHG/CHH methylation reinforces H3K9me1/2 and vice versa: CMT3/CMT2 DNA methylases read the H3K9me2 mark, while SUVH4/5/6 histone methylases (see below) read the mCHG/mCHH signal[8,9]. This mechanical coupling creates a self-amplifying loop between CHH/CHG methylation and histone methylation to silence transposable elements. H3K9me1/2 is catalyzed by KYP/SUVH4, SUVH5, and SUVH6 lysine-methyltransferases, which recognize mCHG/mCHH sites in constitutive heterochromatin[10,11]. (The abundance of H3K9me3 modification is low in *Arabidopsis*.) The H3K9me1/2 signal is recognized by specific reader proteins, of which AGDP1 (also known as ADCP1) has only recently been identified[12,13] and further members await exploration.

Another crucial repressive histone mark is H3 lysine 27 methylation (H3K27me) that occurs independently of DNA methylation pathways[14]. This suggests that the two repressive histone modifying systems act autonomously on heterochromatin formation. H3K27me1 is typically located in chromocenters and deposited by ATXR5/6, while H3K27me2/3 form heterochromatic patches in chromosome arm-associated gene loci, catalyzed by the Polycomb Repressive Complex 2 (PRC2). Polycomb Group (PcG) proteins also form the Polycomb Repressive Complex 1 (PRC1) that has H3K27me3-reader and histone H2A E3 ubiquitin ligase activities[15] and mediates gene repression through histone H2A lysine 119 mono-ubiquitination (H2AK119ub1). Several factors have emerged in recruiting PRC1 to chromatin, such as the PRC2 component CLF[16], the CAF-1 histone chaperone[17], the AL6 PhD finger protein[18], the LHP1 chromodomain protein[16], and more recently, the homeodomain protein NDX (NODULIN HOMEOBOX), which is the only transcription factor to date that interacts with PRC1[19].

NDX is a nuclear transcription factor that appeared early in evolution and is present in vascular plants[20]. It contains an atypical homeobox domain (HD) and two additional NDX-A/B domains, of which the HD and NDX-B domains were shown to bind double-stranded and single-stranded DNA in vitro[19,21,22]. The specific role of these domains is not known in detail. NDX has been shown to associate with the PRC1 core components RING1A/B, making a functional link with the E3 ubiquitin ligase module of the complex[19]. Previously, NDX was shown to regulate the expression of *FLOWERING LOCUS C* (*FLC*), a central integrator of flowering transition, by stabilizing an R-loop structure at the 3′-end of the locus[21]. This region serves as a terminator of sense (*FLC*) transcription and a promoter of antisense (*COOLAIR*) transcription, the latter being inhibited by an unknown mechanism that suppresses *FLC* transcription and accelerates flowering transition. A recent study identified FCA (an RNA binding protein) and FY (a component of the cleavage polyadenylation specificity factor complex) to recognize and resolve the R-loop in *FLC* and a subset of other loci[23,24]. However, the role of NDX in these processes was not studied in a mechanistic detail and was omitted from the working model of co-transcriptional chromatin silencing by R-loop resolution[23]. The genome-wide association of NDX with R-loops and its global effect on ncRNA transcription remains a fundamental question that needs to be addressed. In addition, NDX may also affect gene expression by modifying spatial chromatin organization, as enhanced DNA looping interactions were observed at *FLC* in an *ndx* mutant[21]. A similar mechanism has been hypothesized on the regulation of *ABI4* through spatial interactions between NDX and PRC1[19].

The possibility of NDX-mediated genome structural changes seems realistic given that in animals PRC1/RING1B (i) causes chromatin compaction through its non-enzymatic function[25,26], (ii) mediates long-range promoter-promoter interactions between developmentally regulated genes[27], and (iii) orchestrates estrogen-induced enhancer-promoter looping interactions[28]. In *Drosophila*, the NDX-related zeste protein[22] was shown to be required for long-range communication between promoters and for insulator bypass during gene activation[29,30]. In addition, a recent study in *Arabidopsis* showed that removal of the PRC1 component LHP1 from chromatin (through *APOLO* lncRNA-mediated R-loop formation) induces 3D chromatin conformational changes[31].

Herein, we characterized the global role of NDX in chromatin-based gene regulatory processes using cytological, molecular, and high-throughput approaches. Our results show that NDX is a key regulator of heterochromatin accessibility and chromatin packing in pericentromeric regions that are functionally coupled to non-CG methylation pathways.

Results

NDX is a heterochromatin-associated factor

To understand the genome-wide regulatory roles of NDX, we first analyzed its genomic distribution. For this, we performed chromatin immunoprecipitation sequencing (ChIP-seq) in 10-day old *Arabidopsis* seedlings expressing N-terminally and C-terminally tagged NDX fusion proteins (flag-NDX/*ndx1-1*(FRI)/*flc-2* and NDX-GFP/*ndx1-1*(FRI)/*flc-2*, respectively) expressed from their endogenous promoter[21,32]. The two ChIP-seq profiles were highly correlated (Pearson $r > 0.85$, Supplementary Fig. 1), and 2243 flag-NDX and 583 NDX-GFP binding sites were identified (Fig. 1a) showing a statistically significant overlap compared to random overlaps ($p < 0.0001$, prop.test). Enrichment of NDX was validated by ChIP-PCR at selected genomic loci (Fig. 1b). Importantly, binding of NDX to both the *ABI3* and *ABI4* downstream regions[19] and the *FLC* 3′-terminator[21] was also confirmed (Fig. 1c), indicating that our analysis is robust and reliable. We then visualized the density of ChIP peaks along the entire length of *Arabidopsis* chromosomes and revealed a significant colocalization of NDX with centromeric and pericentromeric regions (Fig. 1d), in marked contrast to the five chromosomal arms and telomeres, where the density of peaks was much lower (16 peaks/Mb in arms *vs.* 84 peaks/Mb in pericentromeres). This non-random distribution suggests that NDX is mainly associated with gene-poor heterochromatic regions and is rare in transcriptionally active euchromatin.

To confirm the genomic binding of NDX by a sequencing-independent approach, we applied confocal laser scanning microscopy (CLSM) coupled to fluorescence correlation spectroscopy (FCS) and fluorescence recovery after photobleaching (FRAP). CLSM performed in live root tips (expressing NDX-GFP) showed strong green fluorescence near the nuclear periphery (typically associated with perinuclear heterochromatin) and in the nucleolus (Fig. 2a and Supplementary Fig. 2). Nucleolar staining was visible in <20% of cells and it is currently unclear whether this fraction represents a functional or non-functional population that acts as a "storage depot". In contrast, peripheral nuclear localization was consistently observed and was further reinforced by the overlap between NDX binding sites and NUCLEOPORIN1 (NUP1)-enriched chromatin[33], which marks nuclear periphery (Supplementary Fig. 3). Of the NUP1 peaks classified as pericentromeric and arm-associated, only the former group showed significant colocalization with NDX. Pericentromeric enrichment was confirmed microscopically in a subset of cells that showed a typical chromocenter structure (Fig. 2b), which represents a small fraction of

https://doi.org/10.1038/s41467-022-32709-y

dc_1977_21

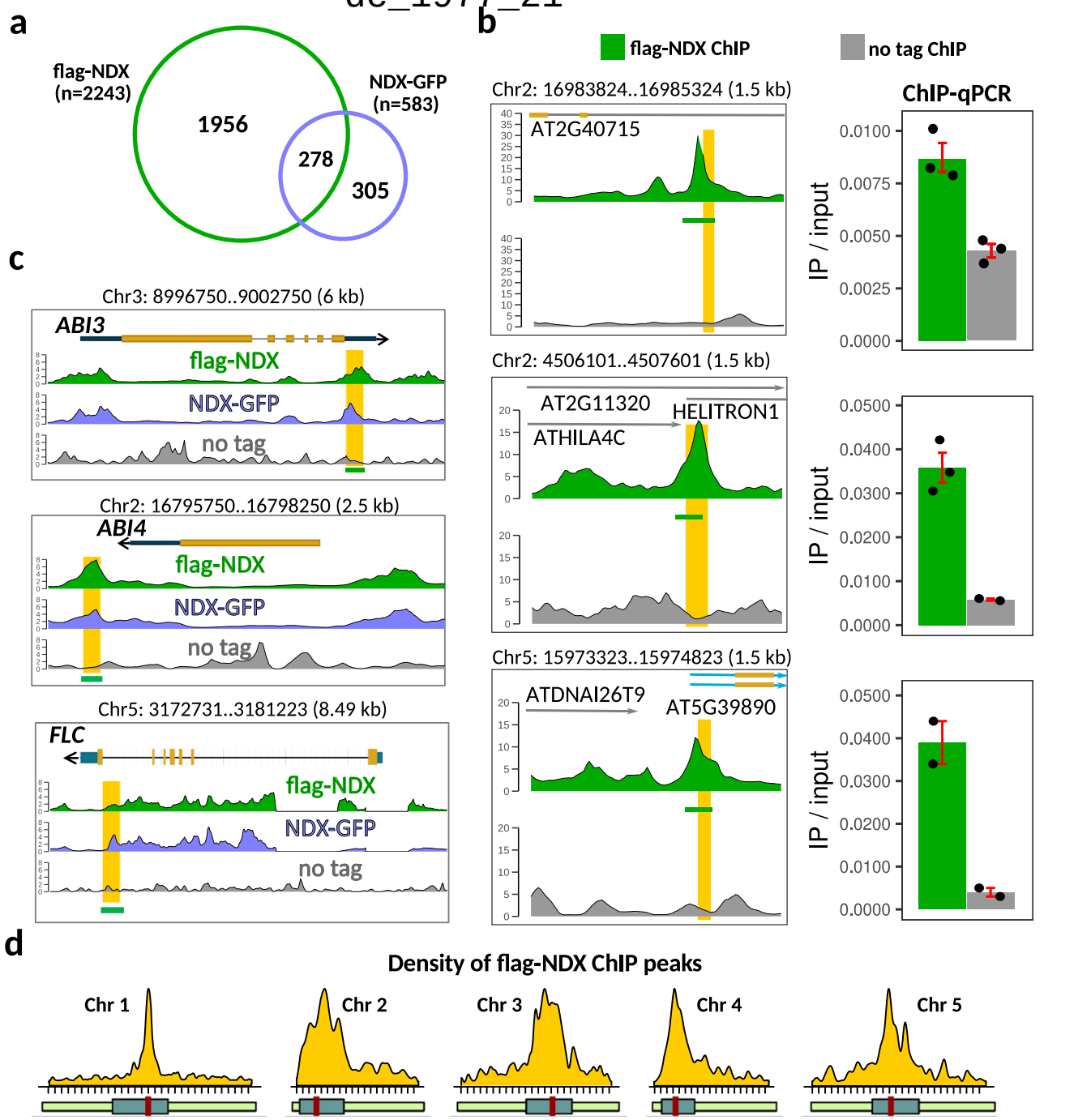


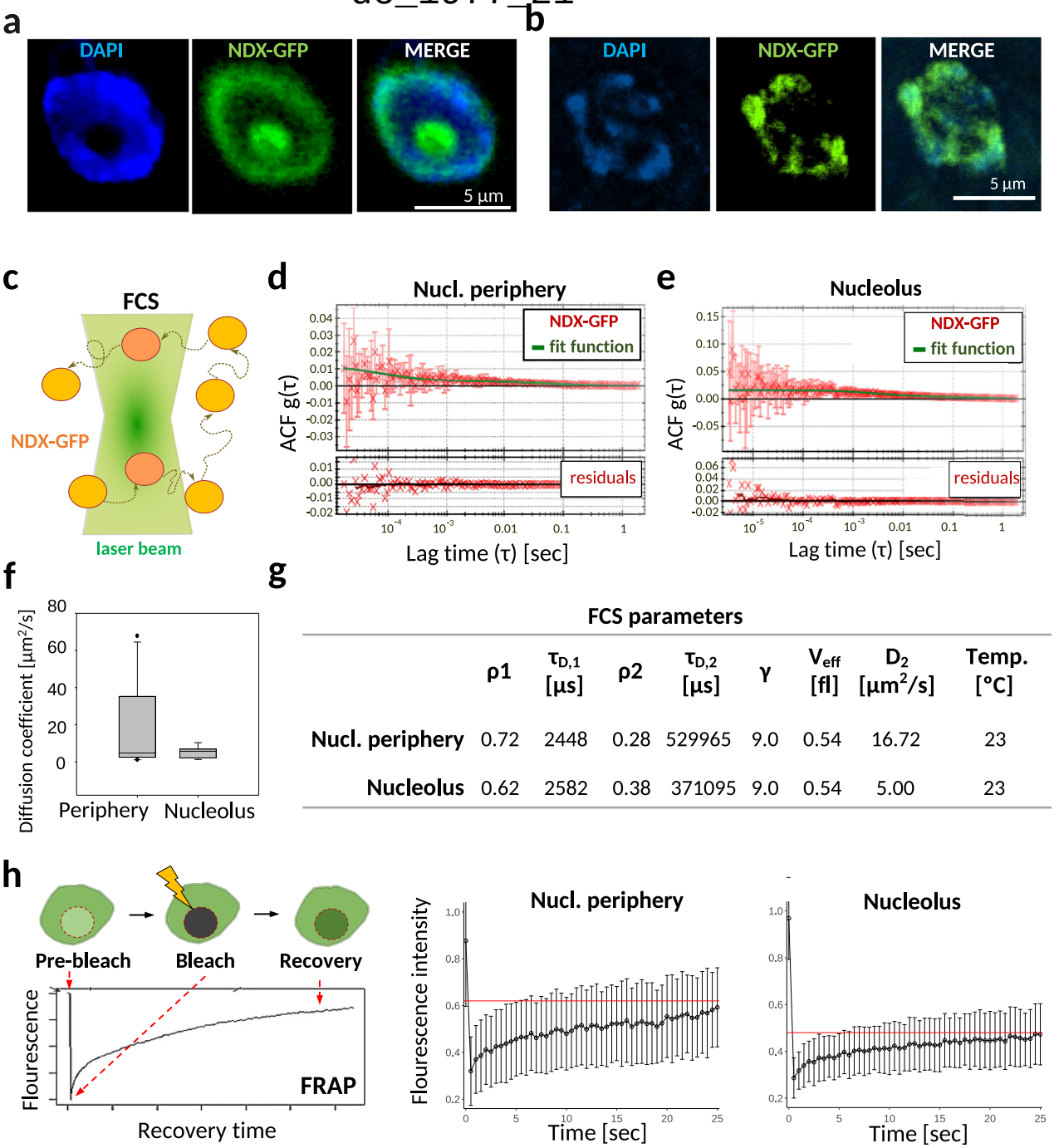
Fig. 1 | Genome-wide mapping of chromosomal binding sites of NODULIN HOMEOBOX (NDX). a Overlap of ChIP-seq peaks identified in N-terminally and C-terminally tagged flag-NDX and NDX-GFP lines, respectively. The proportion of common peaks is significantly higher than expected from a computer randomized distribution (***$p < 0.0001$, prop.test, two-sided; Bonferroni correction). **b** ChIP-qPCR validation of representative flag-NDX binding sites from different chromosomes. Genomic positions are indicated on the top of each panel. Specific ChIP signal and background signal ("no tag") is shown in green and black, respectively. Positions of qPCR amplicons are highlighted in yellow. Error bar: SEM. Sample size $n = 3$ biologically independent replicates. **c** Genome browser tracks showing the distribution of flag-NDX and GFP-NDX ChIP signals over *ABI3*, *ABI4*, and *FLC* loci that were previously shown by qPCR to bind NDX. Positions of qPCR amplicons are highlighted in yellow. Sample size $n = 2$ biologically independent samples for each track. **d** Density of ChIP peaks along the five *Arabidopsis* chromosomes (peak count/Mb). Most NDX peaks are enriched near centromeres, while less peak density is characteristic of arms.

root tip cells in which chromocenters can be visually detected. The NDX signal was weak in the nuclear interior, which typically coincides with euchromatin.

We then used the FCS approach to calculate the effective diffusion coefficient of the mobile fraction of NDX (Fig. 2c). From the FCS parameters, the time-dependent autocorrelation function was derived (G(τ) or ACF; Fig. 2d, e) from which we calculated the average diffusion coefficient at the nuclear periphery ($16.72 \pm 7.1\,\mu m^2$/sec) and the nucleolus ($5.00 \pm 0.9\,\mu m^2$/sec; Fig. 2f, g). Since the diffusion parameters did not differ significantly between nuclear and nucleolar NDX, these compartments are likely to have closely related NDX-binding properties. In the FRAP setting (Fig. 2h, left), GFP fluorescence was bleached within the nuclear periphery and nucleolus, and then fluorescence intensities were tracked (i) in the bleached region (F_{bleach}), (ii)

https://doi.org/10.1038/s41467-022-32709-y

dc_1977_21



g FCS parameters

	ρ1	$\tau_{D,1}$ [µs]	ρ2	$\tau_{D,2}$ [µs]	γ	V_{eff} [fl]	D_2 [µm²/s]	Temp. [°C]
Nucl. periphery	0.72	2448	0.28	529965	9.0	0.54	16.72	23
Nucleolus	0.62	2582	0.38	371095	9.0	0.54	5.00	23

Fig. 2 | Microscopic localization and molecular diffusion of NDX in Arabidopsis root tips. a, **b** Representative confocal images showing the pattern of NDX-GFP distribution in DAPI stained nuclei. The experiments were repeated four times with similar results. **a** The GFP signal is enriched in the nuclear periphery and in the central (DAPI-negative) region, corresponding to the nucleolus. Nucleolar staining is detected in 5–20% of cells. **b** A minor fraction of cells with typical chromocenter structure (DAPI foci) shows strong NDX-GFP enrichment at chromocenters. **c** Scheme of fluorescence correlation spectroscopy (FCS) measurement. **d**, **e** Representative time-dependent autocorrelation functions (ACFs) with the estimated diffusion parameters and residuals. ACF curves represent the correlation coefficient between the fluctuation from the mean GFP fluorescence intensity at time I(0) and the fluctuation from the mean intensity at some time later I(t). The curves were fitted with a two-component 3D normal diffusion model. Sample size $n = 16$ individual cells examined over three independent experiments. **f** Distribution of diffusion coefficients (D) of fast components of NDX-GFP at the nuclear periphery and nucleolus. D values do not differ significantly ($p = 0.820$, Mann–Whitney test, two sided). Bounds of boxes describe the interquartile range with the median; whiskers indicate 95% confidence interval; dots show outliers. **g** Main FCS parameters and calculated values. ρ1: slow FCS component; $\tau_{D,1}$: diffusion time of slow component; ρ2: fast FCS component; $\tau_{D,2}$: diffusion time of fast component; γ: structure factor of the confocal volume; V_{eff}: effective measurement volume; D_2: diffusion coefficient of the fast component in µm²/s unit. The number of cells analyzed were 50. **h** Left: scheme of fluorescence recovery after photobleaching (FRAP) measurement. Right: FRAP curves showing slow fluorescent recovery of NDX-GFP at both the nuclear periphery and the nucleolus. Immobile (non-moving) fractions correspond to the area above the horizontal red line. Sample size $n = 56$ individual cells examined over 3 independent experiments. Error bars: SD.

dc_1977_21

in total nuclei (F_{total}), and (iii) in randomly selected regions outside the nucleus, used for background subtraction (F_{bckgr}). Recovery curves were obtained by normalizing the background-subtracted signal to the mean prebleach signal and, at the same time, considering the decrease of total fluorescence due to the initial high-intensity laser pulse and bleaching upon post-bleach imaging[34]. For the mobile fraction of NDX, we obtained a very slow recovery rate that recapitulates the slow kinetics and tight chromosome binding of nuclear core histone (H3/H4) proteins[35]. GFP fluorescence did not return to the initial pre-bleach value, leaving 40–60% of NDX in the slowly exchanging (static) fraction (Fig. 2h, area above the horizontal red line). The latter immobile fraction most likely represents NDX molecules that bind directly and permanently to chromatin. Taken together, the above quantitative microscopic data obtained in living cells are consistent with our ChIP-seq results and extend them to different spatial resolutions and timescales, indicating that NDX is a chromatin-binding factor that is stably incorporated into chromosomes.

NDX shows opposite genomic distribution to R-loops

To address the possibility that NDX is a general R-loop regulator, we compared the chromosomal distribution of NDX binding sites with the genomic profile of R-loops. For this, RNA-DNA hybrids were immunoprecipitated with the S9.6 antibody[36,37] and mapped by DNA-RNA hybrid immunoprecipitation (DRIP) sequencing[38,39]. The specificity of S9.6 immunoselection was monitored by RNaseH treatment, which degrades the RNA strand of RNA-DNA hybrids (Fig. 3a–c). Our peak prediction identified 14124 R-loops in Col-0 seedlings that followed an antagonistic distribution to NDX (Fig. 3a, c) as 97% of DRIP peaks ($n = 13798$) and 85% of NDX peaks ($n = 1910$) showed no overlap (Fig. 3b, c). Differential association of NDX and R-loops was also analyzed in gene-rich euchromatic regions by a peak-independent (metagene) approach: strong DRIP signal enrichment was found over promoters and 5′-UTRs while NDX was enriched at transcriptional termination sites (Fig. 3d). The predicted R-loops were validated by DRIP-qPCR at selected loci that appeared as positive (enriched) and negative (depleted) sites according to our analysis (Fig. 3e). The R-loop structure previously described at the 3′-end of *FLC*[21] was also detected and further validated by qPCR (Supplementary Fig. 4), confirming the quality of our DRIP-seq data. Importantly, despite the opposite genomic distribution of NDX binding sites and R-loops, 15% of NDX peaks ($n = 333$, green) and 2.3% of DRIP peaks ($n = 326$, red) still overlapped (Fig. 3b, c, "common peaks"). This overlapping fraction is statistically significant compared to random enrichment ($^*p < 0.001$, prop.test) and suggests that the NDX-stabilized R-loop model described for the *FLC* locus[21] may still be true in some cases, though not in general. Accordingly, secondary ssDNA structure prediction showed that NDX binding sites have a lower propensity to form ssDNA than DRIP peaks or common NDX/DRIP binding sites (Fig. 3f), which tend to adopt a single-stranded conformation. In other words, NDX binds primarily to double-stranded DNA regions and, to a lesser extent, to the classical R-loop structure containing ssDNA (represented by the common DRIP/NDX peak set). This is consistent with previous reports demonstrating a strong binding affinity of NDX for dsDNA templates[19,22]. However, mechanistic understanding of NDX binding to dsDNA, ssDNA, and R-loop structures remains unknown.

NDX binds to het-siRNA loci associated with pericentromeric heterochromatin

Functional annotation of the identified ChIP and DRIP peaks revealed a significant enrichment of flag-NDX and NDX-GFP in pericentromeric regions and transposable elements (Fig. 4a). The Knob region, a constitutive heterochromatin domain located on the short arm of Chr4 rather than the pericentromere[40,41], also showed preferential NDX enrichment, suggesting that NDX is generally associated with this type of chromatin. Of the annotated TEs, some of the families showed significant NDX-enrichment characterized by diverse transposition mechanisms, internal structure, and chromosomal distribution (Fig. 4b). NAD transposons, which are localized cytologically in the nucleolus and participate in forming so-called 'nucleolus associated chromatin domains' (NADs)[42], also showed NDX association (Fig. 4a), consistent with our microscopic data. We also performed a family-independent transposon analysis based on TE targeting by non-CG methylation pathways[43] and compared TE groups controlled by RdDM (RdDM TEs) to those targeted by the CMT2 pathway (CMT2 TEs) or both pathways (intermediate TEs) (Fig. 4c). The above TE categories were further classified by chromosomal location as pericentromeric and arm-associated. The results show that NDX is primarily enriched at TEs regulated by the CMT2 pathway as well as the common pathway and is depleted from RdDM-targeted TEs, regardless of pericentromeric or arm location. These functional relationships link NDX to non-CG methylation mediated by CMT2 and/or the CMT2/RdDM common pathway. This conclusion is reinforced by the association of NDX with repressive chromatin modifications (H3K9me2, H3K27me1) required to silence TEs (Fig. 4a, b). Of note, there was no colocalization between NDX and H2A ubiquitination (H2Aub) mediated by PRC1, which is not unexpected as this repressive mark is constantly removed by ubiquitin proteases to drive PRC2-catalyzed H3K27 trimethylation[44].

R-loops were mostly associated with 5′-UTRs, but also with promoters, specific transposons (e.g. NAD/En-Spm/MuDR/Copia), and euchromatic histone marks located in active chromatin (Fig. 4a–d). It is noteworthy that NAD TEs and "canonical" TEs (located outside the nucleolus) show differential R-loop association as the former group is enriched with DRIP peaks while the latter show depletion (Fig. 4a). Thus, R-loops appear to associate with nucleolar TEs and have less preference for extranucleolar TEs. These associations are consistent with previous research that found high R-loop abundance in the nucleolus[45–49]. Interestingly, NAD TEs associated with NDX and DRIP peaks overlap significantly ($p < 0.001$, prop.test), suggesting their possible functional interaction on a subset of nucleolar transposons.

Next, we correlated the localization of NDX and DRIP peaks with nine different chromatin states of *Arabidopsis* (Fig. 4e, left) reconstituted from the combinatorial pattern of epigenomic landscapes and transcriptome maps (based on ref. 50). NDX binding sites showed preferential associations with transcriptionally silent heterochromatin (states 8–9) enriched in H3K9me2, H3K27me1, CG/CHG/CHH methylation, transposons and transposon genes (Fig. 4e, right). In contrast, RNA-DNA hybrids mainly localized to transcriptionally active euchromatin, which corresponds to state 1 (high levels of transcription, H3K4me2/3, H3K36me3, H2Bub, H3.3, H2A.Z, and H3K9ac), state 2 (high levels of H3K4me2/3, H3.3, H2A.Z, and H3K27me3; moderate levels of active chromatin marks)[50]. Notably, NDX binding sites were highly anticorrelated with states 1–2 (where R-loop structures are typically enriched). Interestingly, states 1/2 and 8/9, the most extreme chromatin states, lie furthest apart on the linear scale of the genome[50], consistent with the reciprocal chromosomal distribution of NDX peaks and R-loops (Fig. 5a, top panel).

Functionally distinct small RNAs (sRNAs) were previously classified into nine functional groups and linked to different chromatin states[51]. Utilizing this sRNA database, NDX-binding sites showed a particularly strong colocalization with class 6–9 sRNA loci (Fig. 5a, b) that code for centromeric and pericentromeric 21–24nt siRNAs participating in CHH/CHG methylation via the CMT2/3 pathway (classes 6,8,9) or the RdDM pathway (class 7)[51]. However, NDX peaks were anticorrelated with class 1–3 euchromatic sRNA loci that tend to associate with genes. Nevertheless, NDX enrichment at class 6–9 sRNA loci was rather heterogeneous as NDX bound strongly to a subset of targets (Fig. 5c). Our metaanalysis shows that NDX does not bind to

https://doi.org/10.1038/s41467-022-32709-y

dc_1977_21

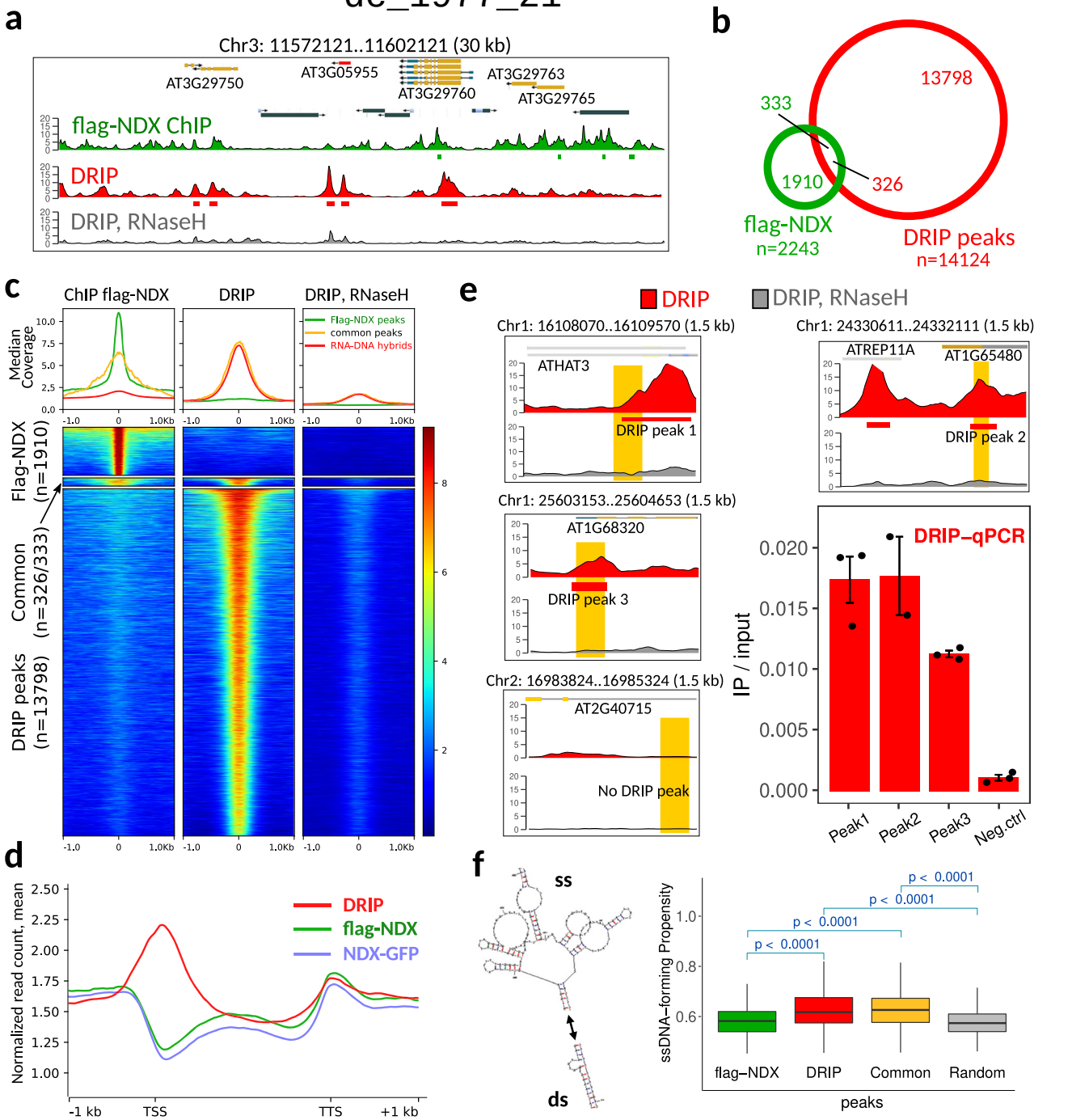


Fig. 3 | Antagonistic genomic distribution of NDX binding sites and R-loops. **a** Genome browser snapshot showing the enrichment of flag-NDX and DRIP signals along chromosome 3. RNase H-treated sample represents the background DRIP signal. RNA-DNA hybrids were immunoprecipitated by S9.6 monoclonal antibodies. **b** Venn diagram showing the overlap of flag-NDX peaks and DRIP peaks (R-loops). The number of common peaks is not equal because ChIP peaks can overlap with multiple DRIP peaks and vice versa. **c** Pile up plot showing the antagonistic distribution of ChIP and DRIP signal intensities over the predicted peak positions. RNase H treatment is shown as a negative control for DRIP. Color scale corresponds to RPGC (reads per genomic content) values. Peak summits were aligned to zero positions. **d** Metagene profile of flag-NDX (light blue), NDX-GFP (light green) and DRIP signal (red) intensities (mean values) over protein coding ORFs. TSS: transcription start site. TTS: transcription termination site. **e** DRIP-qPCR validation of representative DRIP peaks. Specific DRIP signal and background (+RNase H) signal are shown in red and black, respectively. Genomic positions and qPCR amplicons are indicated. Error bars represent SEM. **f** Secondary ssDNA structure prediction over NDX binding sites and RNA-DNA hybrids. The plot shows the propensity of single-strand formation of individual flag-NDX peaks and DRIP peaks, common NDX/DRIP peaks, and randomly selected regions. Sample size $n = 300$ peaks randomly selected from DRIP and ChIP peak lists, and $n = 300$ random peaks. The ssCount values were calculated from the primary nucleic acid sequence of the peaks using the mfold algorithm. Statistical significance and p values are indicated (Mann-Whitney U test, two-sided). Bounds of boxes describe the interquartile range with the median; whiskers indicate minimum and maximum values; outliers are not shown.

https://doi.org/10.1038/s41467-022-32709-y

dc_1977_21

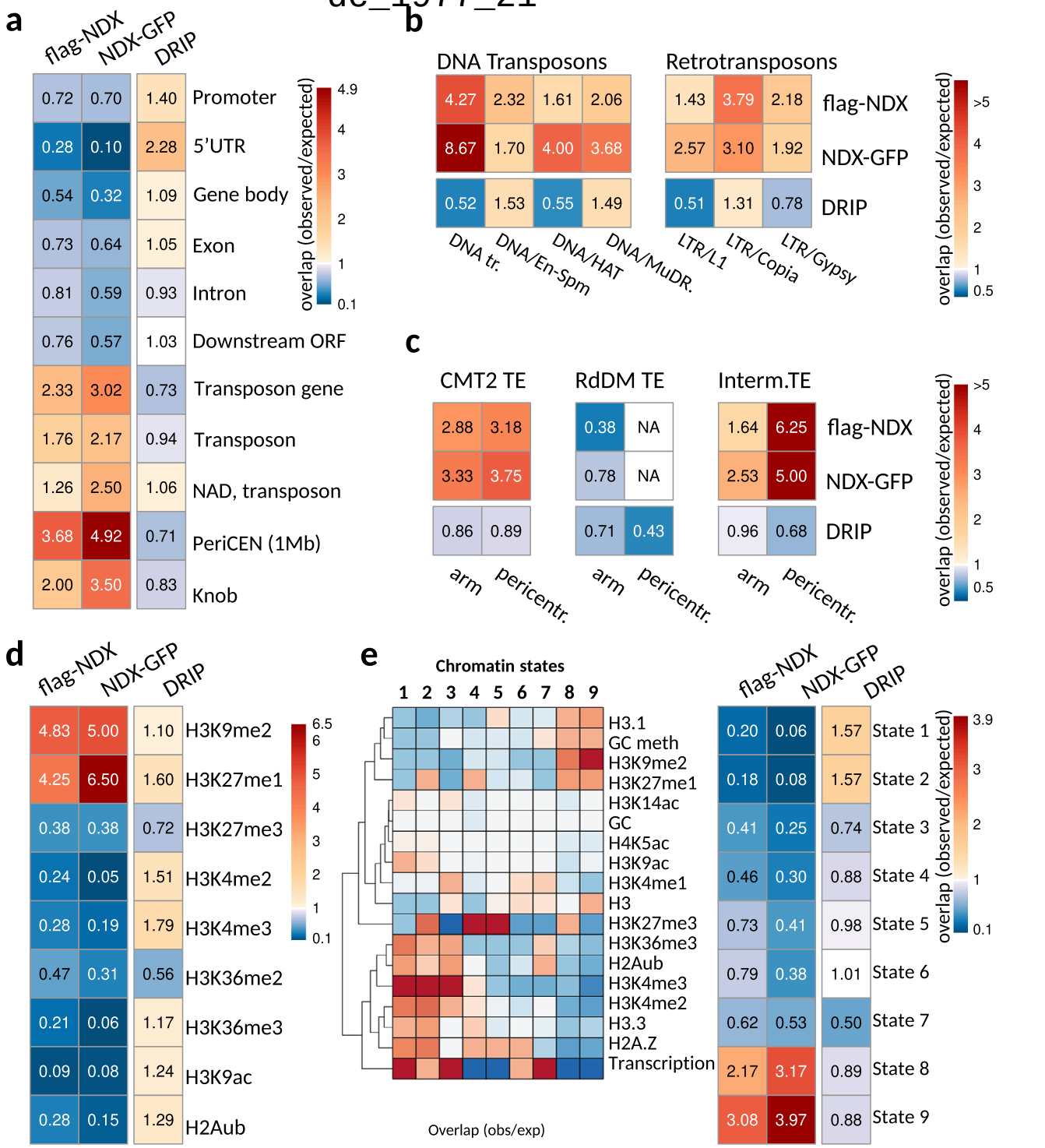


Fig. 4 | NDX is preferentially associated with constitutive heterochromatin. **a** Functional annotation of flag-NDX, NDX-GFP, and DRIP (R-loop) binding sites over genomic features. Cells contain observed/expected ratios for peak counts. Warmer colors represent higher enrichment. **b** DNA transposon and retrotransposon families associated with NDX and DRIP peaks. (Only families showing significant enrichment of ChIP or DRIP peaks are shown.) **c** Enrichment of NDX and DRIP peaks over the functional categories of TEs. CMT2 TEs were compared to those controlled by the RdDM pathway or both pathways (intermediate TEs). TE categories were further classified as pericentromeric and arm-associated. NA: lack of statistical power due to very low peak count in the indicated category. **d** The same annotation as above for histone modifications, and **e** chromatin states. The nine chromatin states were defined by the combinatorial pattern of epigenetic modifications and transcription[50].

sRNA #1–5 loci (as expected), nor does it bind to sRNA #6–9 loci located on chromosomal arms, however, pericentromerically localized sRNA #6–9 loci bind significantly but heterogeneously to NDX (Fig. 5c). NDX enrichments were validated at selected class 6–9 sRNA loci using ChIP-qPCR (Fig. 5d). In contrast to NDX, RNA-DNA hybrids were mainly associated with sRNA classes 1–2 that are related to protein coding ORFs and promoters[51]. The genome-wide enrichment of NDX over het-siRNA loci is fully consistent with previous data that identified two sRNA molecules at the *FLC* locus (24 nt and 30 nt) that colocalized with NDX in a heterochromatic patch of H3K9me2[21,52].

https://doi.org/10.1038/s41467-022-32709-y

dc_1977_21

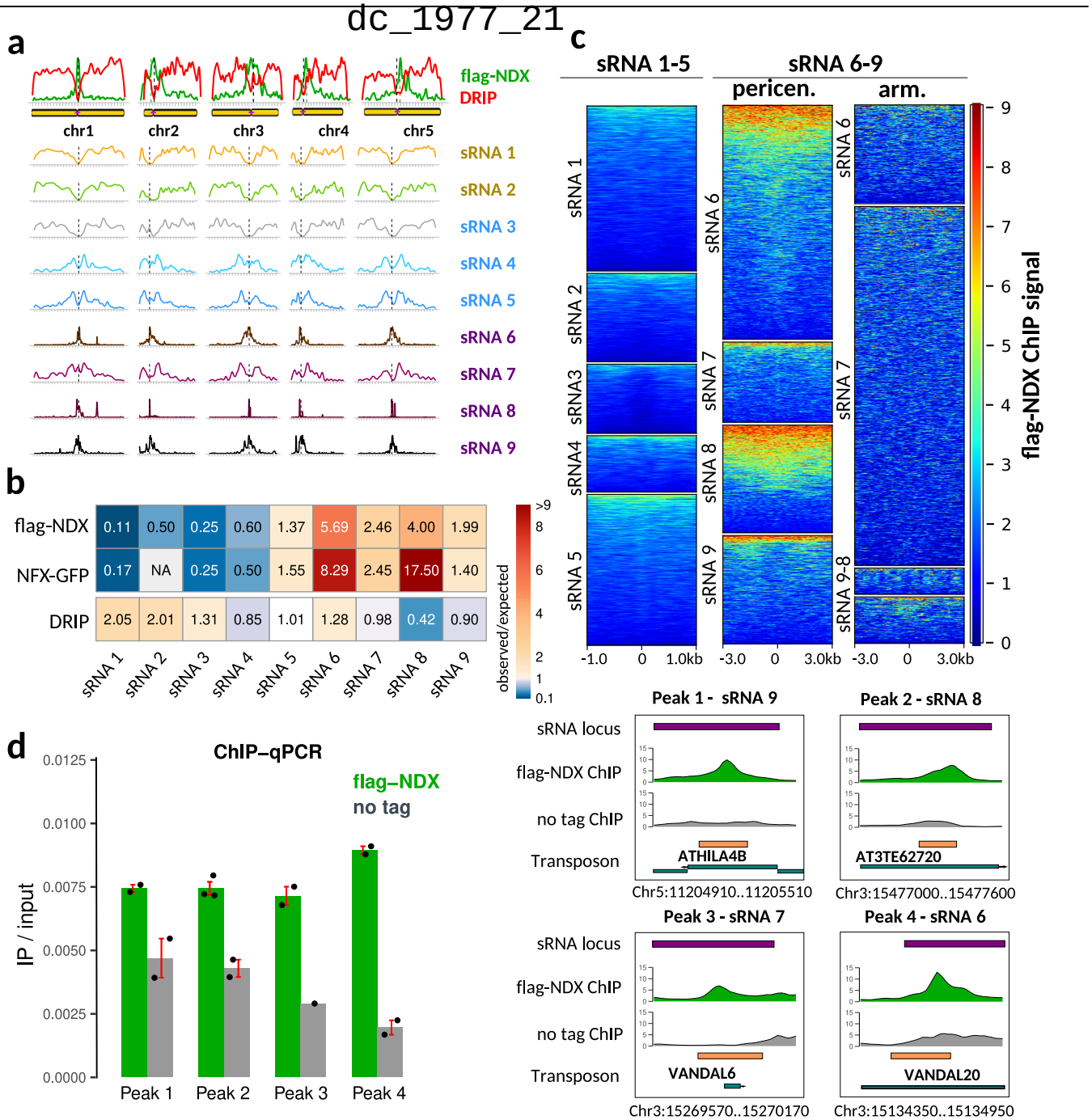


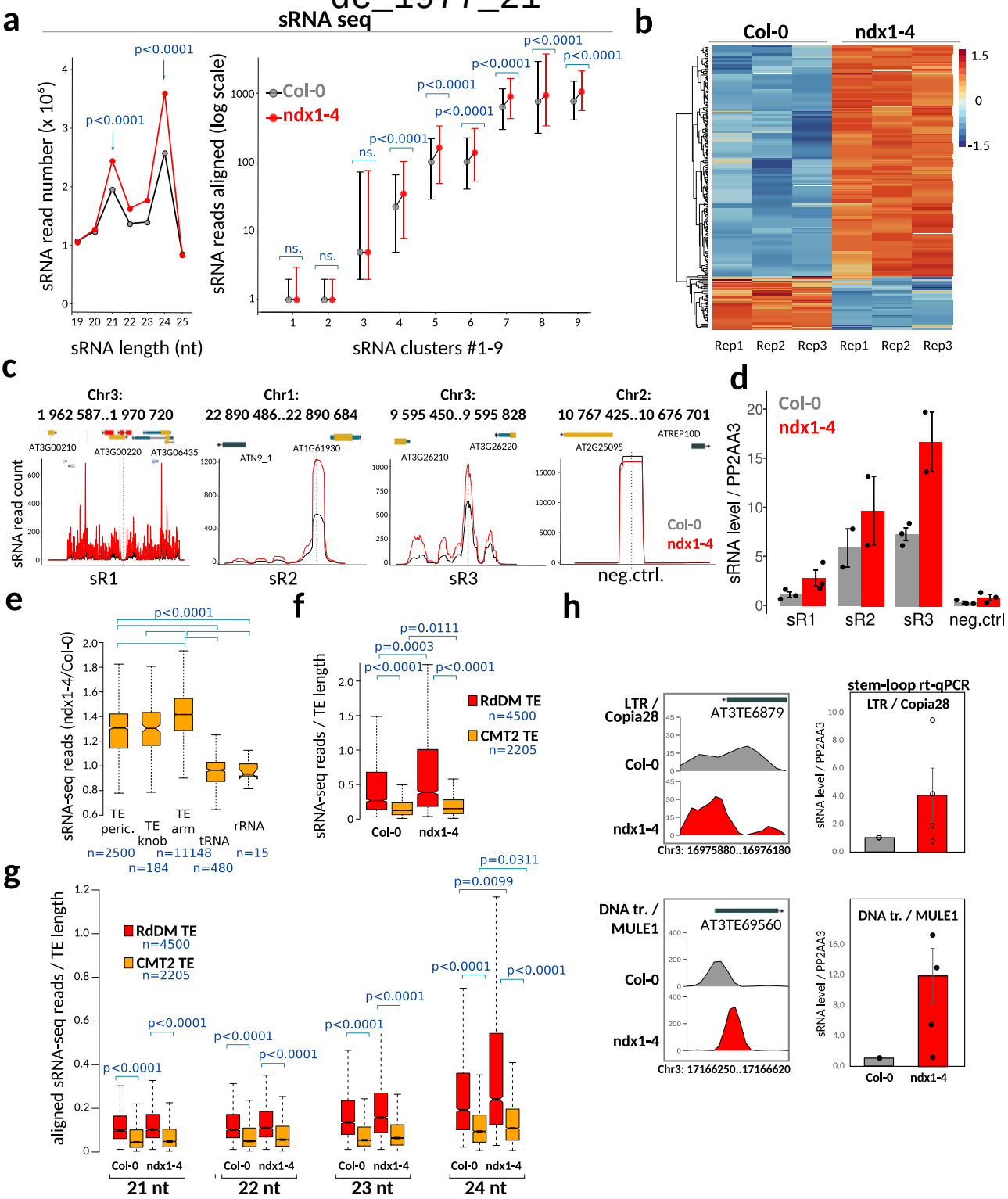
Fig. 5 | NDX binds to heterochromatic small RNA loci (het-siRNA). **a** Chromosomal distribution of flag-NDX binding sites (green), RNA-DNA hybrids (red), and small RNA loci. sRNA loci are classified into nine functional clusters based on their genetic and epigenetic relationships and expression profiles[51]. NDX-binding sites correlate with cluster 6–9 sRNA loci while R-loops are correlated with cluster 1–2 sRNA loci. **b** Overlap of flag-NDX, NDX-GFP, and DRIP peaks with sRNA clusters relative to random associations. Cells contain observed/expected ratios for peak counts. Warmer colors indicate higher degree of association. **c** Metaplot analysis of NDX ChIP signal over sRNA loci 1–5 and 6–9. sRNA loci 6–9 have been divided by chromosomal location as pericentromeric and arm associated. The NDX ChIP signal is preferentially enriched in pericentromerically located sRNA 6–9 loci, however, their binding affinity for NDX is heterogeneous. **d** Left: ChIP-qPCR validation of representative sRNA loci that are bound by NDX. Specific ChIP signal and background signal ("no tag") is shown in green and grey. Error bar: SEM. Sample size $n = 2$ biologically independent experiments. Right: JBrowse snapshot of the test regions.

Loss of NDX function results in increased het-siRNA expression and R-loop formation

To gain functional insights into the association of NDX with het-siRNAs, we performed sRNA deep sequencing (sRNA-seq) in an *ndx* T-DNA insertion line (*ndx1–4*) to assess global changes in the production of sRNAs (Fig. 6). Our analysis revealed significantly higher sRNA abundance in the *ndx1–4* mutant than in wild-type (Col-0), however, the size distribution and the ratio of dominant sRNA fractions (24nt/21nt) were similar (Fig. 6a, left). Of the nine functional groups of sRNA loci, only class 4–9 showed increased sRNA levels in *ndx1–4*, whereas class 1–3 did not differ from the wild type (Fig. 6a, right). These heterogeneities are in accordance with the primary division between sRNA loci that occurs between classes 1–3 and 4–9, determined by their associations with protein coding ORFs and promoters (sRNAs 1–3), and with epigenetically activated small interfering RNAs (easiRNAs) and non-CG methylation pathways (sRNAs 4–9)[51]. Differential expression analysis

https://doi.org/10.1038/s41467-022-32709-y

dc_1977_21



identified 286 upregulated and 165 downregulated sRNA loci ($n = 451$) that showed significantly increased or decreased sRNA levels in the *ndx1–4* mutant (Fig. 6b). About 85% of these sRNAs belong to classes 4–9 and thus depend on Pol IV activity (Supplementary Fig. 5) such that the proportion of classes 4 and 5 was doubled compared to non-differential ones. These associations demonstrate that Pol IV-derived/dependent siRNAs are specifically altered in the *ndx1–4* mutant.

Regarding the expression of annotated microRNAs (that are known to act post-transcriptionally), we found only 10 up- and 2 downregulated miRNAs from all known 326 precursor hairpin RNAs[53]. The small effect of *ndx1–4* mutation on miRNA transcription and the lack of association between NDX ChIP peaks and miRNA loci preclude the possibility that NDX is a central regulator of miRNA expression. In contrast, differential expression of class 6–9 siRNAs is probably due to the direct binding of NDX to these sRNA loci, which follows from their particularly strong association (Fig. 5b). Representative sRNA loci upregulated in *ndx1–4* and their validation by stem-loop rt-qPCR[54] are shown in Fig. 6c, d.

Regarding transposon-derived siRNA production, TE siRNAs also showed a clear increase in the absence of NDX, independent of their

dc_1977_21

Fig. 6 | Loss of NDX function induces increased sRNA expression and R-loop formation. a Left: sRNA expression levels are significantly increased in the *ndx1–4* mutant compared to Col-0. The size distribution of sRNA molecules peaks at 21-nt and 24nt, respectively. Sample size $n = 3$ biologically independent experiments. Statistical sifnificance is indicated (prop.test, two sided, Bonferroni correction). Right: Distribution of sRNA levels in Col-0 and *ndx1–4* plants over the nine functional sRNA clusters identified by[51]. Box plots show the medians and 95% confidence intervals. Clusters 1–3 are associated with protein coding ORFs and miRNA loci and sRNA levels show no difference between *ndx1–4* and Col-0. Clusters 4–9 are associated with Pol IV, and in part, Pol V and sRNA levels are significantly increased in *ndx1–4* ($p < 2.2e\text{-}16$, Kruskal–Wallis test with Mann–Whitney test (for multiple pairwise comparison) and Benjamini–Hochberg correction). Sample size $n = 3$ biologically independent experiments. **b** Heatmap showing upregulated (orange) and downregulated (blue) sRNA loci in *ndx1–4*. sRNA reads were aligned to sRNA loci from[51]. Three independent biological replicates are shown in the diagram. $P < 0.0001$, Mann–Whitney rank sum test. **c**, **d** Representative examples of upregulated sRNA loci and their validation by stem loop rt-qPCR. The bar chart shows positive strand expression levels normalized to PP2AA3 RNA expression. Sample size $n = 3$ biologically independent experiments. Error bar: SEM. **e** Increase of siRNA-levels in the absence of NDX function. The ratio of aligned siRNA reads (*ndx1–4*/Col-0) is significantly increased in pericentromeric TEs, Knob TEs and chromosome arm TEs relative to "negative control" regions (tRNA and rRNA genes). Statistics: Kruskal–Wallis test with Mann–Whitney test (for multiple pairwise comparison) and Benjamini–Hochberg correction. Sample size (n) is indicated. **f** The number of aligned sRNA reads show a significant increase in *ndx1–4* over the functional categories of TEs (RdDM TEs vs. CTM2 TEs). sRNA read counts were normalized to TE length. The siRNA levels from RdDM loci are also significantly higher than those from CMT2-only loci (both in Col-0 and *ndx1–4*; Statistics: Kruskal–Wallis test with Mann–Whitney test (for multiple comparison) and Benjamini–Hochberg correction. Sample size (n) is indicated. **g** The same as **f** but sRNAs were sized as 21, 22, 23, 24 nt siRNAs. In the 24nt class, there is a statistically significant difference between the expression status of *ndx1–4* and Col-0 samples. Bounds of boxes in **e**–**g** describe the interquartile range with the median; whiskers indicate minimum and maximum values; outliers are not shown. **h** Stem-loop rt-qPCR validation of increased sRNA production from Copia28 and MULE1 transposons (normalized to PP2AA3 expression). Error bar: SD. Sample size $n = 4$ biologically independent replicates.

pericentromeric, Knob, or arm association (Fig. 6e). We observed a significant increase in siRNA levels over RdDM TEs and CTM2 TEs such that higher expression was detected in the *ndx1–4* mutant relative to Col-0 (Fig. 6f; *ndx1–4* vs. Col-0 comparison). (We note that siRNA levels from RdDM TEs were substantially higher than those from CMT2 TE both in Col-0 and *ndx1–4* backgrounds, which is expected from the known mechanism of sRNA control, in agreement with published data[55].) When siRNAs were sized into 21, 22, 23, 24 nt classes, 24 nt siRNAs (representing het-siRNAs generated by Pol IV) were specifically upregulated in the *ndx1–4* mutant at both RdDM TEs and CMT2 TEs, as opposed to 21–22 nt siRNAs (so called epigenetically activated siRNAs, easiRNAs) that are related to TE expression (Fig. 6g). These associations further reinforce the link between NDX and Pol IV-dependent siRNA expression from TEs. The above siRNA changes were confirmed by stem-loop rt-qPCR at two representative transposons (Fig. 6h), suggesting that NDX is involved in transcriptional silencing of these loci. Interestingly, the pericentromeric ncRNA *siR1003*, expressed from the silenced 5 S rDNA array on chr3 (cytologically in the chromocenter), showed mildly increased siRNA levels by northern blot hybridization (Supplementary Fig. 6). This suggests that NDX not only affects the sRNA expression profile of transposons but may also silence other gene types in pericentromeric regions.

To elucidate whether the observed siRNA changes are associated with similar (or opposite) changes in the abundance of R-loops, we globally mapped RNA-DNA hybrids in the *ndx1–4* mutant (Fig. 7a). The DRIP-seq signal detected in both *ndx1–4* and Col-0 samples was reduced to background by RNaseH treatment, indicating the specificity of RNA-DNA hybrid detection. The amount of R-loops showed a moderate but statistically significant increase in the absence of NDX (fold change: >1.5; $p < 0.0001$), as evidenced by the higher intensity of DRIP peaks in the *ndx1–4* mutant, calculated by AUC analysis (Fig. 7b). Elevated genomic R-loop levels were also confirmed by slot blot hybridization (Fig. 7c, d) and DRIP-qPCR at selected genomic regions (Fig. 7f). Based on these data, NDX appears to prevent, rather than stabilize or stimulate, the formation of R-loops in the *Arabidopsis* genome. Moderately increased R-loop levels in *ndx1–4* could be due to an indirect consequence of the mutation rather than the direct effect of NDX binding, since R-loops and NDX follow opposite genomic distributions.

Loss of NDX results in global transcriptional changes that may directly or indirectly affect heterochromatin status

To assess the global effect of NDX loss on nuclear transcription, including euchromatin- and heterochromatin-derived transcripts, we performed an mRNA transcriptome analysis.

This analysis identified 1984 differentially expressed protein coding transcripts ($p < 0.01$) involving 864 up- and 1120 downregulated genes in the *ndx1–4* mutant compared to Col-0 control (Supplementary Fig. 7). GO term analysis showed that upregulated genes were primarily involved in the formation of ribonucleotide complexes, cold-response, flowering, and organ development, while downregulated genes were implicated in general stress response, post-embryonic development, lipid storage, and binding of ribonucleotides (Supplementary Fig. 8). Our results were highly correlated with previously published data[19].

We also analyzed the activity of transposons at a global level. The NGS results show that loss of NDX had a mild but significant effect on TE activity: 28 and 21 transposons were up- or downregulated in the *ndx1–4* mutant (DESeq2 analysis; $p < 0.05$; Supplementary Fig. 9). This may be an underestimate since mRNA-seq captures only a portion of active TEs due to poly(A) selection[56]. Notwithstanding, the number of reactivated transposons is comparable to the number of upregulated TEs detected in *cmt2* and *drm1/drm2* heterochromatin mutants[9].

In addition, we sought to find chromatin regulators that are differentially expressed in *ndx1–4*. Indeed, we identified a group of genes that may play a direct or indirect role in the regulation of heterochromatin status. Overexpressed genes involve known RNAi factors such as (1) NRPD1b, NRPE3b, and NRPB/D/E9a, which represent the structural and regulatory subunits of Pol IV and Pol V, the core *trans* factors of RdDM; (2) RDR1, which participates in non-canonical RdDM[57]; (3) IDN2, required for siRNA accumulation and binding to dsRNA and lncRNA[58]; (4) AGO9, which is normally expressed in the ovule to interact with siRNAs transcribed from pericentromeric retrotransposons[59]; (5) ROS1, which demethylates several genomic targets to restrict non-CG methylation activity[60]. These RdDM genes are typically silent in seedlings (or expressed at low level) but become induced at high levels in the *ndx1–4* mutant. Downregulated chromatin factors include HTA4 (histone H2A), HON4 (linker histone like protein), and HMGB1 (high mobility group B1), which are involved in the assembly of nucleoprotein complexes. RT-qPCR validation of representative genes is shown in Supplementary Fig. 10. Misregulation of the above heterochromatin regulators are likely to contribute to the molecular phenotype of *ndx1–4*, however, causative relationships remain to be explored.

NDX is required for proper non-CG methylation levels

Since loss of NDX influences heterochromatin behavior (siRNA accumulation at hundreds of genomic loci and TEs), we performed bisulfite sequencing (BS-seq) to reveal the involvement of NDX in DNA

https://doi.org/10.1038/s41467-022-32709-y

dc_1977_21

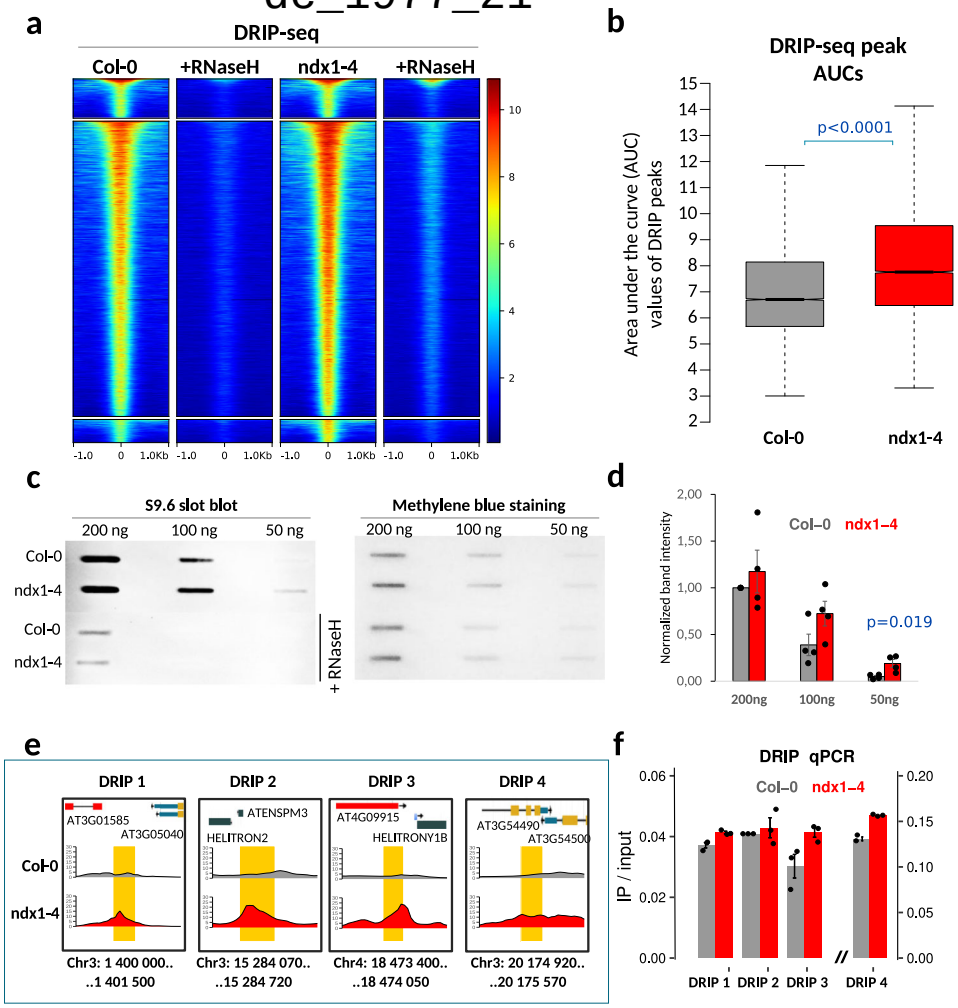


Fig. 7 | Loss of NDX is associated with increased rather than decreased R-loop levels. a Pile up plots showing DRIP signal intensities over the identified peaks in Col-0 and *ndx1–4* plants. RNase H-treated samples are also shown. Color scale corresponds to RPGC (reads per genomic content) values. **b** Statistical evaluation of DRIP peak intensities in Col-0 and *ndx1–4* plants. AUC (area under curve) values were calculated for each peak and their distributions were compared. AUCs, which are proportional to DRIP peak intensities, were significantly higher in the *ndx1–4* mutant ($p < 0.0001$, Mann–Whitney *U* test, two-sided). Bounds of boxes describe the interquartile range with the median; whiskers indicate minimum and maximum values; outliers are not shown. Sample size $n = 15{,}049$ peaks identified in Col-0 and *ndx1–4* plants. **c** Detection of RNA-DNA hybrids in Col-0 and *ndx1–4* plants using slot blot hybridization. 200 ng, 100 ng, and 50 ng of gDNA were slotted onto a nitrocellulose membrane with and without RNase H-treatment, stained with S9.6 antibody and goat anti-mouse-HRP secondary antibody. Equal loading was determined by methylene blue staining. **d** Band intensities were quantified by ImageJ (average values are shown). Error bars: SD. Statistics: Student's t test, two-sided. Sample size $n = 3$ biologically independent replicates. **e**, **f** Genome browser tracks and DRIP-qPCR validation of DRIP peaks identified in Col-0 (red) and *ndx1–4* (blue) samples. Error bar: SEM. Sample size $n = 2$ biologically independent experiments.

methylation (Supplementary Figs. 11, 12). Comparison of *ndx1–4* and Col-0 samples revealed 2449 hypomethylated and 1597 hypermethylated regions (DMRs) in the CHH/CHG sequence context (Fig. 8a), of which hypo-CHH/CHG sites were co-localized with NDX (Fig. 8b, c). In the CG context, 1353 hypo- and 1624 hyper-CG DMRs were detected, showing no enrichment for NDX binding (Fig. 8b). As expected, CHG/CHH DMRs significantly overlapped with TEs located in pericentromeric regions (Fig. 8d), and with chromatin states 8-9 that mark constitutive heterochromatin (Fig. 8e). In contrast, CG DMRs followed a more even distribution between the different annotation categories and chromatin states (Fig. 8d, e). These associations suggest that NDX influences transposon CHH/CHG methylation in pericentric heterochromatin regions. Grouping transposons by their regulatory classes and genomic positions highlighted decreased CHH and CHG methylation in CMT2 TEs in the *ndx1–4* mutant, regardless of pericentromeric or arm association (Fig. 8f). However, RdDM TEs showed reduced DNA methylation only in the CHG context located in chromosomal arms (mCG levels did not change in either class). Since NDX binds directly only to CMT2 TEs (Fig. 4c), the above changes are likely due to the direct effect of NDX, whereas CHG methylation changes observed in RdDM TEs may be the indirect effect of *ndx1–4* mutation. Importantly, sRNA expression changes detected at hypo CHH/CHG DMRs appear to be independent of NDX binding, since there was no difference between hypo CHH/CHGs classified as "NDX-enriched" and "non-enriched" (Supplementary Fig. 13). This suggests that the *ndx1–4* mutation plays an indirect role in the production of siRNAs at hypomethylated CHH/CHG sites, and that the DNA methylase and RNAi systems can function without NDX binding (underscoring their epistatic relationship).

To explore hierarchical relationships between the DNA methylome changes of *ndx1–4* and heterochromatin mutants, we analyzed BS-seq data from a collection of mutants using the hcDMR pipeline[61]. The identified high-confidence DMRs confirmed well-known genetic interactions (e.g., between mutants operating in the RdDM and CMT2 pathways; Supplementary Fig. 14) and also identified new associations between *ndx1–4*, *drm2*, and *met1* as these groups were clustered

https://doi.org/10.1038/s41467-022-32709-y

dc_1977_21

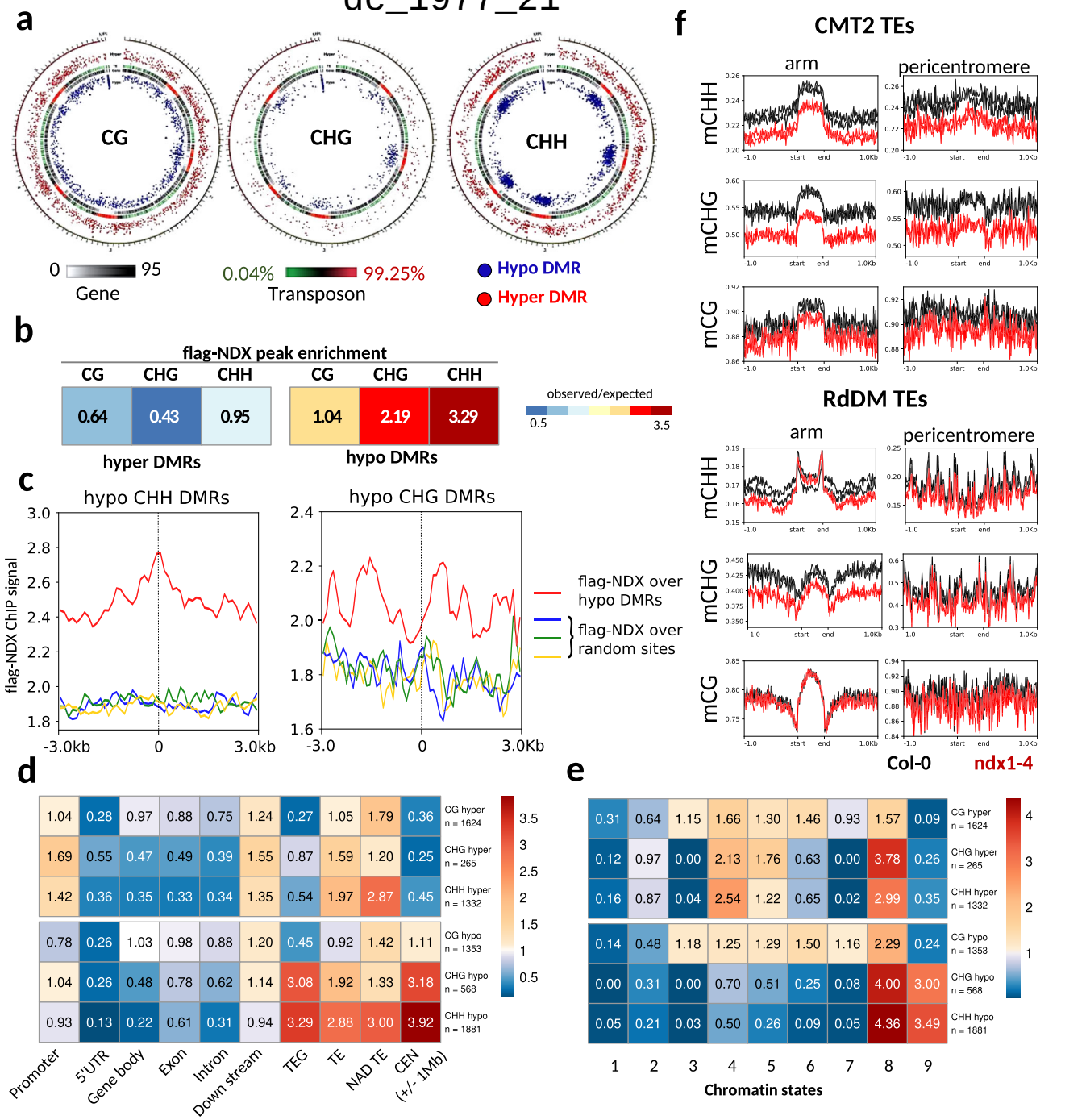


Fig. 8 | Non-CG DNA methylation is affected in the absence of NDX. a Circos plot showing the genomic distribution of differentially methylated regions (DMRs) in CG/CHG/CHH nucleotide contexts. Red and blue spots indicate hyper- and hypo-methylated DMRs in the *ndx1–4* mutant. Outer circle represents the five chromosomes and the mitochondrial and plastid genomes. Color and black-and-white heatmaps show gene density and transposon (TE) density. **b** Overlap of flag-NDX ChIP peaks with hyper- and hypo-DMRs. Cells contain observed/expected ratios for peak counts. Warmer colors represent higher enrichment. **c** Anchor plot of flag-NDX ChIP signal over hypo CHH and CHG DMRs (red line) and random sites (blue, green, yellow lines). The NDX signal is enriched in the middle of the CHH DMRs and near the center of hypo CHG DMRs. **d**, **e** Annotation of hyper- and hypo-DMRs on **d** functional genomic features and **e** chromatin states. Cells contain observed/expected ratios for peak counts. Warmer colors represent higher enrichment. **f** CMT2 TEs and RdDM TEs were classified by chromosomal location as pericentromeric and arm associated. Metaplots show average DNA methylation levels in CHH/CHG/CG contexts in Col-0 (black) and *ndx1–4* (red) plants. CHH and CHG methylation is significantly decreased CMT2 TEs regardless of their arm- or pericentromeric location (mCG levels do not change). CHG methylation levels show reduction at RdDM TEs located in chromosome arms. (mCHH and mCG levels do not change.).

together. However, we could not clearly determine if NDX influences non-CG methylation via the CMT2 or RdDM pathways, or, more possibly, by a less characterized independent pathway (e.g., Knot linked silencing). As DNA methylation pathways show significant redundancies, appropriate classification is not obvious in terms of NDX function. For instance, MET1 (the main CG methylase) and DRM2 (the main RdDM methylase) are both required to maintain CHH methylation in CMT2 targeted heterochromatin. The potential crosstalk between DRM2, MET1, and NDX remains to be explored by combining classical and molecular genetics.

dc_1977_21

Loss of NDX results in rearrangement of chromatin organization in pericentromeric regions

Given the marked changes in siRNA accumulation, RNA transcription, R-loop formation, and DNA methylation in the *ndx1–4* mutant, we asked whether the observed molecular phenotype is related to underlying chromatin structural changes. We therefore examined the genome-wide chromatin conformations of *ndx1–4* and Col-0 plants using in situ Hi-C, and sequenced ~200 million valid reads for each genotype. Normalized contact matrices resembled published Hi-C maps[62–64] and showed similar patterns for *ndx1–4* and Col-0 at first glance (Supplementary Fig. 15). This indicates that loss of NDX does not lead to extensive restructuring of three-dimensional chromosome architecture in *Arabidopsis*. Nevertheless, we found remarkable quantitative changes between the two Hi-C maps during differential analysis of contact matrices (Supplementary Fig. 16). Several genomic regions were intensified (red) or diminished (blue) in the *ndx1–4* mutant, indicating a global rearrangement of spatial interactions in the absence of NDX. Circos plot analysis of the identified differential contacts revealed the topography of weakened (blue) and enhanced (red) Hi-C interaction network, highlighting that intrachromosomal contacts are mostly decreased, while interchromosomal contacts tend to increase in the *ndx1–4* mutant (Fig. 9a). Another striking change was the rearrangement of the Knot structure, which consists of an entanglement of ten genomic regions in wild-type plants[62,65]. Knot-forming regions showed increased interchromosomal contact frequencies in *ndx1–4*, with the entry of new regions and parallel loss of other intra- and interchromosomal connections (Fig. 9b). Interestingly, the T-DNA insert inactivating the *NDX* locus was also anchored to the Knot in *ndx1–4* plants, consistent with recent Hi-C data showing that transgene integration can induce novel Knot interactions[65]. We also analyzed Hi-C data from available epigenetic mutants (*met1*, *ddm1*, *clf28 swn7*) and compared them to *ndx1–4*. The results show that the Hi-C pattern of *ndx1–4* was similar to that of the *met1* and *ddm1* mutants, however, overall interaction frequency changes were milder in *ndx1–4* (Fig. 9b and Supplementary Fig. 16b, c). Several new regions joined the Knot in all three mutants that do not appear in Col-0 (Fig. 9b); however, in the *ddm1* and *met1* mutants additional regions also interact with the Knot that are not present in *ndx1–4* (Fig. 9c). Therefore, chromatin structure changes observed in *ddm1* and *met1* are more severe than in *ndx1–4*. The *clf28 swn7* mutant lacks all H3K27me3[66] and previous Hi-C analysis showed that spatial interactions of H3K27me3-enriched minidomains were reduced in *clf28 swn7*, however, no Hi-C change was found in the Knot[63]. Our analysis gave similar results as the Hi-C pattern of the *clf28 swn7* mutant was different from all three mutants (Supplementary Fig. 16d) and the de novo Knot interactions were absent from *clf28 swn7*. Therefore, in terms of genome organization, *clf28 swn7* appears to operate in a different pathway than *ddm1*, *met1*, and *ndx1–4*, which all showed somewhat similar Knot interaction patterns. These associations reinforce the link between NDX and DNA methylation pathways. Another important Hi-C change in *ndx1–4* occurred at pericentromeric regions, with a general decrease in local intra-centromeric interactions (Fig. 9e, upper panel) and an increase in inter-centromeric interactions (except for CEN chr1/CEN chr4, Fig. 9e, lower panel). The same trend was true for all intra- and inter-chromosomal interactions of centromeres that were significantly weakened (blue) or enhanced (red) in the *ndx1–4* mutant (Fig. 9f). We propose that reduced intracentromeric interactions in the absence of NDX results in chromatin decompaction at pericentromeric regions (where NDX is enriched in wild-type plants), which strongly correlates with the location of sRNA expression changes and CHH/CHG methylation changes observed in the *ndx1–4* mutant. Whether the above chromatin changes are directly or indirectly mediated by NDX remains unclear. Nevertheless, chromatin compactness analysis[67] suggests that NDX has a direct effect on the condensation state of hypomethylated CHH regions (Supplementary Fig. 17) as NDX-enriched hypo-CHHs show significantly higher Hi-C compactness than hypo CHHs not bound by NDX (in wild-type plants). This indicates that NDX binding at these sites can directly promote chromatin condensation, which is inversely changed in the *ndx1–4* mutant due to the loss of NDX (see Hi-C data).

Finally, it should be noted that the chromatin binding of NDX does not scale with 3D chromatin compactness, i.e., NDX shows preferential enrichment in pericentromeric heterochromatin that is not due to the condensed state or high local DNA concentration of this chromatin type. To demonstrate this, we compared the Hi-C compactness of pericentromeric, Knob, and chromosomal arm regions in terms of NDX binding (Supplementary Fig. 18). The analysis showed that the compactness of all three regions differ significantly such that the most condensed is the heterochromatic Knob region, followed by pericentromeres and chromosome arms. Pericentromeric and Knob regions appear to bind more NDX per unit length than chromosome arms, however, the Knob showed similar NDX enrichment as pericentromeric regions despite its greater compactness. This suggests that NDX binding is disproportionate to compactness. When comparing regions with similar compactness (by random sampling from the above regions), chromosome arms still showed significantly lower NDX enrichment despite their same condensation state (Supplementary Fig. 18c). It follows that NDX preferentially interacts with constitutive heterochromatin, regardless of local chromatin density.

Discussion

Heterochromatin is an essential structural feature of eukaryotic genomes, conferring special functional properties on different chromosomal regions[68]. The ability of heterochromatin to restrain DNA recombination events and limit the activity of transposons is critical to maintaining genomic stability. Therefore, it is crucial to identify specific heterochromatin regulators that control the formation and maintenance of repressive chromatin states associated with DNA methylation pathways, siRNA biogenesis, histone modifications, and 3D chromatin structure.

The homeodomain protein NDX was previously described as a transcriptional and/or epigenetic regulator of two euchromatic genes (*FLC*, *ABI4*) located in transcriptionally active chromosome arms[19,21]. In this work, we show that NDX is primarily a heterochromatin regulator that functions in pericentromeric regions to control the production of het-siRNAs and deposition of repressive CHH/CHG DNA methylation. There are multiple lines of evidence to support these claims: (i) ChIP-seq measurements using two different tags in independent transgenic lines (N-terminal flag and C-terminal GFP) show that NDX is strongly associated with pericentromeric regions; (ii) quantitative microscopy in live cells consistently show that NDX is a chromatin-binding factor with very slow nuclear dynamics, stably incorporated into peripheral heterochromatin; (iii) NDX preferentially associates with pericentromeric het-siRNA loci involved in non-CG methylation pathways, with significant sRNA transcriptional changes; (iv) CHH/CHG hypomethylation of pericentromeric regions in *ndx1–4* significantly overlaps with NDX binding sites; and (v) loss of NDX function results in extensive 3D chromatin structural changes in pericentromeric regions.

The previously described NDX-associated *FLC* and *ABI4* loci are found in heterochromatin islands interspersed in euchromatic genomic regions[19,21]. Consequently, NDX may also control the activity of these “heterochromatin-marked” euchromatic genes, consistent with its proposed role as a heterochromatin regulatory protein. Integrating the NDX cistrome data with transcriptomic/epigenomic maps is expected to reveal similar mechanisms for other euchromatic genes. However, our DRIP-seq data strongly argue against the notion that NDX is a common factor that directly regulates widespread genomic R-loops. Instead, NDX appears to negatively regulate R-loop formation throughout the *Arabidopsis* genome, in contrast to its previously proposed positive role as an R-loop stabilizing factor at the *FLC* locus[21].

https://doi.org/10.1038/s41467-022-32709-y

dc_1977_21

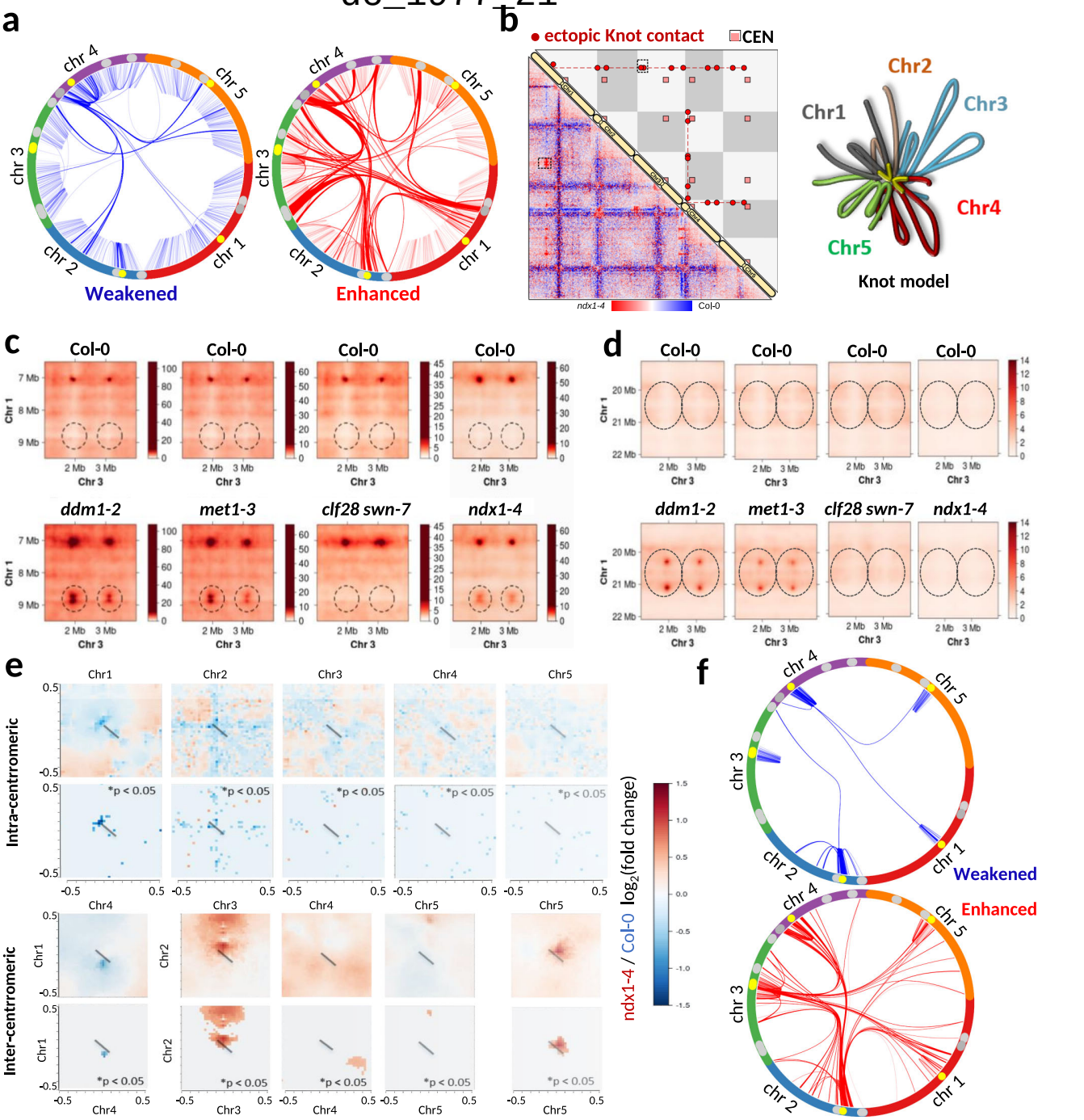


Fig. 9 | Loss of NDX function causes rearrangement of chromatin organization. **a** Circos plot representation of intra- and interchromosomal interactions that are significantly weakened (left) or enhanced (right) in *ndx1–4* ($p < 0.05$, Slitherine pipeline). Color intensity of the arcs is proportional to the fold-change in interaction frequencies (*ndx*1–4/Col-0). Outer circle represents the five *Arabidopsis* chromosomes. Centromeres are highlighted in yellow. Knot engaged elements (KEEs) are shown in gray. **b** Knot interactions in the absence of NDX. Diagonal: five *Arabidopsis* chromosomes. Lower left triangle (below the diagonal): differential Hi-C matrix calculated from the ratio of *ndx1–4* (red) and Col-0 (blue) interaction frequencies in Juicer. Resolution: 250 kb. Upper right triangle (above the diagonal): scheme of Knot interactions in the *ndx1–4* mutant. Knot-associated regions (circles) and centromeres (squares) and are highlighted. The T-DNA insertion site inactivating the *NDX* locus (red line in chr4) also participate in forming an ectopic Knot interaction. Right panel: Proposed 3D model of the Knot in the *ndx1–4* mutant. The structure involves de novo formation of inter-chromosomal contacts (highlighted in yellow). **c**, **d** Representative Knot contacts in *ddm1-2*, *met1-3*, *clf28 swn7*, and *ndx1–4* mutants. Panel c shows de novo formed Knot contacts in the *ddm1-2*, *met1-3*, and *ndx1–4* mutants that are missing from Col-0 and *clf28 swn7*. Panel d shows a *ddm1-2* and *met1-3* specific Knot contact that is missing from *ndx1–4*. **e** Pericentromeric chromatin shows local decompaction and enhanced inter-chromosomal interactions in the absence of NDX. Upper: Differential intra-centromeric Hi-C interactions represented by a pileup plot. Red and blue colors show enrichment or depletion of intrachromosomal interactions of centromeres in the *ndx1–4* mutant. Diagonal bars show the exact locations of centromeric regions. Significantly different regions are shown below (identified by Slitherine; permutation test; p-value adjustment: Benjamini & Hochberg (FDR) method). Lower: The same as above but differential inter-centromeric interactions are shown. **f** Circos plot showing all intra- and interchromosomal interactions of centromeric regions that are significantly weakened or enhanced in the *ndx1–4* mutant ($p < 0.05$, Slitherine pipeline). Color intensity of the arcs is proportional to the fold change in Hi-C interactions (*ndx1–4*/Col-0). Outer circle represents the chromosomes. Centromeres are highlighted in yellow.

dc_1977_21

Because NDX binding sites and R-loops follow an antagonistic genomic distribution, we propose that NDX inhibits R-loop accumulation indirectly. Increased global R-loop levels in *ndx1–4* is likely due to the secondary effect of the mutation, but exact molecular mechanisms remain to be explored. One possibility is that an aberrant Pol II activity or overloading of R-loop resolving pathways may cause increased R-loop levels in the absence of NDX. This is supported by the association of R-loops with gene promoters and transcription terminators. Another possibility arises from the fact that plants contain vast amounts of free RNA molecules[69] that are capable of generating unscheduled R-loops under certain conditions. We speculate that upregulated het-siRNAs may be incorporated into chromatin as homologous RNA-DNA hybrids, resulting in an overall increase in the number of R-loops. Accordingly, when searching for sequence homologies between upregulated siRNAs and R-loops (using the psRNATarget algorithm[70]), we found a significant correlation in the number of siRNA targets homologous to R-loops compared to random sequences ($p < 0.0001$; Supplementary Fig. 19). This is consistent with the observation that class 1–2 sRNAs (where most R-loops are located) do consist not of true sRNA loci, but are targets for other sRNAs acting in trans[51]. These associations support a model in which upregulated siRNAs hybridize to homologous target sites that in turn raise global R-loop levels. The mechanistic link between NDX activity, siRNA biogenesis, and R-loop production awaits further experimental validation.

Another important question concerns the mechanism of action of NDX and potential causal relationships between structural and functional changes induced by loss of NDX. In *Drosophila*, the NDX-related zeste protein has the capacity to self-associate to form protein aggregates, consisting of several hundreds of monomers[71]. This self-assembly mechanism may also apply to *Arabidopsis* NDX with a potential impact on higher-order chromatin architecture. Based on our Hi-C data, NDX appears to mediate long-range chromatin interactions through chromocenters and transcriptionally silent chromatin. The chromatin structural changes in pericentromeric regions are probably specific for *ndx1–4* plants, as no similar genome architecture rearrangements were detected so far in other *Arabidopsis* mutants. However, the identified Knot interactions reflect genome structural changes observed in epigenetic mutants involving DNA methylation (e.g., *ddm1*, *met1*)[62,63,65], which further strengthen the link between NDX and DNA methylation pathways. It seems relevant in this regard that heterochromatin has been shown to induce significant chromosome compaction at centromeres in *Sc. pombe*, providing basic mechanical constraints for proper genome structure and function[72].

Based on our data, the following model is proposed to summarize the function of NDX in heterochromatin homeostasis: (i) inactivation of NDX results in chromatin decompaction at highly condensed pericentromeric regions (by an as yet unknown mechanism), (ii) relaxed heterochromatin structure leads to het-siRNA accumulation and DNA methylation changes, (iii) consequently, a subset of chromatin regulatory genes and transposons become derepressed or repressed. Increased global R-loop formation is likely due to the collateral effect of NDX loss and chromatin conformational changes. Whether the NDX-mediated nuclear chromatin organization is related to the CMT2 pathway (as may be suggested by some of our observation, Figs. 4c, 8f) or represents a completely independent route, needs to be further investigated. Genetic analysis of the above and other factors is expected to lead to a better understanding of heterochromatin homeostasis as the driver of genome organization and stability and its response to developmental signals and environmental stimuli.

Methods

Plant material

The following *Arabidopsis thaliana* genotypes were used in our study: Col-0, *ndx1–4*, flag-NDX (genotype: flag-NDX/*ndx1-1*(FRI)/*flc-2*), and NDX-GFP (genotype: NDX-GFP/*ndx1-1*(FRI)/*flc-2*). The NDX fusion proteins were expressed from their endogenous promoter. All transgenic plant lines were described previously[21,73]. Seeds were sterilized and sown on agar-MS plates, kept at 4 °C in dark for 2 days. After stratification, plates were placed to 21 °C, long day (LD, 16 h light, 8 h dark) conditions for 10 days. For immunofluorescence, NDX-GFP seedlings were grown for 10 days (14 h light, 10 h dark at 16 °C).

FRAP and FCS microscopy

FRAP measurements were performed using an Olympus FluoView 1000 confocal microscope, based on an inverted IX-81 stand with an UPlanAPO 60 × 1.2-numeric-aperture (NA) water immersion objective. GFP was excited by the 488-nm argon-krypton laser line and fluorescence was detected through a 500–550 nm band-pass filter. For the optimal bleach depth, laser power was set to maximal intensity (~900 μW) with 500 ms duration. Cells were selected randomly in the root tip region of the plant and were bleached (in rectangular ROIs for 25 seconds) separately either in the nucleolus or in the nuclear periphery, where the NDX-GFP protein was highly expressed. Post-bleach images were acquired in 30 s intervals using low laser intensities. FRAP curves were normalized using Phair's double-normalization protocol[34,74]. In the FCS measurements, we recorded images and carried out autocorrelation analysis (at room temperature) within selected femtoliter-sized confocal volumes using an Olympus FluoView 1000 confocal microscope combined with FCS module[75]. NDX-GFP molecules passing through the confocal volume were excited with the 488 nm line of an Ar-Kr laser with minimized laser intensity (~0.9 μW). The emitted photons were detected from 500 nm to 550 nm. Fluorescence autocorrelation curves were calculated online by an ALV-5000E correlator card, which recorded the time course of the signal and calculated the autocorrelation function real-time. Data acquisition time was set to 8 × 10 s at each selected point. Autocorrelation analysis was performed by the QuickFit 3.0 software (Krieger, Jan; http://www.dkfz.de/Macromol/quickfit/) applying a 3D normal diffusion model for two-component fitting. The triplet state formation and slower dark states due to protonation were considered in the model function[76].

DNA-RNA hybrid immunoprecipitation

2 g of 10-day-old seedlings of Col-0, and *ndx1–4* plants were harvested, and cross-linked under a vacuum chamber by phosphate buffered saline (PBS) containing 1% formaldehyde for 2 × 5 min. Excess formaldehyde was quenched by adding glycine to a final concentration of 0.125 M. Chromatin was purified and resuspended in Nuclei lysis buffer (50 mM Tris-HCl, 10 mM EDTA, 20 mM Hepes KOH pH 7.4, 10 mM $MgCl_2$, 0.5% Triton X-100, 5 mM DTT and Protease Inhibitors) followed by Proteinase K treatment (Thermo Fisher Scientific) and crosslink-reversal in the same step, at 65 °C, overnight. Total nucleic acid was prepared by phenol-chloroform extraction and precipitated with 2-volumes of ice cold isopropanol and 1/10 volume 4 M ammonium-acetate. Samples were then divided into two equal parts: half of the samples were treated with 8 μl of RNaseH (5000 U/ml, New England Biolabs) at 37 °C, overnight. Nucleic acid preps were diluted 4-fold in a ChIP dilution buffer (16.7 mM Tris-HCl pH 8, 1.2 mM EDTA, 1.1% Triton X-100, 167 mM NaCl) and DNA was fragmented by sonication to an average fragment size of 250–500 bp (low intensity, 3 ×5 min; 30 sec ON, 30 sec OFF, Bioruptor UCD-300, Diagenode). Dynabeads Protein A magnetic beads (Thermo Fisher Scientific) were pre-blocked with PBS/EDTA containing 0.5% BSA. The RNA-DNA hybrid specific S9.6 monoclonal antibody was prepared from the Hb-8730 mouse hybridoma cell line (Atcc) using protein A/G purification and stored in PBS/0.05% (w/v) Sodium Azide buffer. To immobilize the S9.6 antibody, 50 μl pre-blocked Dynabeads Protein A slurry was incubated with 10 μg of S9.6 antibody in IP buffer (50 mM Hepes/KOH at pH 7,5; 0,14 M NaCl; 5 mM EDTA; 1% Triton X-100; 0,1% Na-Deoxycholate, ddH_2O) at 4 °C for 4 h, with rotation. Six micrograms of fragmented genomic DNA were added to the mixture and gently rotated at 4 °C, overnight. Beads were

dc_1977_21

washed two times with 1 ml of ChIP lysis buffer (low salt, 50 mM Hepes/KOH pH 7.5, 0.14 M NaCl, 5 mM EDTA pH 8, 1% Triton X-100, 0.1% Na-Deoxycholate), 1 ml of high-salt ChIP lysis buffer (50 mM Hepes/KOH pH 7.5, 0.5 M NaCl, 5 mM EDTA pH 8, 1% Triton X-100, 0.1% Na-Deoxycholate), 1 ml of wash buffer (10 mM Tris-HCl at pH 8, 0.25 M LiCl, 0.5% NP-40, 0.5% w/v sodium deoxycholate, 1 mM EDTA at pH 8) and 1 ml of TE buffer (100 mM Tris-Cl at pH 8, 10 mM EDTA at pH 8) at 4 °C. Elution was performed in elution buffer (50 mM Tris-HCl pH 8, 10 mM EDTA, 1% SDS) for 15 min at 65 °C. After purification by a NucleoSpin Gel and PCR Clean-up kit (Macherey-Nagel), DRIP samples were analyzed by quantitative real-time PCR (qPCR) performed with a LightCycler 480 SYBR Green I Master mix (Roche) and a QuantStudio 12 K Flex Real-Time PCR System (Thermo Fisher Scientific). Primer sequences are listed in Supplementary Data 14. qPCR data were analyzed using the comparative CT method. RNA-DNA hybrid enrichment was calculated based on the IP/Input ratio.

Chromatin immunoprecipitation (ChIP)

2 g of 10-day old flag-NDX (flag-NDX/*ndx1-1* (FRI)/*flc-2*), NDX-GFP (NDX-GFP/*ndx1-1*(FRI)/*flc-2*), and *ndx1–4* seedlings (no tag control) were harvested, and cross-linked under a vacuum chamber for 2 × 5 min in PBS containing 1% formaldehyde. Excess formaldehyde was quenched by adding glycine to a final concentration of 0.125 M. Nuclei were isolated and lysed in 300 µl of Nuclei lysis buffer (50 mM Tris-HCl, 10 mM EDTA, 20 mM Hepes KOH pH 7.4, 10 mM $MgCl_2$, 0.5% Triton X-100, 5 mM DTT and Protease Inhibitors) followed by a 1-minute centrifugation. The pellet was resuspended in Nuclei lysis buffer and the chromatin was fragmented by sonication to an average fragment size of 250–500 bp (low intensity, 3 × 5 min; 30 sec ON, 30 sec OFF, Bioruptor UCD-300, Diagenode). Immunoprecipitation and reverse crosslinking were performed as described previously[39]. For ChIP, 2 µg of antibodies against flag and GFP epitopes were used (anti-flag antibody - #2044 New England Biolabs; anti-GFP antibody - #ab290 Abcam). Dynabeads Protein A magnetic beads (Thermo Fisher Scientific) were pre-blocked with PBS/EDTA containing 0.5% BSA. To immobilize the anti-flag or anti-GFP antibody, 50 µl pre-blocked Dynabeads Protein A slurry was incubated with 2 µg of anti-flag or anti-GFP antibody in IP buffer (50 mM Hepes/KOH at pH 7,5; 0,14 M NaCl; 5 mM EDTA; 1% Triton X-100; 0,1% Na-Deoxycholate, ddH_2O) at 4 °C for 4 h, with rotation. Six micrograms of fragmented chromatin were added to the mixture and gently rotated at 4 °C, overnight. Beads were washed two times with 1 ml of ChIP lysis buffer (low salt, 50 mM Hepes/KOH pH 7.5, 0.14 M NaCl, 5 mM EDTA pH 8, 1% Triton X-100, 0.1% Na-Deoxycholate), 1 ml of high-salt ChIP lysis buffer (50 mM Hepes/KOH pH 7.5, 0.5 M NaCl, 5 mM EDTA pH 8, 1% Triton X-100, 0.1% Na-Deoxycholate), 1 ml of wash buffer (10 mM Tris-HCl at pH 8, 0.25 M LiCl, 0.5% NP-40, 0.5% w/v sodium deoxycholate, 1 mM EDTA at pH 8) and 1 ml of TE buffer (100 mM Tris-Cl at pH 8, 10 mM EDTA at pH 8) at 4 °C. Elution was performed in elution buffer (50 mM Tris-HCl pH 8, 10 mM EDTA, 1% SDS) for 15 min at 65 °C. Reverse cross-linking was carried out by adding 200 mM NaCl and 5 mM EDTA to the eluted product, incubated at 65 °C, overnight. After RNA and protein digestion by RNaseA and Proteinase K, DNA was isolated by a NucleoSpin Gel and PCR Clean-up kit (Macherey-Nagel). QPCR validations were performed by quantitative real-time PCR (qPCR) performed with a LightCycler 480 SYBR Green I Master mix (Roche) and a QuantStudio 12 K Flex Real-Time PCR System (Thermo Fisher Scientific).

ChIP-seq and DRIP-seq library preparation

ChIP-seq and DRIP-seq libraries were prepared according to Illumina's TruSeq ChIP Sample Preparation protocol except that the uracil tolerant KAPA HiFi HotStart DNA Polymerase (Kapa Bio) was used. Briefly, the enriched DRIP and ChIP DNA samples were end-repaired and indexed adapters were ligated to the inserts. Purified ligation products were then amplified by PCR. Libraries were sequenced using 150 nt paired end read with Illumina HiSeq 2500 (EMBL Genomics Core Facility, Heidelberg, Germany) and 150 nt paired end read with Illumina NextSeq500 (Genomic Medicine and Bioinformatics Core Facility of the University of Debrecen).

ChIP-seq and DRIP-seq data analysis

Sequenced reads were aligned to the *A. thaliana* reference genome (TAIR10; NCBI; Ecotype: Columbia-0) using the bowtie2[77] algorithm. Samtools[78] was used for making.bam files and indexing. Low mapping quality and PCR-duplicated reads were omitted from downstream analysis (http://broadinstitute.github.io/picard/). Deeptools[79] bamCoverage was used to create RPKM (Reads Per Kilobase per Million mapped reads) normalized coverage files (.bedgraph and.bigwig). Read densities were calculated for 20 bp bins (–binSize 20–operation ratio–smoothLength 60 –normalizeUsing RPKM) and IP/input bigwig ratios were made by bigwigCompare. Heatmaps were generated for 10 bp bins with computeMatrix and were plotted by plotHeatmap functions of deepTools. For metagene profiles, mean normalized read coverage were calculated in 20 bp windows with computeMatrix and were plotted using plotProfile (deepTools). MACS2[80] was used with default settings to identify ChIP peaks in input normalized flag-NDX, NDX-GFP, and DRIP peaks in input normalized and RNaseH-corrected Col-0 and *ndx1–4* samples, respectively. The identified peaks are listed in Supplementary Data 2–5. ChIP and DRIP data were visualized in JBrowse.

Genomic annotation of DRIP and ChIP peaks

Enrichment of ChIP and DRIP peaks over functional genomic categories was performed by bedtools intersectBed to calculate overlap ratios[81]. Random peak sets were generated using the bedtools shuffleBed function. Observed/expected (random) ratios were plotted over genomic categories as a heatmap (in R project). The analyzed genomic categories were downloaded from Araport11 (.gff3 file). List of sRNA loci and sRNA clusters[51], mircoRNAs (https://www.mirbase.org/), chromatin states[50] were downloaded from the corresponding publications.

Calculating ssDNA forming propensity

Propensity to form single-stranded DNA was estimated for NDX binding sites and RNA-DNA hybrids. We took a random sample of flag-NDX peaks, DRIP peaks and common (NDX/DRIP) peaks ($n = 300$ each) and then resized each peak to 400 bp. We also selected 400 bp long random regions as control ($n = 300$). We used the command line version of mfold software[82] to predict secondary structures for the nucleotide sequences of each peak. Overall ssDNA forming propensity was defined as the average proportion of individual bases adopting a single-stranded conformation, considering all computed models.

Small RNA sequencing

Total RNA was purified from Col-0 and *ndx1–4* seedlings (3-3 biological replicates) using phenol-chloroform extraction. The quality of total RNA samples was checked by an Agilent bioanalyzer (RIN > 9). Small DNA libraries were prepared according to Illumina's NEBNext® Multiplex Small RNA Library Prep protocol. sRNA-seq libraries were sequenced using an Illumina NextSeq500 instrument with 1 × 50 bp reads. The results were analyzed using the sRNAnalyzer pipeline[83] as follows: Illumina adapters were trimmed by Cutadapt and reads were then size selected in the 19–25 nt range. Using sRNAnalyzer, size selected reads were aligned to the miRbase database (https://www.mirbase.org/) and a recently published comprehensive sRNA locus database[51]. Differential sRNA expression between Col-0 and ndx1–4 samples was determined by DEseq2 ($p < 0.05$, abs(log2(fc)) >log2(1.5)) and results were plotted in R. For sRNA target site prediction of differentially expressed sRNA reads over the identified DRIP-

dc_1977_21

seq peaks, the psRNATarget algorithm was applied using default settings[70].

Stem-loop qRT-PCR analysis of sRNA expression

For sRNA qRT-PCR, we treated 5 µg total RNA with DNaseI (NEB, M0303) according to the manufacturer's instructions. RNA samples were precipitated in absolute ethanol and resuspended in sterile water. One microgram of total RNA was reverse transcribed using an sRNA-specific primer and U6- or PP2AA3 reference RNA specific primer according to the manufacturer's instructions (NEB, E6560). First, the stem-loop RT primer was hybridized to the sRNA molecule and then reverse transcribed in a pulsed RT reaction. Next, the RT product was PCR amplified using an sRNA-specific forward primer and a universal reverse primer. Specific primers for LTR/Copia28 and MULE1 transposon-derived sRNA molecules and reference RNA were designed according to Varkonyi-Gasic and Hellens (2010). qPCRs were performed by a LightCycler 96 Real-Time PCR machine (Roche) using Master Mix NEB M3003. Data was processed with a LightCycler® 96 software (Version 1.1.0.1320). The specific qPCR primers for sRNAs and reference RNA detection can be found in[84].

mRNA sequencing (RNA-seq) and rt-qPCR

Total RNA was purified from Col-0 and *ndx1–4* seedlings using the phenol-chloroform extraction method. The quality of total RNA samples was checked by an Agilent bioanalyzer (RIN > 9). cDNA libraries were prepared according to Illumina's TruSeq RNA Library Preparation Kit v2 protocol. NGS libraries were sequenced using an Illumina MiSeq instrument with 1 × 50 bp reads (Biological Research Center of Szeged, Hungarian Academy of Sciences). We used Salmon to map RNA-seq reads and get transcript quantities for genes and transposable elements. Reads were mapped to the *A. thaliana* reference transcriptome (Araport11) to quantify gene expression whilst reads were mapped to transposable elements (Araport11) themselves to estimate their expression. using Salmon. Transcript quantities were corrected for GC bias to reduce isoform quantification errors. Differentially expressed genes and differentially expressed transposable elements (ndx1–4 vs. Col-0 wild type) were identified by DESeq2. The level of significance was defined by the adjusted *p*-values with independent hypothesis weighting[85] at *p*(adjusted) <0.01 for differentially expressed genes and at *p*(adjusted) <0.05 for differentially expressed transposable elements (DESeseq2). For data visualization, RNA-seq reads were aligned to the TAIR10 reference genome using HISAT2 allowing for reporting spliced alignments. We used deepTools bamCoverage to create RPKM (Reads Per Kilobase per Million) normalized bedgraph files. For rt-qPCR validation of differential mRNA expression changes, contaminating DNA was eliminated from total RNA preps by DNaseI digestion (RQ1 RNase-Free DNase, Promega). DNA-free RNA samples were reverse transcribed with random hexamers using SuperScript IV reverse transcriptase (Thermo Fisher Scientific). Real time qPCR was performed with a LightCycler 480 SYBR Green I Master mix (Roche) using a QuantStudio 12 K Flex Real-Time PCR System (Thermo Fisher Scientific). mRNA expression levels were normalized to UBQ10 gene expression.

Gene ontology (GO) analysis

Singular enrichment analysis (SEA) of differentially expressed genes was performed by agriGO. Plant GO slim terms with a probability of $p < 0.01$ were called significant.

Northern blot

For sRNA northern blot hybridizations, total RNA was extracted from 30 mg 10-day old seedlings. Plant material was homogenized, resuspended in an extraction buffer (0.1 M glycine-NaOH, pH 9.0, 100 mM NaCl, 10 mM EDTA, 2% SDS). The extracts were treated with phenol pH 4,3, phenol-chloroform and chloroform, precipitated in ethanol and resuspended in sterile water. 30 µg of total RNA was separated by 12% PAGE (8.6 M urea, 1× Tris-borate-EDTA). The RNA was transferred onto Hybond-NX membranes and fixed by chemical crosslinking. The membrane was sequentially hybridized with 32P-ATP end-labeled (T4 PNK, NEB) cDNA oligos (IDT) that detect siR1003 siRNA and U6 snRNA, respectively. Quantification of band intensities was performed by ImageJ such that siR1003 expression was normalized to U6 levels.

S9.6 slot blot

Genomic DNA was purified from 14-old Col-0 and *ndx1–4* seedlings using phenol/chloroform extraction. 50, 100, and 200 ng of gDNA preps were slotted onto nitrocellulose membranes with and without RNaseH treatment (Hybond-N + , GE Amersham) using a Bio-Dot SF Microfiltration Apparatus. Membranes were crosslinked with UV(2000) two times, and blocked in milk-TTBS (5% milk in 1× TTBS; 17 mM Tris, 130 mM NaCl, PH7.5, 1%Triton-X-100) for 1 h at room temperature. The S9.6 antibody was diluted in milk-TTBS at 1:1500 and membranes were incubated overnight at 4°C. Membranes were rinsed in milk-TTBS 3 times, each time for 10 min. The secondary antibody (goat anti-mouse-HRP) was added for 1 h at room temperature, followed by washes in milk-TTBS for 5 min and 1× TTBS 3 times, each time for 10 min. The S9.6 signal was detected by ECL reagent on the film. Equal loading of samples was determined by methylene blue staining of the membranes.

Whole genome bisulfite sequencing (BS-seq)

Genomic DNA was prepared by standard phenol-chloroform extraction from 14-old seedlings (Col-0 and ndx1–4). 1 µg of gDNA was sent for whole genome bisulfite sequencing (BS-seq) to Novogene Ltd. Two independent biological replicates were analyzed from each background. DNA samples were fragmented into 200–400 bp using Covaris S220 and end-repaired, dA-tailed and ligated to sequencing adaptors containing only methylated cytosines. Then the DNA fragments were sodium-bisulfite treated with EZ DNA Methylation Gold Kit (Zymo Research) after which cytosines without methylation changed to U (after PCR amplification to T), while cytosines with methylation remained unchanged. 1 ng/l of PE150 library was prepared and sequenced on an Illumina NovaSeq 6000 instrument resulting in 15–17 million raw reads. The Bismark software[86] was used to align the bisulfite-treated reads to the TAIR10 reference genome. For the methylated sites, the methylation level is calculated using the following formula: ML = mC/(mC+umC), where ML represents the methylation level, mC and umC represent the number of methylated and unmethylated cytosines, respectively. Methylation levels were determined for CpG, CHH, and CHG sequence contexts as percentage of methylated cytosines in their contexts and integrated with JBrowse. Differentially methylated regions (DMRs) were identified by the DSS-single (DSS) pipeline considering the variance among biological replicates. To compare the DNA methylome changes of *ndx1–4* plants to a large collection of DNA methylation mutants functioning in different pathways, we used the hcDMR pipeline[61]. High-confidence DMRs were identified by comparing the methylation changes of each mutant to the methylation levels of 54 control libraries. HcDMRs were clustered with the S-MOD method (statistical measurement of overlapping of DMRs), allowing the identification of hierarchical relationships between *ndx1–4* and DNA methylation mutants. SRA IDs of the mutants used in the analysis are listed in Supplementary Data 15.

In situ Hi-C

Hi-C experiments were performed by the Arima HiC Kit (Arima Genomics). 10-day old Columbia wild type (Col-0) and *ndx1–4* seedlings were crosslinked with 1% formaldehyde as described in DRIP-seq experiments. According to the Arima User Guide for Plant tissue (https://arimagenomics.com/), 1 g of crosslinked plant tissue was used as a starting material for isolation of nuclei. Plant samples were ground

dc_1977_21

in liquid nitrogen using a mortar. 20 ml of PTNI buffer was then added (250 mM Sucrose, 20 mM HEPES pH 8.0, 5 mM KCl, 1 mM MgCl2, 40% glycerol, 0.1 mM PMSF, 1% protease inhibitor cocktail (Sigma), 0.25% Triton X-100, 0.1% mercaptoethanol) and samples were purified two times by using a double-layer Mirachlot (Merk). Nuclei were washed several times in PTNI buffer as described in the Arima-HiC User Guide. Proximity ligated in situ Hi-C libraries were then constructed according to the Arima User Guide. Hi-C samples were fragmented to an average size of 400 bp (Bioruptor UCD-300, Diagenode; low intensity mode, 3 × 5 min; 30 sec ON, 30 sec OFF). NGS libraries for Illumina sequencing were prepared using the Accel-NGS 2 S Plus DNA Library Kit and Accel-NGS 2 S Indexing Kit (Swift Biosciences). To estimate the proper number of PCR cycles for library amplification, Arima-QC2 values were determined using the KAPA Library Quantification Kit (Roche). NGS library amplifications were performed until the estimated PCR cycle numbers (usually between 5 and 9) using the Kapa Library Amplification Kit (Roche). NGS libraries were then sequenced on an Illumina NextSeq 500 platform (2 × 150 nt, paired end reads) using a NextSeq 500 High Output v2 kit (Illumina). Raw Hi-C libraries (.fastq files) were processed by the Juicer toolbox (Durand et al. 2016) using default parameters, except that appropriate restriction enzyme cutting sites - provided by the Arima-HiC kit (GATC and GANTC) - were introduced into the script files. Hi-C reads were mapped to the *A. thaliana* TAIR10 reference genome and MAPQ ≥ 30 reads were retained for further analysis. Hi-C alignment statistics is summarized in Supplementary Data 12. We used the HiGlass tool[87] for visual exploration and analysis of the interaction maps. Quantitative differences in Hi-C interactions (Col-0 (wild type) vs. *ndx1–4* mutant) were identified at 25 kb resolution by the Slitherine pipeline (https://gitlab.pasteur.fr/gmillot/slitherine). Slitherine runs the Serpentine tool[88] to smooth local noise in Hi-C interaction maps and identifies statistically significant differences between Hi-C contact matrices. Regions that have significantly different Hi-C interactions in wild type vs. *ndx1–4* mutant are listed in Supplementary Data 13. For chromatin compactness analysis, we calculated the Hi-C condensation states of chromosomal regions following the method of Zhu et al.[67], with the difference that average Hi-C interaction frequencies were used within the specified genomic ranges to calculate compactness (instead of the sum of Hi-C interactions).

Reporting summary

Further information on research design is available in the Nature Research Reporting Summary linked to this article.

Data availability

Datasets generated for this study can be accessed in Supplementary Data 1–16 and via JBrowse (https://geneart.med.unideb.hu/pub/2021-ndx). Raw data are available at GEO GSE201841. External datasets: TAIR10 gene annotation files (gene list, splice junctions) were obtained from The *Arabidopsis* Information Resource. Promoter and downstream regions were defined as the arbitrary extension of transcription start sites (TSS) and termination sites (TTS) by 2000 base pairs. All other datasets used in this study are summarized in Supplementary Data 15. Source data are provided with this paper.

References

1. Zhang, H., Lang, Z. & Zhu, J. K. Dynamics and function of DNA methylation in plants. *Nat. Rev. Mol. Cell Biol.* **19**, 489–506 (2018).
2. Matzke, M. A. & Mosher, R. A. RNA-directed DNA methylation: an epigenetic pathway of increasing complexity. *Nat. Rev. Genet.* **15**, 394–408 (2014).
3. Zhang, X. et al. Genome-wide high-resolution mapping and functional analysis of DNA methylation in arabidopsis. *Cell* **126**, 1189–1201 (2006).
4. Lister, R. et al. Highly integrated single-base resolution maps of the epigenome in arabidopsis. *Cell* **133**, 523–536 (2008).
5. Gao, Z. et al. An RNA polymerase II-and AGO4-associated protein acts in RNA-directed DNA methylation. *Nature* **465**, 106–109 (2010).
6. Zheng, B. et al. Intergenic transcription by RNA polymerase II coordinates Pol IV and Pol V in siRNA-directed transcriptional gene silencing in Arabidopsis. *Genes Dev.* **23**, 2850–2860 (2009).
7. Zhong, X. et al. Molecular mechanism of action of plant DRM de novo DNA methyltransferases. *Cell* **157**, 1050–1060 (2014).
8. Lindroth, A. M. et al. Requirement of CHROMOMETHYLASE3 for maintenance of CpXpG methylation. *Science* **292**, 2077–2080 (2001).
9. Stroud, H. et al. Non-CG methylation patterns shape the epigenetic landscape in Arabidopsis. *Nat. Struct. Mol. Biol.* **21**, 64–72 (2014).
10. Du, J. et al. Dual binding of chromomethylase domains to H3K9me2-containing nucleosomes directs DNA methylation in plants. *Cell* **151**, 167–180 (2012).
11. Du, J. et al. Mechanism of DNA methylation-directed histone methylation by KRYPTONITE. *Mol. Cell* **55**, 495–504 (2014).
12. Zhang, C. et al. Arabidopsis AGDP1 links H3K9me2 to DNA methylation in heterochromatin. *Nat. Commun.* **9**, 4547 (2018).
13. Zhao, S. et al. Plant HP1 protein ADCP1 links multivalent H3K9 methylation readout to heterochromatin formation. *Cell Res.* **29**, 54–66 (2018).
14. Mathieu, O., Probst, A. V. & Paszkowski, J. Distinct regulation of histone H3 methylation at lysines 27 and 9 by CpG methylation in Arabidopsis. *EMBO J.* **24**, 2783–2791 (2005).
15. Cao, R., Tsukada, Y. I. & Zhang, Y. Role of Bmi-1 and Ring1A in H2A ubiquitylation and hox gene silencing. *Mol. Cell* **20**, 845–854 (2005).
16. L, X. & WH, S. Polycomb silencing of KNOX genes confines shoot stem cell niches in Arabidopsis. *Curr. Biol.* **18**, 1966–1971 (2008).
17. Jiang, D. & Berger, F. DNA replication–coupled histone modification maintains Polycomb gene silencing in plants. *Science* **357**, 1146–1149 (2017).
18. AM, M., Z, B., Y, Y. & WH, S. Arabidopsis AL PHD-PRC1 complexes promote seed germination through H3K4me3-to-H3K27me3 chromatin state switch in repression of seed developmental genes. *PLoS Genet.* **10**, e1004091 (2014).
19. Zhu, Y. et al. The arabidopsis nodulin homeobox factor AtNDX interacts with AtRING1A/B and negatively regulates abscisic acid signaling. *Plant Cell* **32**, 703–721 (2020).
20. Mukherjee, K., Brocchieri, L. & Bürglin, T. R. A comprehensive classification and evolutionary analysis of plant homeobox genes. *Mol. Biol. Evol.* **26**, 2775–2794 (2009).
21. Sun, Q., Csorba, T., Skourti-Stathaki, K., Proudfoot, N. J. & Dean, C. R-loop stabilization represses antisense transcription at the Arabidopsis FLC locus. *Science* **340**, 619–621 (2013).
22. Jørgensen, J. E. et al. A new class of plant homeobox genes is expressed in specific regions of determinate symbiotic root nodules. *Plant Mol. Biol.* **40**, 65–77 (1999).
23. Xu, C. et al. R-loop resolution promotes co-transcriptional chromatin silencing. *Nat. Commun.* **12**, 1790 (2021).
24. Baxter, C. L., Šviković, S., Sale, J. E., Dean, C. & Costa, S. The intersection of DNA replication with antisense 3′ RNA processing in Arabidopsis FLC chromatin silencing. *Proc. Natl Acad. Sci. USA* **118**, 2–4 (2021).
25. Francis, N. J., Kingston, R. E. & Woodcock, C. L. Chromatin compaction by a polycomb group protein complex. *Science.* **306**, 1574–1577 (2004).
26. Eskeland, R. et al. Ring1B compacts chromatin structure and represses gene expression independent of histone ubiquitination. *Mol. Cell* **38**, 452–464 (2010).

dc_1977_21

27. Schoenfelder, S. et al. The pluripotent regulatory circuitry connecting promoters to their long-range interacting elements. *Genome Res* **25**, 582–597 (2015).
28. Zhang, Y. et al. The polycomb protein RING1B enables estrogen-mediated gene expression by promoting enhancer–promoter interaction and R-loop formation. *Nucleic Acids Res*. gkab723 https://doi.org/10.1093/NAR/GKAB723 (2021)
29. Shidlovskii, Y. V et al. Subunits of the PBAP chromatin remodeler are capable of mediating enhancer-driven transcription in Drosophila. *Int. J. Mol. Sci. Artic. Dros. Int. J. Mol. Sci*. https://doi.org/10.3390/ijms22062856 (2021)
30. Kostyuchenko, M. et al. Zeste can facilitate long-range enhancer-promoter communication and insulator bypass in Drosophila melanogaster. *Chromosoma* **118**, 665–674 (2009).
31. Ariel, F. et al. R-loop mediated trans action of the APOLO long noncoding RNA. *Mol. Cell* **77**, 1055–1065.e4 (2020).
32. Michaels, S. D. & Amasino, R. M. FLOWERING LOCUS C encodes a novel MADS domain protein that acts as a repressor of flowering. *Plant Cell* **11**, 949–956 (1999).
33. Bi, X. et al. Nonrandom domain organization of the Arabidopsis genome at the nuclear periphery. *Genome Res*. **27**, 1162–1173 (2017).
34. Phair, R. D., Gorski, S. A. & Misteli, T. Measurement of dynamic protein binding to chromatin in vivo, using photobleaching microscopy. *Methods Enzymol*. **375**, 393–414 (2004).
35. Kimura, H. & Cook, P. R. Kinetics of core histones in living human cells: little exchange of H3 and H4 and some rapid exchange of H2B. *J. Cell Biol*. **153**, 1341–1353 (2001).
36. Hu, Z., Zhang, A., Storz, G., Gottesman, S. & Leppla, S. H. An antibody-based microarray assay for small RNA detection. *Nucleic Acids Res*. **34**, 1–7 (2006).
37. Bou-Nader, C., Bothra, A., Garboczi, D. N., Leppla, S. H. & Zhang, J. Structural basis of R-loop recognition by the S9.6 monoclonal antibody. *Nat. Commun*. **13**, 1–14 (2022).
38. Ginno, P. A., Lott, P. L., Christensen, H. C., Korf, I. & Chédin, F. R-loop formation is a distinctive characteristic of unmethylated human CpG island promoters. *Mol. Cell* **45**, 814–25 (2012).
39. Halász, L. et al. RNA-DNA hybrid (R-loop) immunoprecipitation mapping: an analytical workflow to evaluate inherent biases. *Genome Res*. **27**, 1063–1073 (2017).
40. Fransz, P. F. et al. Integrated cytogenetic map of chromosome arm 4S of A. thaliana: structural organization of heterochromatic knob and centromere region. *Cell* **100**, 367–376 (2000).
41. Grob, S., Schmid, M. W., Luedtke, N. W., Wicker, T. & Grossniklaus, U. Characterization of chromosomal architecture in Arabidopsis by chromosome conformation capture. *Genome Biol*. **14**, 1–19 (2013).
42. Pontvianne, F. et al. Identification of nucleolus-associated chromatin domains reveals a role for the nucleolus in 3D organization of the A. thaliana genome. *Cell Rep*. **16**, 1574–1587 (2016).
43. Choi, J., Lyons, D. B. & Zilberman, D. Histone H1 prevents non-CG methylation-mediated small RNA biogenesis in Arabidopsis heterochromatin. *Elife* **10**, e72676 (2021).
44. Kralemann, L. E. M. et al. Removal of H2Aub1 by ubiquitin-specific proteases 12 and 13 is required for stable polycomb-mediated gene repression in Arabidopsis. *Genome Biol*. **21**, 1–19 (2020).
45. Manzo, S. G. et al. DNA Topoisomerase I differentially modulates R-loops across the human genome. *Genome Biol*. 1–18 (2018) https://doi.org/10.1186/s13059-018-1478-1.
46. Velichko, A. K. et al. Hypoosmotic stress induces R loop formation in nucleoli and ATR/ATM-dependent silencing of nucleolar transcription. *Nucleic Acids Res*. **47**, 6811–6825 (2019).
47. El Hage, A., French, S. L., Beyer, A. L. & Tollervey, D. Loss of topoisomerase I leads to R-loop-mediated transcriptional blocks during ribosomal RNA synthesis. *Genes Dev*. **24**, 1546–1558 (2010).
48. Hegedüs, É. et al. Endogenous single-strand DNA breaks at RNA polymerase II promoters in saccharomyces cerevisiae. *Nucleic Acids Res*. **46**, 10649–10668 (2018).
49. Székvölgyi, L. et al. Ribonucleoprotein-masked nicks at 50-kbp intervals in the eukaryotic genomic DNA. *Proc. Natl Acad. Sci. USA* **104**, 14964–14969 (2007).
50. Sequeira-Mendes, J. et al. The functional topography of the arabidopsis genome is organized in a reduced number of linear motifs of chromatin states. *Plant Cell* **26**, 2351–2366 (2014).
51. Hardcastle, T. J., Müller, S. Y. & Baulcombe, D. C. Towards annotating the plant epigenome: the Arabidopsis thaliana small RNA locus map. *Sci. Rep*. **8**, 1–15 (2018).
52. Swiezewski, S. et al. Small RNA-mediated chromatin silencing directed to the 3′ region of the Arabidopsis gene encoding the developmental regulator, FLC. *Proc. Natl Acad. Sci*. **104**, 3633–3638 (2007).
53. Kozomara, A., Birgaoanu, M. & Griffiths-Jones, S. MiRBase: from microRNA sequences to function. *Nucleic Acids Res*. **47**, D155–D162 (2019).
54. Varkonyi-gasic, E. & Hellens, R. P. qRT-PCR of small RNAs. *Methods Mol. Biol*. **631**, 109–122 (2010).
55. He, L. et al. Pathway conversion enables a double-lock mechanism to maintain DNA methylation and genome stability. *Proc. Natl. Acad. Sci. USA* **118**, e2107320118 (2021).
56. Maturana, B. V., Torres, F., Carrasco, M. & Tapia, J. C. Differential regulation of transposable elements (TEs) during the murine submandibular gland development. *Mob. DNA* **12**, 1–10 (2021).
57. Cuerda-Gil, D. & Slotkin, R. K. Non-canonical RNA-directed DNA methylation. *Nat. Plants* **2**, 16163 (2016).
58. Böhmdorfer, G. et al. RNA-directed DNA methylation requires stepwise binding of silencing factors to long non-coding RNA. *Plant J*. **79**, 181–191 (2014).
59. Olmedo-Monfil, V. et al. Control of female gamete formation by a small RNA pathway in Arabidopsis. *Nature* **464**, 628–632 (2010).
60. Gong, Z. et al. ROS1, a repressor of transcriptional gene silencing in Arabidopsis, encodes a DNA glycosylase/lyase. *Cell* **111**, 803–814 (2002).
61. Zhang, Y. et al. Large-scale comparative epigenomics reveals hierarchical regulation of non-CG methylation in Arabidopsis. *Proc. Natl Acad. Sci. USA* **115**, E1069–E1074 (2018).
62. Grob, S., Schmid, M. W. & Grossniklaus, U. Hi-C analysis in arabidopsis identifies the KNOT, a structure with similarities to the flamenco locus of Drosophila. *Mol. Cell* **55**, 678–693 (2014).
63. Feng, S. et al. Genome-wide Hi-C analyses in wild-type and mutants reveal high-resolution chromatin interactions in arabidopsis. *Mol. Cell* **55**, 694–707 (2014).
64. Liu, C. et al. Genome-wide analysis of chromatin packing in Arabidopsis thaliana at single-gene resolution. *Genome Res*. **26**, 1057–1068 (2016).
65. Grob, S. & Grossniklaus, U. Invasive DNA elements modify the nuclear architecture of their insertion site by KNOT - linked silencing in Arabidopsis thaliana. *Genome Biol*. **20**, 1–15 (2019).
66. Lafos, M. et al. Dynamic regulation of H3K27 trimethylation during arabidopsis differentiation. *PLoS Genet*. **7**, e1002040 (2011).
67. Zhu, W. et al. Altered chromatin compaction and histone methylation drive non-additive gene expression in an interspecific Arabidopsis hybrid. *Genome Biol*. **18**, 1–16 (2017).
68. Allshire, R. C. & Madhani, H. D. Ten principles of heterochromatin formation and function. *Nat. Rev. Mol. Cell Biol*. **19**, 229–244 (2018).
69. Dalmadi, Á., Gyula, P., Bálint, J., Szittya, G. & Havelda, Z. AGO-unbound cytosolic pool of mature miRNAs in plant cells reveals a novel regulatory step at AGO1 loading. *Nucleic Acids Res*. **47**, 9803–9817 (2019).

dc_1977_21

70. Dai, X., Zhuang, Z. & Zhao, P. X. PsRNATarget: A plant small RNA target analysis server (2017 release). *Nucleic Acids Res.* **46**, W49–W54 (2018).
71. Bickel, S. & Pirrotta, V. Self-association of the Drosophila zeste protein is responsible for transvection effects. *EMBO J.* **9**, 2959–2967 (1990).
72. Mizuguchi, T. et al. Cohesin-dependent globules and heterochromatin shape 3D genome architecture in S. pombe. *Nature* https://doi.org/10.1038/nature13833 (2014).
73. Shindo, C., Lister, C., Crevillen, P., Nordborg, M. & Dean, C. Variation in the epigenetic silencing of FLC contributesto natural variationin Arabidopsis vernalization response. *Genes Dev.* **20**, 3079–3083 (2006).
74. Karányi, Z. et al. Nuclear dynamics of the Set1C subunit Spp1 prepares meiotic recombination sites for break formation. *J. Cell Biol.* **217**, 3398–3415 (2018).
75. Hetey, S. et al. Biophysical characterization of histone H3.3 K27 M point mutation. *Biochem. Biophys. Res. Commun.* **490**, 868–875 (2017).
76. Haupts, U., Maiti, S., Schwille, P. & Webb, W. W. Dynamics of fluorescence fluctuations in green fluorescent protein observed by fluorescence correlation spectroscopy. *Proc. Natl Acad. Sci. USA* **95**, 13573–13578 (1998).
77. Kim, D., Langmead, B. & Salzberg, S. L. HISAT: a fast spliced aligner with low memory requirements. *Nat. Methods* **12**, 357–360 (2015).
78. Li, H. et al. The sequence alignment/map format and SAMtools. *Bioinformatics* **25**, 2078–2079 (2009).
79. Ramírez, F. et al. deepTools2: a next generation web server for deep-sequencing data analysis. *Nucleic Acids Res.* **44**, W160–W165 (2016).
80. Zhang, Y. et al. Model-based analysis of ChIP-Seq (MACS). *Genome Biol.* **9**, R137 (2008).
81. Quinlan, A. R. & Hall, I. M. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**, 841–842 (2010).
82. Zuker, M. Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Res.* **31**, 3406–3415 (2003).
83. Wu, X. et al. SRNAnalyzer-A flexible and customizable small RNA sequencing data analysis pipeline. *Nucleic Acids Res.* **45**, 12140–12151 (2017).
84. Varkonyi-Gasic, E. & Hellens, R. P. qRT-PCR of small RNAs. *Methods Mol. Biol.* **631**, 41–48 (2010).
85. Ignatiadis, N., Klaus, B., Zaugg, J. B. & Huber, W. Data-driven hypothesis weighting increases detection power in genome-scale multiple testing. *Nat. Methods* **13**, 577–580 (2016).
86. Krueger, F. & Andrews, S. R. Bismark: a flexible aligner and methylation caller for Bisulfite-Seq applications. *Bioinformatics* **27**, 1571–1572 (2011).
87. Kerpedjiev, P. et al. HiGlass: web-based visual exploration and analysis of genome interaction maps. *Genome Biol.* **19**, 125 (2018).
88. Baudry, L., Millot, G. A., Thierry, A., Koszul, R. & Scolari, V. F. Serpentine: a flexible 2D binning method for differential Hi-C analysis. *Bioinformatics* **36**, 3645–3651 (2020).

Acknowledgements

This work was funded by HAS-Lendület-LP2015-9/2015, NKFIH-NNE-130913, GINOP-2.3.2-15-2016-00024, and the Thematic Excellence Programme (TKP2021-EGA-18) of the National Research, Development and Innovation Fund of Hungary. L.Sz was supported by the Bolyai Janos fellowship of the Hungarian Academy of Sciences and the UNKP-21-5-DE-11 and UNKP-22-5-DE-3 new national excellence program of the Ministry For Innovation and Technology from the source of the National Research, Development and Innovation Fund. M.M received support from NKFIH-K137678. T.Cs received grants from NKFIH-K129283, K137722, and K136513. We thank Caroline Dean (John Innes Centre, UK) for providing all the plant lines utilized in this study. We thank Csaba Máthé and Tamás Garda (Dept. of Botany, University of Debrecen) for their contribution to microscopic measurements. We are grateful for the Genomic Medicine and Bioinformatics Core Facility (University of Debrecen) for the NGS service.

Author contributions

Á.M, A.H., B.B., Sz.H., É.N., H.Sz, T.Cs. performed the research, Zs.K, M.M., O.F., T.Cs., L.Sz analyzed the data, L.Sz., T.Cs., I.H. secured funding and supervised the work, L.Sz and T.Cs. wrote the manuscript.

Funding

Open access funding provided by University of Debrecen.

Competing interests

The authors declare no competing interest.

Additional information

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41467-022-32709-y.

Correspondence and requests for materials should be addressed to Tibor Csorba or Lóránt Székvölgyi.

Peer review information *Nature Communications* thanks the anonymous reviewers for their contribution to the peer review of this work. Peer reviewer reports are available.

Reprints and permission information is available at http://www.nature.com/reprints

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit http://creativecommons.org/licenses/by/4.0/.

© The Author(s) 2022

dc_1977_21

Ribonucleoprotein-masked nicks at 50-kbp intervals in the eukaryotic genomic DNA

Lóránt Székvölgyi*, Zsuzsa Rákosy†, Bálint L. Bálint‡, Endre Kókai§, László Imre*, György Vereb*, Zsolt Bacsó*, Katalin Goda*, Sándor Varga¶, Margit Balázs†, Viktor Dombrádi§, László Nagy‡, and Gábor Szabó*‖

*Department of Biophysics and Cell Biology, †Department of Preventive Medicine, Division of Biomarker Analysis, ‡Department of Biochemistry and Molecular Biology, ¶Clinical Research Center, and §Cell Biology and Signaling Research Group of the Hungarian Academy of Sciences, Department of Medical Chemistry, Research Center for Molecular Medicine, Medical and Health Science Center, University of Debrecen, 4032, Debrecen, Hungary

Edited by Mark T. Groudine, Fred Hutchinson Cancer Research Center, Seattle, WA, and approved August 7, 2007 (received for review March 12, 2007)

By using a microscopic approach, field inversion single-cell gel electrophoresis, we show that preformed single-strand discontinuities are present in the chromatin of resting and proliferating mammalian and yeast cells. These single-strand breaks are primarily nicks positioned at ≈50-kbp intervals throughout the entire genome that could be efficiently labeled *in situ* by DNA polymerase I holoenzyme but not by Klenow fragment and terminal transferase unless after ribonucleolytic treatments. The RNA molecules involved appear to comprise R-loops, recognized by the S9.6 RNA/DNA hybrid-specific antibody. By using the breakpoint cluster region of the *Mixed Lineage Leukemia* (*MLL*) gene as a model, we have found that the number of manifest nicks detected by FISH performed after field inversion single-cell gel electrophoresis depends on epigenetic context, but the difference between germ-line and translocated MLL alleles is abolished by protease treatment. Our data imply that the double-stranded genomic DNA is composed of contiguous rather than continuous single strands and reveal an aspect of higher-order chromatin organization with ribonucleoprotein-associated persistent nicks defining ≈50-kbp domains.

chromatin loop | RNA/DNA hybrid | translocation

The concept that eukaryotic chromatin is organized into ≈30- to 150-kbp units anchored to a ribonucleoprotein-containing structure, the enigmatic nuclear matrix/scaffold, has stemmed from microscopic observations of DNA loops emanating from histone-depleted nuclei (for review, see ref 1). Chromatin appears to bind matrix elements through special, although heterogeneous, DNA sequences, scaffold/matrix attachment regions (S/MARs), that remain attached to the remnants of salt-extracted nuclei (nuclear halos) and are thought to represent the boundaries of supercoiled 20- to 150-kbp looped domains (2). Consistent with this model of chromosome architecture, chromatin fragmentation phenomena have been observed that involve the preferential cleavage of DNA, presumably at the bases of loops (3). The global disassembly of chromatin to high-molecular-weight (≥20-kbp) units also takes place upon alkali denaturation after proteinase digestion (4), at exposure to single-strand (ss)-specific nuclease (5), in the early stage of apoptotic DNA fragmentation (6), as well as in the case of healthy nonapoptotic mammalian and yeast cells upon various protein denaturing treatments (7, 8). The DNase I hypersensitivity of mammalian chromatin at every ≈50 kbp (9), also detected in the vicinity of certain S/MARs (10), points to the special vulnerability of the DNA at the borders of supernucleosomal units of this size. The above data raise the question whether special base-unpaired secondary structures or perhaps regularly spaced stably maintained ss discontinuities constitute the predilection points of ≈50-kbp chromatin fragmentation, delimiting higher-order domains. To tackle this issue, based on the conventional comet assay (11), we have developed a microscopic approach, field inversion single-cell gel electrophoresis (FI-SCGE), that provides direct visual evidence for the presence of preformed ss nicks, positioned at regular ≈50-kbp intervals throughout the interphase chromatin of nonapoptotic mammalian and yeast cells. These nicks are shown to be associated with ribonucleoprotein structures that appear to anchor them to the nuclear matrix, delimiting supercoiled domains.

Results

FI-SCGE Reveals Persistent Regularly Spaced ss Discontinuities in the Chromatin of Nonapoptotic Cells. Upon alkaline FI-SCGE, the nuclear halos of healthy nonapoptotic Jurkat cells become completely disassembled to granules (Fig. 1*A*). The same phenomenon could be observed in the case of proliferating or resting cells of several cell lines (e.g., ML-1, HL-60, HeLa, and HPBL) and human peripheral blood lymphocytes (PBLs), and in *Saccharomyces cerevisiae* spheroplasts (Fig. 1*B*). Similar granules were seen in normal (DC-3F) and topoisomerase IIβ-deficient Chinese hamster cells [DC-3F/9-OH-E; supporting information (SI) Fig. 6*A*], and also in wild-type, top2-defective, and Δtop1 yeast strains. The chromatin particles were revealed only in alkaline electrophoretic conditions, i.e., when the DNA migrates in a denatured state; neutral FI-SCGE of nuclear halos yielded parallel stretches of uninterrupted DNA fibers (Fig. 1*C*). As determined by several independent methods, the average DNA content of the particles is compatible with what is assumed for the loop level of higher-order chromatin architecture (12). First, quantitative image analysis of individual fluorescent speckles gave an average DNA fragment size of ≈48 kbp per particle (for calculation, see *Materials and Methods*). Second, FISH was performed on nuclear halos and FI-SCGE comets (FI-SCGE FISH) by using a probe specific for the *Mixed Lineage Leukemia* (*MLL*) gene that is frequently rearranged in childhood and posttherapeutic leukemias (13). The *MLL* breakpoint cluster region (bcr) includes two high-affinity S/MARs implicated as predilection zones of both chromatin fragmentation and translocations (14, 15). Although the chromosome territories were well preserved in the nuclear halos (Fig. 1*D1*), the MLL FISH probe hybridized to the alkaline FI-SCGE comets as several discrete distant spots (Fig. 1*D2* and *3*). The average fragment size, calculated by dividing the overall length of the region recognized by the MLL probe with the observed average frag-

Author contributions: L.S. and G.S. designed research; L.S., Z.R., B.L.B., and E.K. performed research; L.S., L.I., G.V., Z.B., S.V., M.B., V.D., L.N., and G.S. contributed new reagents/analytic tools; L.S., G.V., Z.B., K.G., and G.S. analyzed data; and L.S. and G.S. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

Abbreviations: S/MARs, scaffold/matrix attachment regions; ss, single strand; FI-SCGE, field inversion single-cell gel electrophoresis; PBL, peripheral blood lymphocyte; MLL, mixed lineage leukemia; bcr, breakpoint cluster region; Pol I, polymerase I; TdT, terminal deoxynucleotidyltransferase; gMLL, germ-line MLL; FIGE, field inversion gel electrophoresis.

‖To whom correspondence should be addressed. E-mail: szabog@dote.hu.

This article contains supporting information online at www.pnas.org/cgi/content/full/0702269104/DC1.

© 2007 by The National Academy of Sciences of the USA

Downloaded by guest on November 19, 2021

dc_1977_21

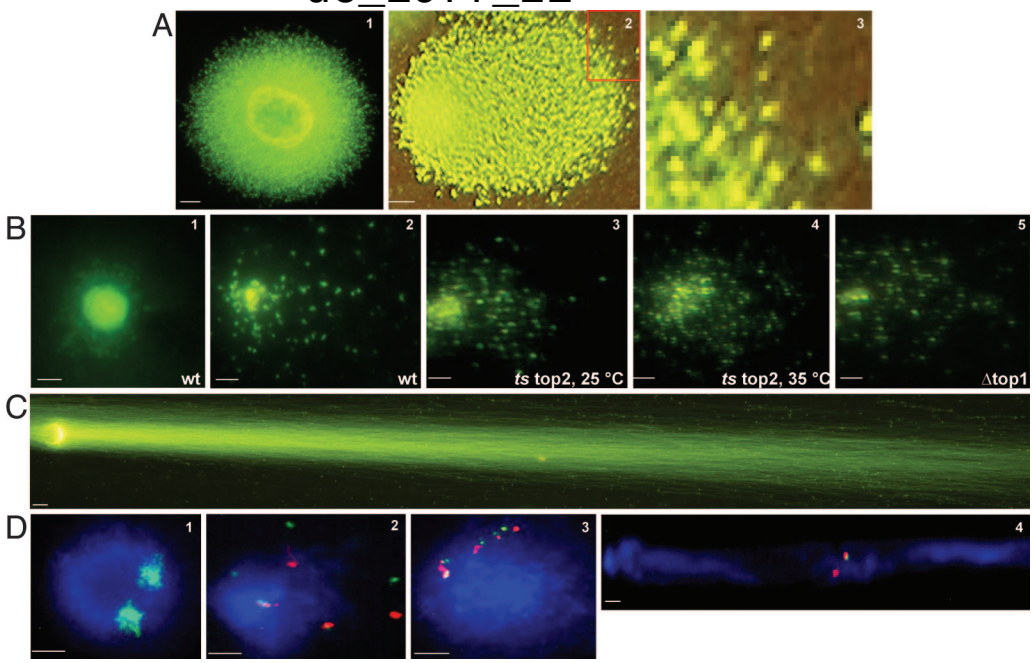


Fig. 1. Single-strand discontinuities in the DNA revealed at ≈50-kbp intervals by FI-SCGE and FISH. (*A*) Nuclear halos of nonapoptotic Jurkat cells before (*1*) and after (*2*) alkaline FI-SCGE; *'3'* shows a magnified portion of *'2'*. (*B*) Nuclear halos prepared from WDHY199 *S. cerevisiae* spheroplast before (*1*) and after (*2*) alkaline FI-SCGE. Alkaline FI-SCGE of *ts* top2-mutant JN394t2–4 yeast spheroplasts grown at permissive temperature (*3*) or after incubation at the restrictive temperature (*4*); alkaline FI-SCGE of top1-mutant JEL1Δtop1 yeast spheroplasts (*5*). (Scale bar, 2 μm.) (*C*) Neutral FI-SCGE of nuclear halos prepared from Jurkat cells. (Scale, 5 μm.) (*D*) FI-SCGE FISH performed on alkaline FI-SCGE comets prepared from quiescent PBLs (*2*) or cycling Jurkat cells (*3*) and on neutral FI-SCGE comets (*4*). *1* shows chromosome 8 detected by a chromosome painting probe. *2–4* shows a 540-kbp region of chromosome 11q23 spanning the *MLL* gene; green, probe covering the centromeric half of *MLL*; red, probe spanning the telomeric half of *MLL*; blue, DNA. (Scale bar, 5 μm.)

ment number (2 × 540 kbp divided by 10.2; SEM = ±0.4; n = 200) was 106 kbp. FISH of neutral FI-SCGE comets has led to compact (unfragmented) signals (Fig. 1*D4*). Third, DNA samples reisolated either from nuclear halos or alkaline FI-SCGE comets and analyzed by nondenaturing field inversion gel electrophoresis (FIGE) have yielded a band focused at ≈50 kbp in each case (Fig. 2*A*).

Similar disintegration of chromatin was also observed when agarose-embedded intact chromosomal DNA samples were analyzed either by alkaline FIGE (Fig. 2 *B* and *D*) or, after S1 nuclease digestion, by neutral FIGE (Fig. 2 *C*, *E*, and *F*). Both methods revealed a massive chromatin disassembly, in the case of untreated and formaldehyde-fixed mammalian cells and *S. cerevisiae* spheroplasts alike, to ≈50-kbp fragments. Very similar fragmentation patterns were seen in normal or topoisomerase-mutant cell lines and yeast strains (Fig. 2 *B–E*), arguing against the possibility that the strand breaks arise from a random access of topoisomerases to DNA during cell lysis. In view of the above results, and because both formaldehyde fixation and rapid alkaline lysis preclude the initiation of enzymatic actions, we assume that the observed ss breaks preexist at ≈50-kbp intervals in the chromatin of live cells.

Nicks Are Present in the Chromatin at ≈50-kbp Intervals, Masked by RNA and Protein. As Fig. 3*A* demonstrates, the nuclear halos could be efficiently nick-labeled by using DNA polymerase I (Pol I), indicating that nicks or gaps ending in free 3′ OH are present in the nuclear halos. Neither terminal deoxynucleotidyltransferase (TdT) nor Klenow enzyme could efficiently label the nuclear halos (Fig. 3 *B* and *D*), arguing against the possibility that the discontinuities are mainly ds breaks and/or ss gaps. The number of nick-translated spots (7×10^4), determined by quantitative image analysis (see *Materials and Methods*), was comparable to the number of disassembled chromatin granules seen after alkaline FI-SCGE. Remarkably, both Klenow and TdT could intensively incorporate biotin-dUTP if the samples were previously exposed to exonuclease III (Exo III), RNase A, or alkali (Fig. 3 *C* and *E–G*). The effect of RNase A was abolished in the presence of a specific RNase inhibitor (Fig. 3*H*). The finding that ribonucleolytic treatment is required for efficient TdT and Klenow labeling suggests that access of TdT to these discontinuities and incorporation of nucleotides by Klenow are both constrained by RNA. By their RNase H activity, Pol I and Exo III can degrade the RNA strands in RNA/DNA hybrids; thus the RNase-sensitive structures limiting labeling at the nicks are likely to comprise RNA/DNA hybrids. This conclusion was further corroborated by using the anti-RNA/DNA hybrid-specific monoclonal antibody S9.6 (16), revealing a crowd of fluorescent speckles representing RNA/DNA hybrids in the nuclear halos (Fig. 3*M*). The number of RNA/DNA hybrid speckles was comparable to the number of nick-translatable particles or to that of the disassembled granules in the alkaline FI-SCGE comets ($\approx 10^5$).

In line with the preformed nature of the breaks, the chromatin of HCHO-fixed cells (Fig. 3 *I–L*) could also be labeled by both Pol I and TdT; because labeling in the case of these samples was observed only after extensive proteinase K digestion, the accessibility of the free 3′-OH termini must also be hindered by proteins staying attached to the bases of looped domains forming the halos.

The nicks as well as the RNA/DNA hybrids are primarily localized in the central, probably matrix-associated, area of the halos; the DNA loops emanating from the center are barely labeled. As shown in Fig. 4, upon addition of increasing concentrations of the intercalating dye ethidium bromide (based on ref. 17), the radius of halos increased to a maximum, because of stretching out of negative supercoils, then at high concentrations of the dye, it gradually decreased as a result of overwinding. Both

Downloaded by guest on November 19, 2021

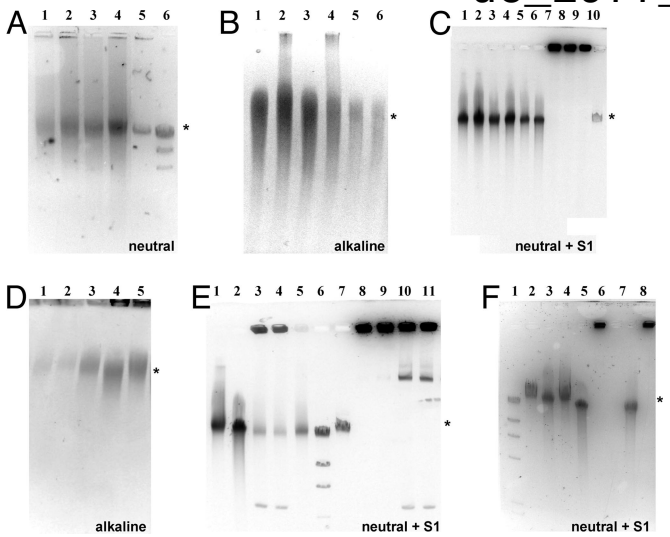


Fig. 2. Demonstration of single-strand discontinuities in the DNA by FIGE. (*A*) FIGE of DNA reisolated from the nuclear halos (lanes 1 and 3) and alkaline FI-SCGE comets (lanes 2 and 4) of Jurkat (lanes 1 and 2) and ML-1 (lanes 3 and 4) cells. Lane 5, λ DNA; lane 6, λ DNA HindIII marker. (*B*) Alkaline FIGE of intact DNA molecules prepared from untreated (lanes 1, 3, 5, and 6) and HCHO-fixed (lanes 2 and 4) Jurkat (lanes 1 and 2), ML-1 (lanes 3 and 4), DC-3F (lane 5), and DC-3F/9-OH-E (lane 6) cells embedded into agarose plugs. SI Fig. 7*C* shows the Southern blot of this gel hybridized with the 0.2 HB *MLL* probe. (*C*) Nondenaturing FIGE of intact DNA of untreated (lanes 2, 4, and 6) and HCHO-fixed (lanes 1, 3, and 5) Jurkat (lanes 1 and 2), ML-1 (lanes 3 and 4), and DC-3F (lanes 5 and 6) cells, agarose-embedded and lysed as described in *Materials and Methods*. Lanes 1–6, S1 nuclease digested samples; lanes 7–9, undigested control plugs; lane 10, λ DNA. (*D*) Alkaline FIGE of intact DNA molecules prepared from different agarose-embedded *S. cerevisiae* strains. Lanes 1 and 2, *ts* top2-mutant JN394t2–4 grown at permissive temperature and after incubation at restrictive temperature, respectively; lane 3, top1-mutant JEL1Δtop1; lane 4, JN394; lane 5, WDHY199 cells. (*E*) Nondenaturing FIGE of intact DNA of agarose-embedded HCHO-fixed DC-3F/9-OH-E cells (lanes 1 and 2) and different *S. cerevisiae* spheroplasts (lanes 3–5), lysed and S1-digested as described in *Materials and Methods*. Lanes 3 and 4, *ts* top2-mutant JN394t2–4 strain grown at permissive temperature and after incubation at restrictive temperature, respectively; lane 5, top1-mutant JEL1Δtop1 strain; lane 6, λ HindIII marker; lane 7, λ DNA; lanes 8–11, undigested controls of HCHO-fixed DC-3F/9-OH-E cells (lanes 8 and 9) and yeast spheroplasts (lanes 10 and 11), respectively. (*F*) Nondenaturing FIGE of intact DNA of untreated (lanes 3 and 5) and HCHO-fixed (lanes 4 and 7) *S. cerevisiae* spheroplasts, lysed and S1-digested as described in *Materials and Methods*. Lane 1, λ HindIII marker; lane 2, λ DNA; lanes 3 and 4, WDHY199; lanes 5 and 6, JN394 cells; lanes 7 and 8, undigested control plugs. *, 50 kbp.

proteinase and RNase digestion elicited the enlargement of nuclear halo radii, indicating the complete disruption of the nuclear matrix. Similarly to the nicking of DNA by UV irradiation, both treatments eliminated the characteristic increment phase, reflecting the relaxation of negative supercoils.

The Nicks at ≈50-kbp Intervals Are Not Specific for Proliferating or Apoptotic Cells. In Jurkat cells and human PBLs harboring germ-line MLL (gMLL) at chromosome 11q23, FI-SCGE FISH revealed disassembly to discrete granules containing loop-sized DNA (Fig. 5*B1*; see also Fig. 1*D2* and *3*). There was no difference in the fragmentation pattern of gMLL between quiescent PBLs and cycling Jurkat cells (Fig. 1*D2* and *3*), and between BrdU-positive (S-phase) and BrdU-negative cells (Fig. 5*B*). Halo FISH carried out in apoptotic cells showed a markedly different picture of fragmentation, with numerous fine speckles corresponding to oligonucleosomal DNA cleavages (SI Fig 6*D1* and *2*). These results indicate that the occurrence of nicks at ≈50-kbp intervals is not specific either for proliferating or apoptotic cells.

Comparison of Germ-Line and Translocated MLL Alleles. The possible effect of chromatin structure on the incidence of nicks or on their manifestation in FI-SCGE FISH was assessed by using ML-1 cells harboring rearranged *MLL* (t*MLL*) at chromosome 6q27 (18). As clearly demonstrated by reverse-transcriptase QPCR, ChIP, and DNase I-sensitivity assays, t*MLL* is transcriptionally repressed and lies in a tightly packed chromatin structure, as opposed to the transcriptionally active g*MLL* (Fig. 5*A*). FI-SCGE FISH showed that in a majority of cells (80%; $n = 120$), t*MLL* was present in a much less fragmented state than g*MLL* in PBLs or Jurkat cells, encompassing ≥540-kbp DNA (Figs. 5*B1′*). This result suggests that either the incidence of ss breaks or their manifestation in these experiments depends on chromosomal context. In line with this conclusion, the centromeres of chromosome 8 also exhibited a distinct pattern of fragmentation (SI Fig. 6*C*) with an average fragment size of 330 kbp, consistent with the loop periodicity estimated for centromeres (19).

As Fig. 5*B2′* shows, proteinase digestion of the nuclear halos before FI-SCGE has led to the disassembly of t*MLL* as well, overruling the above difference between the germ-line and translocated alleles. Consistent with this result, we have found no difference between g*MLL* and t*MLL* in the number or distribution of ss breaks in alkaline and neutral Southern blots, the latter performed on S1-digested DNA (SI Fig. 7 *B–D*) or in primer extension mapping on purified DNA samples (SI Fig. 7*E*). Comparing the size distributions of the 8.3-kbp bcr (SI Fig. 7 *B–D*) to that of bulk DNA (Fig. 2 *B* and *C*), we estimate approximately one nick per MLL bcr on average, randomly distributed both in the case of g*MLL* and t*MLL*, as opposed to the average one nick per 50 kbp in bulk DNA.

Discussion

Our findings are reminiscent of early observations by Weintraub (5), demonstrating the release of discrete supranucleosomal "a" particles containing ≈20- to 40-kbp dsDNA upon mild nuclease digestion of chromatin and of those by Werner *et al.* (20–22) on the presence of protease-induced S1 nuclease- and alkaline-sensitive regions arranged at ≈13.5- and 27-kbp intervals, respectively, in Ehrlich ascites tumor cell DNA. In the latter experiments, additional labile sites of different configuration may have been revealed (23), because the ends of those shorter DNA fragments, observed at more intensively denaturing conditions, were not labeled by Pol I. Alternatively, secondary changes might have been generated in that or in our system, leading to different end structures. Notwithstanding the differences, our observations and those of refs. 20–22 may be closely related. Compared with the above reports, we argue for the preformed nature of the discontinuities, exclude apoptosis as their source, extend these observations to other cell types including yeast spheroplasts, implicate RNA/DNA hybrids in the molecular structures masking the nicks, and directly tackle the possible role of topoisomerases. Furthermore, the microscopic approach developed has led to an experimental system readily amenable to molecular analysis also at the single-cell level.

Our data provide evidence that the chromosomes of nonapoptotic cells contain ss discontinuities positioned at ≈50-kbp intervals all over the entire genome. As demonstrated by alkaline FIGE and *in situ* labeling of nuclear halos and HCHO-fixed cells, the revealed discontinuities are persistent nicks, probably representing the predilection points of high-molecular-weight dsDNA fragmentation. As shown by the characteristic changes of halo radius upon addition of an intercalator dye (Fig. 4), the fragments bordered by the nicks are indeed supercoiled DNA loops; these loops are apparently tethered at the nicks so that spontaneous relaxation is prevented by anchoring structures. Thus, we hypothesize that the presence of ribonucleoprotein-

Downloaded by guest on November 19, 2021

dc_1977_21

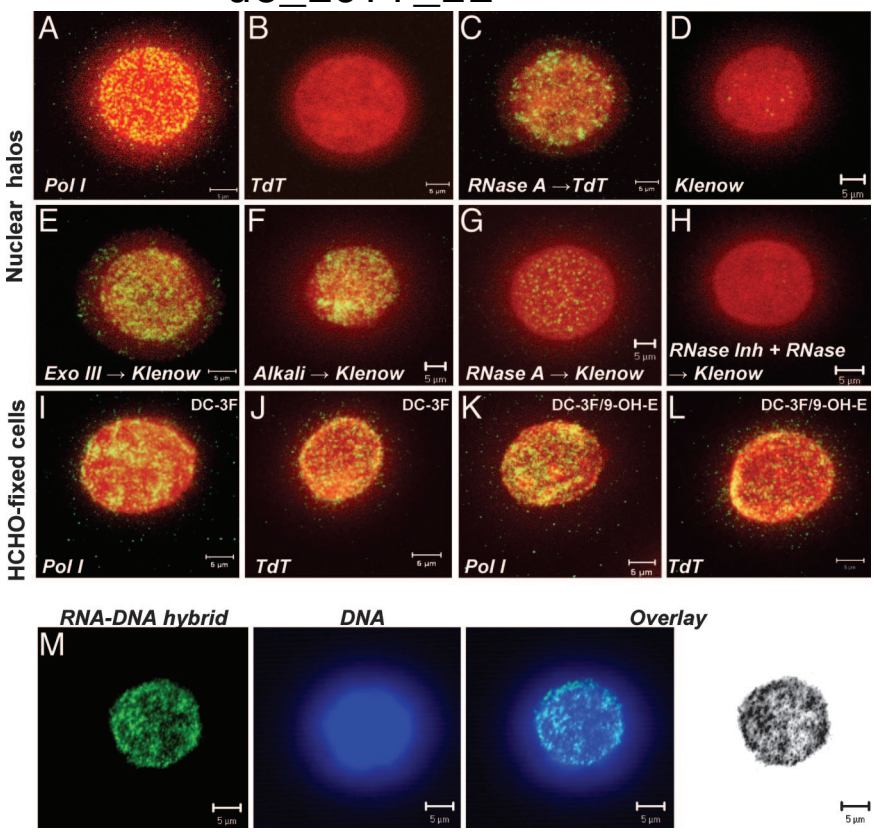


Fig. 3. *In situ* labeling characteristics of the DNA termini. *In situ* labeling of nicks in nuclear halos (*A–H*) and HCHO-fixed cells (*I–L*). Green, nick-translated regions; red, DNA. (*A*) Breaks in the nuclear halos labeled by DNA Pol I. (*B*) Lack of labeling by TdT. (*C*) Nuclear halos pretreated with RNase A, labeled by TdT. (*D*) Slight labeling by Klenow enzyme. (*E*) Nuclear halos pretreated with exonuclease III, labeled by Klenow fragment. (*F*) Nuclear halos prepared under alkaline conditions, labeled by Klenow fragment. (*G*) Nuclear halos pretreated with RNase A, labeled by Klenow enzyme. (*H*) RNase inhibitor prevents the effect of RNase A on labeling with Klenow. HCHO-fixed DC-3F (*I–J*) and DC-3F/9-OH-E (*K–L*) cells labeled by DNA Pol I (*I* and *K*) and TdT (*J* and *L*). Before nick labeling, RNA and proteins were removed by lysing the HCHO-fixed cells in a buffer containing RNase A and proteinase K/EDTA and Sarkosyl. (*M*) Detection of RNA/DNA hybrids in nuclear halos. Green, RNA/DNA hybrids labeled by the S9.6 antibody; blue, DNA.

masked nicks at ≈50-kbp intervals is related to the formation of chromatin loops (SI Fig. 8).

Because ribonucleolytic treatment of nuclear halos could expose the nicks for *in situ* labeling by Klenow or TdT, access to the free 3′-OH termini detected in our experiments must be hindered by RNA. RNA may be present in RNA/DNA hybrids, a possibility confirmed by using an anti-RNA/DNA-specific antibody, which has identified a similar number of fluorescent speckles, in similar localization as *in situ* nick translation. The following observations suggest that the nicks are associated with proteins as well: HCHO-fixed cells can be nick-labeled only after proteinase treatment (Fig. 3 *I–L*); proteinase K digestion relaxes the supercoiled DNA loops in the nuclear halos (Fig. 4), and the difference in the fragmentation of g*MLL* and t*MLL* is abolished by proteinase added before FI-SCGE FISH (Fig. 5). The similar incidence but different exposition of nicks in the expressed germ line and the transcriptionally inactive translocated *MLL* alleles (SI Fig. 7 and Fig. 5) demonstrate that the masking of nicks depends on the epigenetically controlled chromatin structure. The above findings are in line with the perception of nuclear matrix as a web of ribonucleoproteins associated with the bases of DNA loops (1, 24–28). Identification of the proteins and RNA molecules involved in hiding the described nicks from becoming manifest discontinuities remains a challenge for the future and may lead to an understanding of the mechanism of how the nicks are established.

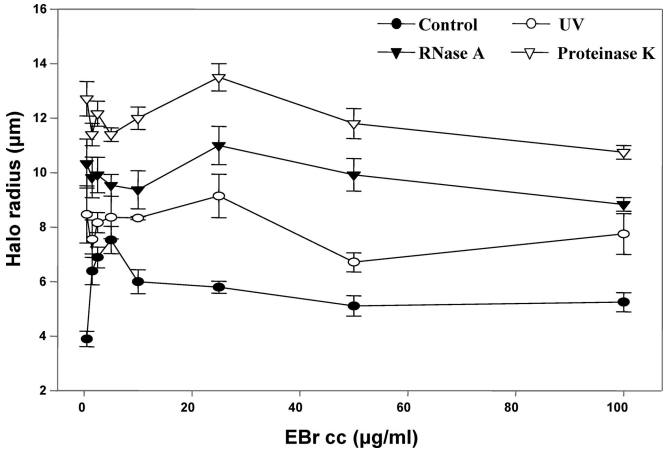


Fig. 4. Dependence of nuclear halo radius on ethidium bromide (EBr) concentration. Samples were stained with 0.5, 1.5, 2.5, 5, 10, 25, 50, and 100 μg/ml EBr. Means are indicated; error bars represent SD. Control, intact nuclear halos containing supercoiled loops of DNA; UV, nuclear halos nicked by UV irradiation; RNase A, nuclear halos digested with RNase A; and Proteinase K, nuclear halos digested by proteinase K.

Our results do not support the involvement of topoisomerase I and II in the observed phenomena; we failed to observe any difference in the case of topoisomerase-mutant mammalian cells and yeast strains (Figs. 1–3), or when cells were treated with the topoisomerase poison ICRF-193 (Figs. 5*B3, 4, 3′*, and *4′*). Yet topoisomerases might play a role by forming at ≈50-kbp inter-

Downloaded by guest on November 19, 2021

dc_1977_21

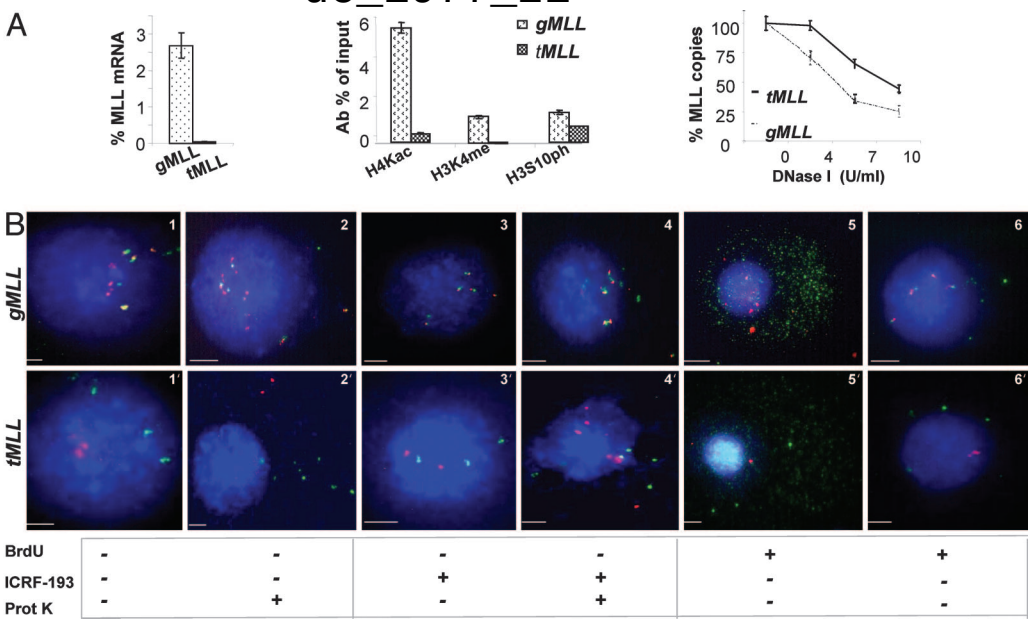


Fig. 5. Effect of chromosomal context on the incidence of nicks. (*A*) Comparison of chromatin structure at the germ line (g*MLL*) and translocated (t*MLL*) MLL bcr. g*MLL* is present at chromosome 11q23 in Jurkat cells, t*MLL* is at 6q27 in ML-1 cells. (*Left*) RT-quantitative real-time PCR (QPCR) analysis of g*MLL* and t*MLL* expression. *y* axis denotes the MLL transcript levels normalized to cyclophylin expression. (*Center*) Histone modifications within g*MLL* and t*MLL*, detected by chromatin immunoprecipitation. *y* axis denotes the background corrected ''% of input DNA'' values. (*Right*) Measurement of DNase I sensitivity of g*MLL* and t*MLL* by QPCR. Solid line, g*MLL*; dashed line, t*MLL*. (*B*) Detection of ss discontinuities by FI-SCGE FISH within a 540-kbp region of chromosome 11q23 that spans g*MLL* or 6q27 that spans t*MLL*. FI-SCGE comets were prepared from ML-1 cells harboring t*MLL* and Jurkat cells harboring *gMLL*. Green, probe flanking the centromeric half of *MLL* bcr; red, probe flanking the telomeric half of *MLL* bcr; blue, DNA. *1* and *1′*, control samples; *2*, *2′*, *4*, and *4′*, FI-SCGE comets of proteinase K-digested nuclear halos; *3*, *3′*, *4*, and *4′*, FI-SCGE comets prepared from ICRF-193-treated Jurkat and ML-1 cells; *5*, *5′*, *6*, and *6′*, FI-SCGE comets of BrdU-labeled Jurkat and ML-1 cells. S-phase (*5* and *5′*) and nonreplicating cells (*6* and *6′*) are shown, respectively. Green, BrdU incorporation and FISH probe that flank the centromeric half of MLL bcr. (Scale bar, 5 μm.)

vals stably maintained ''cleavable complexes.'' This possibility would be congruent with the participation of topoisomerases both in transcription and replication, also in view of the suggested links between the two processes (29, 30). Restructuring and activation of certain promoters require topoisomerase II-elicited DNA breaks (31). Recently, both topoisomerase I and II have been shown to compete with ORC2 and induce ss breaks at the lamin B2 origin of replication in a well orchestrated manner throughout the cell cycle, giving an essential contribution to origin definition; in addition, topoisomerase II appeared to be a permanent resident of a neighboring S/MAR (32).

The ribonucleoprotein-masked ss discontinuities described herein might play a role in certain aspects of genomic instability that lead to chromosomal translocations in human malignancies. This aspect has gained a novel perspective recently: rapid ≈50-kbp DNA fragmentation has been observed when recruitment of the ASF/SF2 premRNA splicing factor to the nascent mRNA was prevented (33). The fragmentation events took place in the absence of apoptosis and were accompanied by gene rearrangements; the hypermutator phenotype induced was shown to be the result of the formation of RNA/DNA hybrids; however, it was not resolved how the DNA lesions were introduced into the R-loop regions.

In summary, our data demonstrate an aspect of higher-order chromosome organization: the presence of regularly spaced ss discontinuities closely associated with and masked by RNA, comprising RNA/DNA hybrids and proteins. These structures, positioned at every ≈50 kbp throughout the genome, may anchor supercoiled chromatin domains to the nuclear matrix. In view of the persistent nature of these discontinuities and their distribution, we propose that the eukaryotic genomic DNA is composed of contiguous rather than continuous ss, interrupted at the boundaries of interphase chromatin loops.

Materials and Methods

Nuclear Halo Preparation and FI-SCGE. The cells were washed twice with PBS (pH 7.4), embedded into 1% low melting-point (LMP)-agarose prepared in TE (10 mM Tris·l/1 mM EDTA, pH 8.0), and pipetted onto the surface of glass slides precoated with LMP agarose. Cell numbers were set to 5–8 $\times$ 10^5 per slide. Where indicated, before agarose encapsulation, the cells were treated with (*i*) the Top II drugs ICRF-193 (100 μM for 15 min; Fig. 5) or etoposide (40 μM for 6 h; SI Fig. 6*D*) or (*ii*) bromodeoxyuridine (100 μM for 40 min; Fig. 5). Nuclear halos were prepared by lysing the cells in a high-salt buffer (2 M NaCl/10 mM Tris·Cl, pH 8/50 mM EDTA/0.5% Triton X-100/0.2% Sarkosyl/10% DMSO) for 7 min, on ice. The disassembly of the nuclear halos to granules containing ≈50-kbp ssDNA was visualized by FI-SCGE by using an MJ Research (Waltham, MA) PPI 200 power inverter; electrophoresis was carried out in alkaline (300 mM NaOH/1 mM EDTA) or neutral (1× TAE) conditions for 12 h at 4°C; parameters were set to maximize resolution in the ≈50- to 500-kbp range. Where indicated, nuclear halos were extensively digested by proteinase K before FI-SCGE (Fig. 5). After running, each sample was neutralized in 1 M Tris·HCl, pH 8/10 mM EDTA for 5 min and then dehydrated in an ascending series of ethanol for 2 min, each at 4°C, and finally in methanol and acetone for 15–15 min, each at −20°C. In the experiment shown in Fig. 2*A*, the nuclear halos and FI-SCGE comets were scraped off from the surface of the glass slides, and, after β-agarase and proteinase K digestion, the DNA molecules were reisolated by phenol-chloroform extraction/ethanol precipitation and analyzed by FIGE in a 1% agarose gel prepared in 1× TAE.

Because of space limitation, the detailed description of FI-SCGE FISH, *in situ* nick labeling, microscopy, and other methods used in this study can be found in *SI Text*; SI Figs. 6–8 accompany the paper.

Downloaded by guest on November 19, 2021

dc_1977_21

We are grateful to Dr. Peter Nagy for valuable help in image processing and Gyöngyi Dajka for conscientious technical assistance. We thank Drs. S. H. Leppla (Bacterial Toxins and Therapeutics Section, National Institute of Allergy and Infectious Diseases, National Institutes of Health, Bethesda, MD) for the DNA/RNA hybrid-specific mouse monoclonal antibody S9.6, G. Capranico (Department of Biochemistry, University of Bologna School of Pharmacy, Bologna, Italy) for the Δtop1, I. D. Hickson (Cancer Research UK Laboratories, Weatherall Institute of Molecular Medicine, University of Oxford, John Radcliffe Hospital, Oxford, U.K.) for the ts-top2-mutant yeast strains, C. Austin (Institute for Cell and Molecular Biosciences, The Medical School, University of Newcastle-upon-Tyne, Newcastle-upon-Tyne, U.K.) for a gift of ICRF-193, J. Markovits and A. Jacquemin-Sablon (INSERM E362, Université Victor Segalen Bordeaux 2, Bordeaux, France) for the topoisomerase IIβ-deficient Chinese hamster cells, P. D. Aplan (Genetics Branch, Laboratory of Cellular Oncology, Center for Cancer Research, National Cancer Institute, National Institutes of Health) for the ML-1 cells and the 0.2 HB MLL bcr probe, and A. Udvardy (Institute of Biochemistry, Biological Research Center of the Hungarian Academy of Sciences, Szeged, Hungary) and J. Aradi (Department of Biochemistry and Molecular Biology, Debrecen, Hungary) for discussions and advice. This work was supported by the Hungarian Ministry of Education and Ministry of Health (Grants OTKA T046945, OTKA T048742, ETT 614/2003, and ETT 067/2006), the Hungarian Scientific Research Fund (Grants OMFB-01590/2006 and OMFB-01626/2006), and the Hungarian National Office of Research and Technology (in the program Öveges).

1. Nickerson J (2001) *J Cell Sci* 114:463–474.
2. Berezney R (2002) *Adv Enzyme Regul* 42:39–52.
3. Gromova II, Nielsen OF, Razin SV (1995) *J Biol Chem* 270:18685–18690.
4. Hershey HV, Werner D (1976) *Nature* 262:148–150.
5. Weintraub H (1984) *Cell* 38:17–27.
6. Oberhammer F, Wilson JW, Dive C, Morris ID, Hickman JA, Wakeling AE, Walker PR, Sikorska M (1993) *EMBO J* 12:3679–3684.
7. Szabo G, Jr, Bacso Z (1996) *Cell Death Differ* 3:237–241.
8. Szekvolgyi L, Hegedus E, Molnar M, Bacso Z, Szarka K, Beck Z, Dombradi V, Austin C, Szabo G (2006) *Histochem Cell Biol* 125:63–73.
9. Szabo G, Jr (1995) *Exp Cell Res* 221:320–325.
10. Gromova II, Thomsen B, Razin SV (1995) *Proc Natl Acad Sci USA* 92:102–106.
11. Bacso Z, Everson RB, Eliason JF (2000) *Cancer Res* 60:4623–4628.
12. Jackson DA, Dickinson P, Cook PR (1990) *EMBO J* 9:567–571.
13. Aplan PD, Chervinsky DS, Stanulla M, Burhans WC (1996) *Blood* 87:2649–2658.
14. Stanulla M, Wang J, Chervinsky DS, Thandla S, Aplan PD (1997) *Mol Cell Biol* 17:4070–4079.
15. Strissel PL, Strick R, Rowley JD, Zeleznik L (1998) *Blood* 92:3793–3803.
16. Hu Z, Zhang A, Storz G, Gottesman S, Leppla SH (2006) *Nucleic Acids Res* 34:e52.
17. Vogelstein B, Pardoll DM, Coffey DS (1980) *Cell* 22:79–85.
18. Strout MP, Mrozek K, Heinonen K, Sait SN, Shows TB, Aplan PD (1996) *Genes Chromosomes Cancer* 16:204–210.
19. Gilbert N, Allan J (2001) *Proc Natl Acad Sci USA* 98:11949–11954.
20. Schroeter D, Werner D, Meinzer P (1981) *Eur J Cell Biol* 24:131–138.
21. Werner D, Krauth W, Hershey HV (1980) *Biochim Biophys Acta* 608:243–258.
22. Werner D, Hadjiolov D, Neuer B (1981) *Biochem Biophys Res Commun* 100:1047–1054.
23. Juodka B, Pfutz M, Werner D (1991) *Nucleic Acids Res* 19:6391–6398.
24. Barboro P, D'Arrigo C, Diaspro A, Mormino M, Alberti I, Parodi S, Patrone E, Balbi C (2002) *Exp Cell Res* 279:202–218.
25. Barboro P, D'Arrigo C, Mormino M, Coradeghini R, Parodi S, Patrone E, Balbi C (2003) *J Cell Biochem* 88:113–120.
26. Fey EG, Krochmalnic G, Penman S (1986) *J Cell Biol* 102:1654–1665.
27. Herman R, Weymouth L, Penman S (1978) *J Cell Biol* 78:663–674.
28. Muller M, Spiess E, Werner D (1983) *Eur J Cell Biol* 31:158–166.
29. Dimitrova DS (2006) *Genes Cells* 11:829–844.
30. MacAlpine DM, Rodriguez HK, Bell SP (2004) *Genes Dev* 18:3094–3105.
31. Ju BG, Lunyak VV, Perissi V, Garcia-Bassets I, Rose DW, Glass CK, Rosenfeld MG (2006) *Science* 312:1798–1802.
32. Abdurashidova G, Radulescu S, Sandoval O, Zahariev S, Danailov MB, Demidovich A, Santamaria L, Biamonti G, Riva S, Falaschi A (2007) *EMBO J* 26:998–1009.
33. Li X, Manley JL (2005) *Cell* 122:365–378.

Downloaded by guest on November 19, 2021

Published online 24 August 2018 dc_1977_21 *Nucleic Acids Research, 2018, Vol. 46, No. 20* ***10649–10668***
doi: 10.1093/nar/gky743

Endogenous single-strand DNA breaks at RNA polymerase II promoters in *Saccharomyces cerevisiae*

Éva Hegedüs[1], Endre Kókai[2], Péter Nánási[1], László Imre[1], László Halász[3], Rozenn Jossé[4], Zsuzsa Antunovics[5], Martin R. Webb[6], Aziz El Hage[7], Yves Pommier[4], Lóránt Székvölgyi[1,3], Viktor Dombrádi[2] and Gábor Szabó[1,*]

[1]Department of Biophysics and Cell Biology, Faculty of Medicine, University of Debrecen, Debrecen, Hungary, [2]Department of Medical Chemistry, Faculty of Medicine, University of Debrecen, Debrecen, Hungary, [3]MTA-DE Momentum Genome Architecture and Recombination Research Group, Department of Biochemistry and Molecular Biology, Faculty of Medicine, University of Debrecen, Debrecen, Hungary, [4]Developmental Therapeutics Branch and Laboratory of Molecular Pharmacology, Center for Cancer Research, National Cancer Institute (CCR-NCI), NIH, Bethesda, MD, USA, [5]Department of Genetics and Applied Microbiology, Faculty of Science and Technology, University of Debrecen, Debrecen, Hungary, [6]The Francis Crick Institute, London NW1 1AT, UK and [7]Wellcome Trust Centre for Cell Biology, University of Edinburgh, Edinburgh, UK

Received February 07, 2018; Revised July 13, 2018; Editorial Decision August 02, 2018; Accepted August 16, 2018

ABSTRACT

Molecular combing and gel electrophoretic studies revealed endogenous nicks with free 3′OH ends at ∼100 kb intervals in the genomic DNA (gDNA) of unperturbed and G1-synchronized *Saccharomyces cerevisiae* cells. Analysis of the distribution of endogenous nicks by Nick ChIP-chip indicated that these breaks accumulated at active RNA polymerase II (RNAP II) promoters, reminiscent of the promoter-proximal transient DNA breaks of higher eukaryotes. Similar periodicity of endogenous nicks was found within the ribosomal rDNA cluster, involving every ∼10th of the tandemly repeated 9.1 kb units of identical sequence. Nicks were mapped by Southern blotting to a few narrow regions within the affected units. Three of them were overlapping the RNAP II promoters, while the ARS-containing IGS2 region was spared of nicks. By using a highly sensitive reverse-Southwestern blot method to map free DNA ends with 3′OH, nicks were shown to be distinct from other known rDNA breaks and linked to the regulation of rDNA silencing. Nicks in rDNA and the rest of the genome were typically found at the ends of combed DNA molecules, occasionally together with R-loops, comprising a major pool of vulnerable sites that are connected with transcriptional regulation.

INTRODUCTION

Recent observations in several mammalian experimental models suggest that transcriptional activation of RNA polymerase II (RNAP II) dependent genes frequently involves formation of DNA strand breaks, supposedly elicited by topoisomerase 2β (Top 2β) (1–7). Indeed, this enzyme has been mapped to the 5′ end of transcriptionally active genes in several experimental systems (8–10). These breakages are generally interpreted in the context of the requirement for topological relaxation during transcription, although elongation by RNAP rather than initiation is readily explained in terms of the twin supercoil domain model (11–13), while strand opening at initiation is facilitated by negative supercoiling (14,15). Transient discontinuities have been detected at the promoters of several genes by 3′OH end-labeling; these are generally perceived as double strand (ds) breaks even though the two subunits of Top2 are known to work independently of each other and the DNA discontinuities generated by this enzyme are partly single-strand (ss) breaks/nicks (9,16). Furthermore, DNA is sensitive to mechanical damage at ss breaks (data in (17)), so ds breaks may be indirectly generated at the site of Top2 activity. Mechanical breakages may include those that arise due to an abrupt release of torsional stress upon deproteinization. On the other hand, non-random ss breaks may also arise when topoisomerases, and perhaps some other nucleases present in a cell lysate, might cleave one strand at accessible sites so that the relaxed DNA becomes resistant to further cleavages (similarly to the observations in (18)), that would result in a single nick in each supercoiled loop. In addition, many restriction enzymes generate ss incisions at imperfect recognition motifs (19,20), that are easily mistaken for endoge-

*To whom correspondence should be addressed. Tel: +36 52 412623; Fax: +36 52 532201; Email: szabog@med.unideb.hu

© The Author(s) 2018. Published by Oxford University Press on behalf of Nucleic Acids Research.
This is an Open Access article distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/4.0/), which permits unrestricted reuse, distribution, and reproduction in any medium, provided the original work is properly cited.

Downloaded from https://academic.oup.com/nar/article/46/20/10649/5078800 by guest on 26 October 2021

dc_1977_21

nous breaks. Furthermore, the 3′ end-labeling procedures may also be misleading since the 3′end of the RNA moiety in R-loops (three-stranded structures composed of an RNA:DNA hybrid and a displaced ss DNA strand (21,22)) could also serve as starting points for either terminal transferase (23) or DNA polymerase activity (24) that may be mistakenly interpreted as nicks. In the light of these experimental challenges, the findings on endogenous breaks are often regarded as controversial and clarification of the questions concerning their origin and function calls for alternative, independent approaches. To overcome these caveats, one possibility would be to have a global view on all of the endogenous nicks in the genome rather than focusing on particular loci, in better known model organisms, using novel methods.

It has been reported in our earlier published studies that conventional procedures involving extensive proteolytic digestion of lysed cells followed by phenol–chloroform extraction or purification based on silica-adsorption, yield ∼50 kb ds DNA fragments (25,26). Remarkably, this average fragment size overlaps that of the DNA loops (27) and coincides with the estimated size of the functional (transcriptional, replicative) units of chromatin (28–32). According to the common view, such loop-size ds fragmentation could be a direct consequence of random mechanical breakage, since it can be avoided when the cells are embedded into agarose plugs before lysis. Importantly however, a similar fragmentation was observed even when the DNA in agarose plugs was treated with S1 nuclease (33), an enzyme that preferentially cleaves at nicks and ss DNA regions (34). Moreover, loop-size fragmentation occurs upon rapid alkaline lysis of cells (26,35), upon urea/heat-denaturation of intact chromatin embedded in agarose (36,37), and also when DNA is isolated from fixed cells (38). Using *in situ* nick-translation and single-cell gel electophoresis we have previously shown that preformed ss discontinuities, i.e. nicks are scattered over eukaryotic chromatin, including that of *Saccharomyces cerevisiae*, delimiting loop-size domains (37). All these observations could be collectively interpreted in terms of persistent, endogenous ss breaks that may yield ds breaks depending on the experimental conditions. We hypothesize that the promoter-proximal, transcriptional activity-related DNA breaks that have been reported in mammalian cells (reviewed in (39)) represent a major subpopulation of the endogenous nicks detected by us at loop-size intervals in various eukaryotic systems (26,33,36–38).

Based on this hypothesis, we exploited the advantages offered by assessing endogenous DNA breaks in a global manner over the genomic DNA (gDNA) of *S. cerevisiae*. To investigate the possible sequence-related and sequence-unrelated determinants of the breaks, the rDNA cluster containing naturally amplified units of identical sequence was also included in our study. The ∼1–2 Mb rDNA cluster is located at the *RDN1* locus in the right arm of chromosome XII (chr XII) in budding yeast and consists of a tandem array of 100–200 repeated transcription units. The rDNA organizes the nucleolus wherein transcription of ribosomal RNA (rRNA) and the early steps of ribosome biogenesis take place (40,41). A perinucleolar protein network including the replication fork barrier (RFB)-associated protein Fob1 and the topoisomerase 1 enzyme (Top1) tethers the rDNA repeats to the nuclear membrane, thus separating the locus from the bulk DNA inside the nucleus, and also ensures repeat stability (42). Top1mediates ss nicking and re-ligation in a process involving stabilized Top1 covalent cleavage complexes (Top1cc) formed *via* the 3′OH of the nicked DNA (17).

The 9.1 kb rDNA units (see Figure 4A) harbour the genes for 5.8S, 25S and 18S rRNAs which are transcribed by RNAP I as a single precursor (i.e. 35S pre-rRNA). Each unit contains also a 5S rRNA gene transcribed by RNAP III from the opposite strand. A replication origin (rARS element) is positioned outside the transcribed regions within IGS2. Although DNA replication begins bidirectionally, the rightward-moving fork is arrested at the RFB within IGS1, while the other fork proceeds through about five repeats until it terminates at a stalled rightward-moving fork of the closest replication unit. These RFBs prevent the replisome from head-on collison with the RNAP I transcription apparatus, so that rRNA transcription can proceed even during S-phase (43). About half of the units are transcriptionally active at the same time (44–46) and are transcribed at high efficiency by ∼50 RNAP I complexes per 35S pre-rRNA (47). About 20% of all available rARSs are used as replication origins during a single S phase (48). There are also a few genes that are transcribed by RNAP II within the rDNA units. Accordingly, RNAP II can be detected by chromatin immunoprecipitation (ChIP) in the corresponding promoter regions (49). A preferential association of DNA breaks with the rARS regions or with the regulatory regions of any of the transcribed genes would indicate that the mechanism that generates the DNA breaks is associated with rDNA replication or transcription, respectively.

Here, we made use of enzymatic labeling of endogenous free 3′OH ends in conjunction with molecular combing, microarray analyses, and a novel reverse Southwestern (rSW) blotting procedure in order to characterize the endogenous nicks. The nicks were mapped genome wide, including the repetitive rDNA locus. The specific localization of the nicks together with the comparison of mutant yeast strains suggest that their biological function is associated with transcription. Importantly also, the picture that emerges from our observations implicates these DNA breaks in genomic instability.

MATERIALS AND METHODS

For *cultivation and synchronization of S. cerevisiae and S. pombe and description of the common gel electrophoretic techniques* see Supplementary Materials and Methods. The budding yeast strains used in the present study are listed in Table 1.

Preparation of agarose plugs containing yeast chromosomes

Preparation of *S. cerevisiae* agarose-plugs was carried out as previously described (33). Briefly, *S. cerevisiae* cells were harvested and washed twice in 50 mM EDTA (pH 8.0), then resuspended in digestion solution (0.9 M sorbitol, 0.125 M EDTA,100 mM dithiothreitol (DTT)) containing 2 mg/ml lyticase enzyme (Sigma-Aldrich). Samples

Downloaded from https://academic.oup.com/nar/article/46/20/10649/5078800 by guest on 26 October 2021

Table 1. List of *S. cerevisiae* strains used in this study

Strain	Genomic background	Genotype	Source
WDHY199	W303	Mat a, leu2-3,112, trp1-283, ura3-52, his7-2, lys1-1	Wolf-Dietrich Heyer lab
BY4741	WT	Mat a, his3Δ1, leu2Δ0, met15Δ0, ura3Δ0	EUROFAN
top1Δ	BY4741	Mat a, his3Δ1, leu2Δ0, met15Δ0, ura3Δ0, top1Δ::KanMX	EUROFAN
bar1Δ	BY4741	Mat a; his3D1; leu2D0; met15D0; ura3D0; YIL015wΔ::KanMX4	EUROFAN
fob1Δ	BY4741	Mat a, his3Δ1, leu2Δ0, met15Δ0, ura3Δ0, fob1:: KanMX4	EUROFAN
sir2 Δ	BY4741	Mat a, his3Δ1, leu2Δ0, met15Δ0, ura3Δ0, sir2:: KanMX4	EUROFAN

were mixed with an equal volume of 1.5% low melting point (LMP) agarose (Sigma-Aldrich) dissolved in 0.9 M sorbitol/0.125 M EDTA. Aliquots were allowed to harden in sample molds at 4°C for 5 min, and then placed into 0.9 M sorbitol/0.125 M EDTA at 37°C for 6 h. Each plug contained ∼3 × 10^8 cells. The plugs containing yeast spheroplasts were digested with 0.5 mg/ml Proteinase K (Thermo Fisher Scientific) in lysing solution (0.5 M EDTA, 10 mM Tris–HCl, 1% SDS, pH 8.0) at 55°C for 2 days, then washed with TE (10 mM Tris–HCl, 2 mM EDTA, pH 8.0) and treated by 0.75 μM phenyl-methyl-sulfonyl-fluoride (PMSF, Sigma-Aldrich) at 37°C for 10 min in order to inactivate residual proteinase activity. Finally, the plugs were washed with TE and stored in the same buffer at 4°C.

In-gel enzyme digestion and labeling of yeast gDNA in agarose plugs

S1 nuclease digestion. Plugs were washed in S1 buffer (0.2 M NaCl, 50 mM Na-acetate, 1 mM $ZnSO_4$, 0.5% glycerol, pH 4.5) three times for 30 min, then incubated with 500 U/ml S1 nuclease (Promega Biosciences Inc.) in S1 buffer for 1.5 h at 37°C (33).

Restriction endonuclease digestion. Plugs were washed in the appropriate 1× restriction buffer three times for 1 h and then further incubated with 150 U/ml restriction enzyme (Sfi I, Sma I, Mlu I, Pvu II or Stu I; Thermo Fisher Scientific) in 1× restriction buffer for 16 h at 30, 37 or 55°C according to the manufacturers' recommendations.

In situ nick-translation using DNA polymerase I and biotinylated nucleotides. Plugs were used to incorporate biotin-dUTP by nick-translation performed under limiting conditions ('limiting nick-translation') that restricted incorporation to <200 bp regions (see Supplementary Figure S6). After washing in 1× DNA polymerase I buffer (50 mM Tris–HCl (pH7.5), 10 mM $MgCl_2$, 1 mM DTT) three times for 20 min, the plugs were incubated with 150 U/ml DNA polymerase I (Thermo Fisher Scientific) in DNA polymerase I buffer containing 1 μM dNTP mix (1 μM biotin–dUTP, 1 μM dATP, 1 μM dCTP and 1 μM dGTP) and 5 μM ddNTP mix (5 μM ddATP, 5 μM ddTTP, 5 μM ddCTP and 5 μM ddGTP) for 30 min on ice to allow equilibration, then for 20 min at 37°C. The optimal dNTP/ddNTP ratio was determined by using a PCR product nicked at a specific site (Supplementary Figure S6). In the case of standard, 'non-limiting nick-translation', the ddNTP mix was omitted from the reaction.

Terminal deoxynucleotidyl transferase labeling. The reaction was performed using 260 U/ ml Terminal deoxynucleotidyl Transferase (TdT; Thermo Fisher Scientific) in its own buffer and 1 μM biotin-dUTP. The other conditions of the reaction were similar to those of the nick-translation.

Combined RNase treatment (RNase HI, A, H2). Agarose plugs were digested with 12.5 U/plug RNase HI (Thermo Fisher Scientific) in 1× RNase HI buffer (20 mM Tris–HCl (pH 7.8), 40 mM KCl, 8 mM $MgCl_2$, 1 mM DTT) at 37°C overnight, subsequently with 5 μl/plug of 10 mg/ml RNase A (Thermo Fisher Scientific) in TE for 1 h at room temperature then with 5 μl/plug of human RNase H2 enzyme (gift from Martin Reijns, University of Edinburgh) in RNase H2 buffer (60 mM KCl, 50 mM Tris–HCl pH 8, 10 mM $MgCl_2$, 0.01% BSA, 0.01% Triton X-100) at 37°C overnight. Before each digestion, the plugs were equilibrated with the appropriate enzyme buffer three times for 50 min. All digestions were performed in 150 μl reaction volumes.

Molecular combing

Molecular combing was performed on whole gDNA or on isolated chr XII, as described in (50). Genomic DNA, or chr XII isolated by CHEF, were embedded in agarose plugs and nick-labeled by biotinylated nucleotides (in limiting or non-limiting conditions, described above). chr XII was isolated from the plugs containing the whole genome by running this chromosome into a block of 0.5% LMP agarose inserted into a 1% standard agarose gel, then this block was cut out without EBr staining. To solubilize the whole genomic or chr XII DNA, 1.6 ml 0.1 M MES (pH 6.5) was added to each plug, incubated at 70°C for 20 min, then at 42°C for 10 min. The blocks were dissolved by 8 U Agarase (Thermo Fisher Scientific) treatment at 42°C overnight. In the case of combing of λ phage DNA, site-specific nicks were introduced by Nt.BbvCI nickase cutting 7 times in the phage genome (delimiting 306, 318, 614, 3977, 8013 and 12 451 bp fragments): 1.5 μg λ DNA was incubated with 50 U/ml Nt.BbvCI nickase (New England Biolabs) in 20 μl CutSmart buffer for 30 min at 37°C.

The DNA solutions were placed at room temperature and transferred to disposable reservoirs without pipetting. DNA combing was performed by the combing apparatus of Genomic Vision (France) according to the manufacturer's instructions using 22 × 22 mm vinylsilane coated coverslips (from the same source). After combing, the coverslips were glued to glass slides with cyanoacrylate glue. Nonspecific binding of the antibodies was blocked by incubation with

Downloaded from https://academic.oup.com/nar/article/46/20/10649/5078800 by guest on 26 October 2021

dc_1977_21

30 μl of 5% BSA/1× PBS/0.1% Triton X-100 (temporarily covering the combed sample with a clean coverslip) for 20 min in a humid chamber. The biotin molecules incorporated into the DNA were visualized by indirect immunofluorescent labeling with 1:60 diluted mouse anti-biotin as a primary antibody (Sigma-Aldrich). For R-loop detection, 33 μg/ml RNA:DNA hybrid specific S9.6 primary antibody (used as in (37); hybridoma from ATCC) was applied in 1% BSA/1× PBS/0.1% Triton X-100 for 45 min at room temperature in a humid chamber. After washing 3 times with 3 ml 1× PBS, twice with 30 μl PBS/0.1% Triton X-100 and then once with 30 μl PBS for 5 min, Alexa Fluor 647 conjugated goat anti-mouse antibody (Life Technologies) was used as a secondary antibody at a final concentration of 17 μg/ml in 1% BSA/1× PBS/0.1% Triton X-100, at room temperature in a humid chamber, for 45 min. In some experiments the enzymatically incorporated biotin and the RNA:DNA-hybrids were detected simultaneously on the same sample, using goat anti-mouse Abberior STAR 580 and streptavidin Abberior STAR RED as secondary reagents, respectively. After immunofluorescence labeling the coverslips were washed as before. DNA staining was performed by 30 μl YOYO-1 dye (Thermo Fisher Scientific) diluted 1:5000 in 0.1 M MES (pH 6.5) in a humid chamber, in the dark, for 20 min. The coverslips were covered by ProLong® Gold Antifade using a clean coverslip and placed at 4°C overnight. Imaging was carried out in an Olympus FluoView 1000 confocal laser scanning microscope equipped with 488 and 633 nm lasers, using a 60× oil immersion oil objective.

Genome-wide mapping of nicks (Nick ChIP-chip)

Saccharomyces cerevisiae (BY4741) cells were fixed in 1% formaldehyde (10 min, RT) and excess formaldehyde was quenched with 0.7 M glycine. Fixed cells were embedded into agarose plugs prepared according to standard protocols. Spheroplasts were obtained by lyticase digestion (Sigma-Aldrich) performed at 37°C for 3 h, followed by cell lysis in 0.43 M EDTA, 1% (v/v) Sarcosyl, 0.01 M Tris (pH 8) and 15 U/ml Proteinase K (55°C, 72 hours). At this point, one half of the sample was treated with nicking enzyme Nb.Bpu10I which recognizes CCTNA∧GC sites. Nickase plus and nickase minus samples were processed in parallel in the subsequent steps. Tagging of nicks was performed by incorporating biotin-dCTP/dUTP in the presence of chain terminator ddNTPs by the *E. coli* DNA polymerase I holoenzyme. Plugs were equilibrated in 2 ml of DNA polymerase I buffer for 3 × 50 min and then transferred to ice for 30 min in the nick-translating mix (consisting of 1× DNA polymerase I buffer, 5 μM of each ddNTPs, 1 μM of dATP, dGTP, biotin-dCTP, biotin-dUTP and 150 U/ml of DNA polymerase I). The reaction was initiated by transferring the tubes to 37°C for 30 min with gentle shaking, then stopped by washing the plugs in excessive amounts of 0.5 M EDTA. Agarose plugs were digested/solubilized with 2 U of β-agarase and sonicated (Bioruptor, Diagenode). Nucleic acids were purified by a PCR cleanup kit (Macherey-Nagel). RNA was digested with 10 μg/ml of RNase A in low-salt conditions (10 mM NaCl) at 37°C for 60 min. Immunoprecipitation of nicked DNA was performed according to the standard ChIP protocol for yeast using a monoclonal anti-biotin antibody (Sigma-Aldrich) coupled to Protein G coated Dynabeads (Thermo Fisher). Two-thirds of the immunoprecipitated DNA (IP) and an equal amount of input DNA were amplified by random primer extension followed by PCR, incorporating amino-allyl dUTP for subsequent dye coupling. Input and IP DNAs were labeled with Cy3 and Cy5, respectively, mixed in equal amounts, and hybridized to Agilent 4 × 44K whole genomic microarrays in 1× hybridization buffer for 16 h at 65°C. Microarray slides were scanned with an Axon 4000B scanner (using the GenePix5.1 software). In Figure 2, representative data of two (together with the nickase-digested sample of Supplementary Figure S8: three) biological replicates are shown.

Data were analyzed and nick peaks were called by the COCAS ChIP On Chip Analysis Suite (51). Genomic positions of annotated transcription units were obtained from the sacCer3 genomic assembly. The efficiency of cleavage/labeling was such that about 7% of the Nb.Bpu10I recognition sites were detected (Supplementary Figure S8A). Specificity of labeling was documented by the statistically highly significant coincidence of labeling with the nickase sites (Supplementary Figure S8B and C). The labeled sites in the nickase treated sample that coincide neither with the nickase recognition sites, nor with the endogenous nicks revealed in the sample without nickase treatment may be due to unrevealed endogenous nicks, or, less likely, off-target labeling.

Enrichment analysis

To estimate the enrichment or depletion of nick peaks within annotation categories we chose to follow a previously described permutation test (52). Briefly, fold-change values were calculated by dividing the observed intercepting nucleotide occupancies of each annotation category with the mean occupancy of 1000 computer randomized (simulated) peaks. Random regions were generated with *shuffleBed* (53) with respect to the original peak sizes and chromosomal distribution. Significant difference was assigned with a two-tailed proportion test (http://www.socscistatistics.com/tests/ztest/).

Signal density analysis

Mean RNAPII (54) around nicks (±1000 bp) or Nick ChIP-chip signal intensities around TSSs (±1500 bp) were calculated for 100/300 bp bins using Deeptools (55). The generated profiles were further processed and plotted in R. Random regions were generated with *shuffleBed* (53) with respect to the original peak sizes and chromosomal distributions. Significant difference was assigned with K–S tests or t-tests after checking the normality of distributions using the Shapiro-Wilk test.

Gene expression correlation

For each nick (peak) position ($n = 215$) we assigned the closest protein coding gene with *closestBed* (53) and their expression values were plotted as a boxplot. We randomly sampled 215 genes from the total gene pool 1,000 times

Downloaded from https://academic.oup.com/nar/article/46/20/10649/5078800 by guest on 26 October 2021

and compared their expression levels with that of the nick-associated genes. Overlapping genes were considered based on their average expression level. Significant difference was assigned with the Mann-Whitney test; *P*-values were corrected with the Benjamini & Hochberg method. Expression values for wild-type *S. cerevisiae* genes were obtained from GEO (GSE98435) (56). Correlation with RNAP II chromosomal binding was analyzed using publicly available ChIP-Chip datasets downloaded from GEO (GSE6293) (54).

Mapping of S1-sensitive sites by Southern-blot analyses

Log-phase *S. cerevisiae* cells were fixed in 1% HCHO as described above and used to prepare agarose embedded spheroplasts. The plugs were digested with S1 nuclease and rare cutting restriction enzymes (Sfi I, Sma I, Mlu I or Hind III, single-cutter restriction enzymes in the rDNA units), applied in a sequence indicated in Figures 3–5, S12 and S14, and the restriction fragments were separated by conventional or urea/heat-agarose gel electrophoresis. The gels were blotted to Hybond N^+ membrane (Amersham GE Healthcare Life Science) by a vacuum blotter (Bio-Rad Model 785).

A 1405 bp ds fragment covering the rDNA region shown in Figure 4A, amplified using the primers (pRDs (5′-GGG GAT CGA AGA TGA TCA GA-3′; Integrated DNA Technologies (IDT)) and pRDas (5′-GAA AAG GCC AGC AAT TTC AA-3′; IDT), served as the template for the preparation of single-stranded probes. First, linear amplification was performed using 2.5 U Taq polymerase (Thermo Fisher Scientific), in 50 μl of 1× reaction buffer (10 mM Tris–HCl, 50 mM KCl, 0.08% Nonidet P-40, pH 8.8) containing 3 mM $MgCl_2$, 200 ng template DNA, 20 pmol of either pRDs (sense) or pRDas (antisense) primer, dATP, dTTP, dGTP and dCTP at 0.25 mM concentration (from Promega Life Science, Madison, USA). The probes were purified on Sephadex G-25 spun columns for random primer labeling using [α^{32}P]-dCTP (6000 Ci/mmol, 10 mCi/ml; Institute of Isotopes LTD, Budapest), as described earlier (57). The probes were denatured for 10 min at 100°C and kept on ice for 5 min before hybridization. Note that a small amount of complementary strand (derived from the co-purified original template DNA) was also present in the single stranded probes. The radioactive signal was captured by Phospho-screen (Kodak) and was visualized by a BIO-RAD Phospho-Imager.

For the estimation of ds or ss fragment sizes in Southern blot experiments calibration curves were constructed for ds and ss DNA molecules, based on ds DNA ladders and mixtures of denatured rDNA PCR products, respectively.

Reverse Southwestern blot

rSW blot (58) was carried out following a protocol developed by us, that is described below and is explained in Supplementary Figure S15. Plugs containing *S. cerevisiae* gDNA after limiting *in situ* nick-translation with DNA polymerase I (see above) were digested in lysing solution (0.5 mg/ml Proteinase K, 1% SDS, 0.5 M EDTA, 10 mM Tris; pH 8.0) at 54°C for 30 min, then washed with TE and treated with 0.25 mM PMSF at 37°C for 10 min to inactivate residual proteinase activity. Finally, the plugs were digested with the rare cutting restriction endonucleases Sfi I or Sma I. The DNA fragments were separated on 1% agarose gel, and the 9.1 kb rDNA units were carefully cut out from the gel and were further digested for 16 h at 37°C with the mixture of two different restriction endonucleases (Pvu II + Stu I). The restriction fragments of rDNA units were separated on 1.2% agarose gels and vacuum-blotted on Immobilion-P Transfer Membrane (PVDF, 0.45 μm, Millipore). These were prehybridized for 1 hour at room temperature in 5 ml prehybridization solution (1% BSA, 0.2% Tween-20/1× PBS), then incubated with mouse anti-biotin primary antibody at a dilution of 1:1000 in 5 ml hybridization solution for 16 hours at 4°C. After washing with 0.2% Tween-20/1× PBS for 5 × 5 min, the membranes were incubated with goat anti-mouse IgG antibody conjugated with horseradish peroxidase (dilution 1:2500) at room temperature, for 1.5 h. We found that the choice of the membrane and of the blocking agent were crucial. The use of milk powder as a blocking agent is not satisfactory in the case of biotinylated targets, because its biotin content gives a strong background, therefore BSA is recommended as a blocking agent. The signal was detected by chemiluminescence (KODAK Medical X-ray Processor). The dynamic detection range of the method was calibrated with biotinylated PCR products (Supplementary Figure S16). When the membranes were stripped and re-probed for the detection of R-loops in rDNA units with the RNA:DNA hybrid specific S9.6 antibody (used at 1 μg/ml concentration), the membranes were washed four times for 5 min each in 0.2% Tween-20/1× PBS, then incubated in stripping buffer (62.5 mM Tris base, 2% SDS, 0.7% 2-mercaptoethanol, pH 6.8) for 30 min at 50°C, then washed six times for 5 min in 0.2% Tween-20/1× PBS. Blots were evaluated by the ImageJ software. For quantification, the biotin signal intensities of the fragments relative to their EBr signal intensities were calculated. To compare the intensities in different yeast cells, the calculated biotin/EBr ratios were normalized to the band having the highest ratio on the gel (taken as 1.0 on the Y-axis of the histogram). Since the ratios are proportional to the incidence of nicks in a given amount of DNA (EBr signal) the observed changes are independent of differential loading and/or the copy number of rDNA units.

RESULTS

Endogenous ss breaks in *S. cerevisiae* gDNA are revealed by molecular combing and gel electrophoretic analyses

The purpose of the experiments described below was to detect endogenous DNA breaks via their free 3′OH groups and to determine whether they belong to ss or ds termini. DNA strand breaks with free 3′OH could be visualized in molecular combing experiments when biotinylated nucleotides were incorporated into agarose embedded deproteinized gDNA by limited nick-translation (mixing terminator nucleotides with dNTPs; see Materials and Methods). As shown in Figure 1, the combed DNA molecules were similar to or larger than the combed λ phage DNA (Figure 1A, B versus F), and often carried the 3′-label at the fragment ends, sometimes on both sides (Figure 1A). Similar labeling was detected in α-factor-synchronized G1

Downloaded from https://academic.oup.com/nar/article/46/20/10649/5078800 by guest on 26 October 2021

dc_1977_21

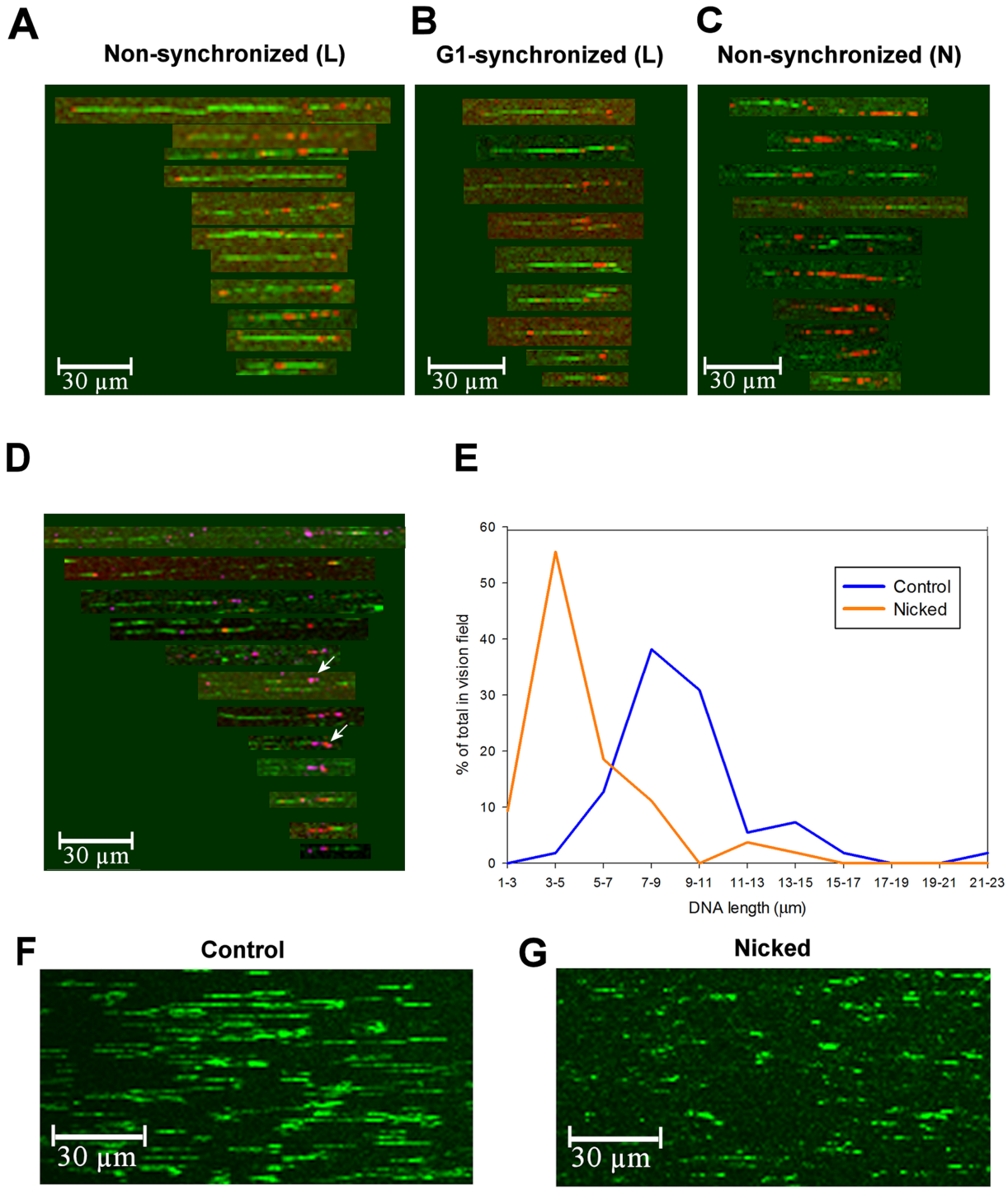


Figure 1. Breakage of *S. cerevisiae* gDNA and λ DNA at preformed ss nicks upon molecular combing. (**A-C**) Molecular combing of nick-translated gDNA from *S. cerevisiae*. Biotinylated nucleotides were incorporated by nick-translation conducted either in limiting (L), or non-limiting/standard conditions (N), into agarose-embedded gDNA of unperturbed, non-synchronized (A and C), or G1-synchronized (B) BY4741 cells. Biotin was detected by AlexaFluor 647-conjugated anti-biotin antibody (red) and DNA molecules were stained with YOYO-1 (green). Panel **D** shows examples of co-localization of nicks labeled with TdT (magenta) and R-loops labeled with the RNA:DNA hybrid specific S9.6 antibody (red), when both entities were visualized in the same sample. The percentage of co-labeled spots was estimated ∼10% of all nick-related DNA associated spots. Arrows indicate examples of co-localization. (**E–G**) Molecular combing of λ DNA. Representative images of YOYO-1 stained (green) control (F) and Nt.BbvCI nickase-treated (G) λ DNA. The size distribution histograms of combed DNA molecules before (blue) and after (orange) nickase treatment are shown in panel E. The full length intact ds λ DNA (48.5 kb) corresponds to 16.2 µm (calculated with 3 bp/nm helical repeat length), i.e. the majority of λ DNA molecules were fragmented after combing alone. Images of DNA fibers were assembled from the fields-of-view analyzed, except for panels F and G which show the original fields-of-view. For statistics see Supplementary Tables S1–S5.

Downloaded from https://academic.oup.com/nar/article/46/20/10649/5078800 by guest on 26 October 2021

cells (Figure 1B), suggesting that endogenous nicks can occur independently of DNA replication. The free 3′OHs revealed by labeling could be either (i) part of preexisiting ds fragment ends, or (ii) belong to preexisiting single-strand (ss) discontinuities, likely nicks, which could serve as predilection points for mechanical breakage of the intact strand during solubilization of the plugs and/or combing. In a model experiment, the average size of combed λ phage DNA nicked at specific sites by a nickase enzyme (see Materials and Methods) was reduced relative to the length of the phage DNA molecules without nickase treatment (Figure 1E–G). Therefore, it is likely that the preferential end-labeling of the ds fragments obtained upon combing was the consequence of mechanical breakage upon solubilization or combing, at pre-existing, labeled nicks, even though no pipetting or mechanical shearing were applied in the whole protocol. The presence of nicks in the agarose-embedded DNA was demonstrated by standard nick-translation experiments (in the absence of terminator nucleotides) where labeled DNA stretches encompassing a few kb were visualized (Figure 1C). Such long stretches could become labeled by the 5′-3′ polymerase activity of the DNA polymerase I (involving either nick-translation or strand-displacement). Incorporation of the biotinylated nucleotide in the case of ds breaks with overhanging 5′ end would result in spot-like end-labeling due to the limited length of the overhang at a ds break (in contrast with two, non-apposed ss breaks). (See Supplementary Tables S1-S5 for statistical analysis of the combing experiments.) In order to exclude the possibility that the labeling originated from free 3′OH of RNA (see (59)) rather than DNA molecules, terminal transferase (TdT) treatment was followed by combined RNase digestion applying subsequently RNase A, RNase HI and human RNase H2 treatments (based on (60–62)). Note that this schedule should preclude labeling of RNase-induced nicks at misincorporated ribonucleotides. RNase H enzymes specifically degrade the RNA portion of the RNA:DNA hybrids, so the RNA-primed fragments should diffuse away from the sites of labeling. Importantly, biotinylated nucleotides incorporated by TdT were not removed by RNase treatment (Supplementary Figure S1A and B). The efficiency of RNase treatment was indicated by the fragmentation of the DNA likely due to unrepaired, misincorporated ribonucleotides in the genome (63–65) and by the disappearance of S9.6 staining (21,37) of the R-loops (Supplementary Figure S1C and D) that also appeared frequently at the ends of the combed fragments (Supplementary Table S2). Remarkably, coinciding R-loop and nick signals were also detected upon co-labeling (Figure 1D), suggesting that these two features are occasionally positioned within ∼1 kb from each-other (in view of the resolution of the microscope). In summary, the DNA breaks giving rise to loop-size fragments in combing experiments may represent endogenous nicks with free 3′OH ends.

In order to determine the overall incidence of nicks along the genome and assess their relative localization on the two strands, we performed gel electrophoretic analyses by FIGE (Supplementary Figures S2 and S3, for *S. cerevisiae* and *S. pombe*, respectively). Treatment of the DNA with S1 nuclease revealed S1-sensitive regions delimiting ∼100 kb intervals on the ds DNA (Supplementary Figure S2A). These may include both endogenous nicks and alternative DNA structures containing ss regions (like hairpins, R-loops, G-quadruplexes, etc.). However, the ss fragments observed after urea/heat (and also alkaline) denaturation offer a *bona fide* indication of DNA breaks (see Supplementary Figure S2B). Note that urea/heat-denaturation ((57); explained in Supplementary Figure S4) is a method which avoids the hydrolysis at misincorporated ribonucleotides that may happen in alkaline gels (64,66). The ss size of the DNA denatured and analyzed after S1 digestion was smaller than without digestion, suggesting that mainly nicks arranged on the two strands in a non-apposed manner are the sites of S1 digestion (compare Supplementary Figure S2B lanes 4–7 with the schemes of Supplementary Figure S2C and D). The size of the urea/heat-denatured ss fragments was very similar to that of denatured T4 phage DNA (∼169 kb; data not shown).

Supplementary Figure S5 shows that the average ss fragment size assessed by S1 digestion or urea/heat-denaturation was not affected by the addition of ethidium bromide (EBr) that relaxes or overwinds DNA depending on its concentration (37,67,68). If nicks on the opposite strands were separated by short (< 25 bp) stretches of DNA, disassembly into loop-size ss fragments upon denaturation would readily occur at ≤80°C (see the Supplementary Figure S5 legend for the calculation), in sharp contrast to what was observed. In line with the conclusion that nicks far away from each other on the complementary strands are responsible for the appearance of the ∼150 kb ss fragments upon denaturation, the 2 μ plasmid that carried no nicks (Supplementary Figure S2B, lanes 2–3) was denatured at a very similar EBr concentration as gDNA (Supplementary Figure S5B).

Collectively, the above data confirm that the nicks on the complementary strands are arranged in a non-apposed manner, i.e. they represent nicks rather than ds breaks, in agreement with the results of the combing experiments. They also suggest that the emergence of the breaks is not related to an abrupt redistribution of twist and writhe along the loops, which is likely to occur upon lysis/deproteinization (69).

Since shearing-related mechanistic origins of a post-lysis generation of the breakages were excluded in our earlier reports (33,36), the discontinuities observed here appear to be pre-exisisting, endogenous breaks, generated *in vivo*. The incidence of labeled sites in the combing experiments was around one in 70–100 kb DNA, so it is plausible to conclude that most of the breaks detected in the gel electrophoretic experiments (i.e. ∼1 in 100 kb) possess free 3′OH ends.

Genome-wide distribution of endogenous nicks: microarray analyses suggest RNAP II promoter-proximal enrichment

To find out if the nicks are scattered randomly or accumulate at particular sites over the *S. cerevisiae* genome, we performed microarray experiments (Figure 2), using normally cycling, non-synchronized cells. As most of the nicks carry free 3′OH ends (see above), nick-translation could be used to label these breaks. After deproteinization of the agarose-embedded and formaldehyde-fixed spheroplasts, these sites were nick-labeled with biotinylated nu-

Downloaded from https://academic.oup.com/nar/article/46/20/10649/5078800 by guest on 26 October 2021

dc_1977_21

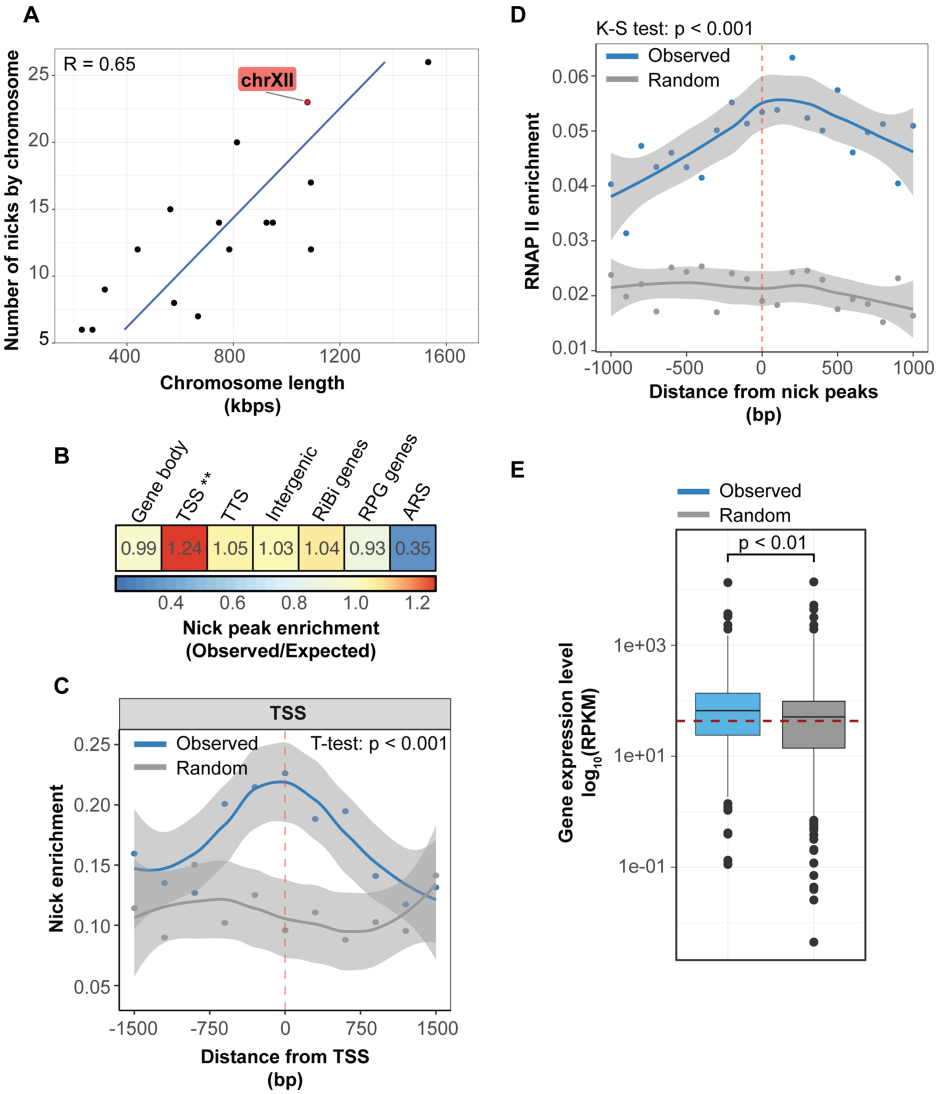


Figure 2. Distribution of nicks with free 3′OH ends along the gDNA of *S. cerevisiae.* (**A**) The frequency of nicks positively correlates with chromosome length. The number of nicks determined in one of the microarray experiments is shown for the *S. cerevisiae* chromosomes. R: Pearson correlation coefficient. Dots indicate the number of nicks per chromosome, chr XII is highlighted. (**B**) Distribution of nicks (peaks) within the functional categories of the *S. cerevisiae* genome (TSS: transcription start sites; TTS: transcription termination sites; RiBi: ribosome biogenesis genes; RPG: ribosomal protein genes; ARS: autonomous replicating sequences). Heatmap shows the ratio of overlapping nucleotides between observed and randomly permutated nick positions within the indicated genomic category. Significant difference was assigned with the two-tailed proportion test (**$P < 0.001$). (**C**) Nick signals are preferentially enriched at transcription start sites (TSSs; n (total gene pool) = 6664). The average binding intensity of Nick ChIP-chip signal is shown as a scatterplot for observed (blue) and randomly permutated (gray) nick peaks (±1500 base pairs). Lines represent the smoothed mean after loess normalization. Faded area shows the 95% confidence interval for predictions from the loess model. Significant difference was assigned using the two-tailed Student's t-test. (**D**) RNAP II tends to accumulate over nick peaks in *S. cerevisiae*. The average binding intensity of RNAP II is shown for observed (blue) and randomly permutated (gray) nick peaks (+/- 1000 base pairs). Lines represent the smoothed means after loess normalization. Faded area shows the 95% confidence interval for predictions from the loess model. Significant difference was assigned using the Kolmogorov–Smirnov (K–S) test (two-tailed). (**E**) Nick-associated genes show elevated mRNA expression. Proportional box plots show the expression levels (RPKM) of nick-associated genes (blue; n = 215) and of random genes (gray; n = 215). Red dashed line represents the genome-wide median mRNA levels. Significant difference was assigned using the Mann–Whitney rank sum test.

cleotides under limiting conditions (see Supplementary Figure S6 for optimization of these conditions). Subsequently, the sonicated DNA samples were immunoprecipitated, amplified and hybridized onto tiling microarrays covering the non-repetitive regions (i.e. excluding rDNA, telomeric regions, tRNA genes and retrotransposons) of the *S. cerevisiae* genome. As Figure 2A shows, the overall frequency of nick overlapping sequences detected by the array was proportional to the length of the chromosomes. The incidence of the endogenous nicks was ∼1/70 kb (calculation based on Figure 2A). Remarkably, the nicks accumulate at the RNAP II transcription start sites (TSSs; Figure 2B, C and Supplementary Figure S7A) as opposed to the transcription termination sites (TTSs; Supplementary Figure S7B), the autonomously replicating sequences (ARSs), and other genetic regions (Figure 2B). RNAP II distribution (data from

Downloaded from https://academic.oup.com/nar/article/46/20/10649/5078800 by guest on 26 October 2021

(54)) peaks around nicks (Figure 2D) and the incidence of nicks significantly correlates with gene activity (Figure 2E; and Supplementary Figure S7A). To assess the sensitivity and accuracy of the whole procedure, some of the samples were also digested with a frequent cutter nickase enzyme of known specificity (Supplementary Figure S8A–C; see Materials and Methods). Annotation of the nicked promoters in different experimental samples shows partially overlapping sets (Supplementary Figure S8D). R-loops, that can be detected by the S9.6 antibody, were absent from the TSSs and accumulated at the TTSs (unpublished data in line with (70,71)), so most probably we deteced true nicks at the TSSs. Similar, TSS-proximal accumulation of nicks was observed using wild-type cells at different metabolic conditions and alpha factor synchronized (G1) *bar1Δ* cells (Supplementary Figure S8D and data not shown). The results of the microarray analyses were confirmed in an independent 'chip-on-beads' experiment according to (72,73): We demonstrated enrichment of Ser5-phosphorylated (i.e. initiating and elongating; see (74)) RNAP II in the <500 bp vicinity of nicks, whereas no such an accumulation of the Ser2-phosphorylated, elongating and terminating species was observed (Supplementary Figure S9). The majority of initiating RNAP II molecules detected in the assay are considered primarily gDNA-bound.

Collectively, these data suggest that endogenous nicks are distributed along the chromosomes in a non-random manner, accumulating at RNAP II promoters, strongly suggesting that they are transcription- rather than replication-related.

Distribution of endogenous nicks within chromosome XII of *S. cerevisiae* defines ∼100 kb loops

In agreement with the microarray data, the overall manifestation of endogenous nicks detected by two dimensional gel ectrophoresis (described in Supplementary Figure S10) was similar for the 16 *S. cerevisiae* chromosomes (Figure 3A). Moreover, the nick distribution within the three *S. pombe* chromosomes, as revealed either by S1 digestion or urea/heat-denaturation, was also similar (Supplementary Figure S11), indicating that the nick-generating mechanisms may be closely related in these two genetically distant model organisms. Interestingly, chr XII (carrying the rDNA cluster) also exhibited a similar fragmentation pattern, suggesting that sequence-unrelated factors determine the incidence of ss breaks. Indeed, the DNA fragments detected on Southern blots by an rDNA specific probe showed a size distribution similar to that of the total gDNA (Figure 3B and C). Sizing was performed using CHEF, yielding a more accurate picture of fragment length distribution than FIGE (used in (25,26,33,36–38)). The ds fragment size range was between 20 and 200 kb, with the median of distribution at ∼100 kb. Thus, every ∼10th of rDNA units of identical sequence harbors nicks on any of the DNA strands. Collectively, these data suggest that sequence-related factors alone do not determine where the nicks are generated.

Endogenous nicks could also be detected in chr XII by molecular combing, as shown in Figure 3D–H. The nicks, revealed by enzymatic labeling of the free 3′OHs, were present in the DNA of G1-synchronized cells, excluding DNA replication as a main source of nick formation. In agreement with the results of whole genome combing experiments (Figure 1C), the stretches of labeled DNA observed in standard, non-limiting nick-translation conditions indicated that ss discontinuities were detected (Figure 3E and G). The quantitative evaluation of the labeling frequencies in the case of gDNA and chr XII is presented in Tables S1, S3 and S5. Remarkably, very few combed chr XII fragments carried label on both ends (Supplementary Table S5). Molecules carrying several labeled sites separated from each-other at ∼10 kb distances were also observed in the case of limiting nick-translation (arrows in Figure 3F). Nicks appear to be less frequent in chr XII than in the whole gDNA, perhaps because smaller fragments bordered by neighboring nicks within a single rDNA unit (see Figure 6) have been lost during combing. We also found that DNA molecules derived from G1 synchronized cells contained nick-labeled spots less frequently than those of asynchronous/non-perturbed cells (Supplementary Table S3), as a result of the absence of Okazaki fragments in G1. Surprisingly, S9.6 labeling was also prevalent at the fragment ends of chr XII (Figure 3H). The identity of R-loops was further confirmed by their co-localization with recombinant, bacterial single-strand binding protein (SSBP, (75); Figure 3H).

Mapping of the nicks within the rDNA units of chromosome XII demonstrates accumulation at discrete RNAP II-related sites

To map the arrangement of nicks within the rDNA cluster, we employed restriction endonucleases that cut only once in each rDNA unit (Figure 4A) followed by Southern blotting. The cleavages in the other parts of the gDNA produced high molecular weight fragments that remained in the compression zone of the gels (e.g. Figure 4B lane 1). The nicked sites were determined after S1 nuclease digestion or urea/heat-denaturation, when the rDNA units were cut out by Sfi I (Figure 3B), Sma I (Figure 4 and Supplementary Figure S12), and Mlu I or Hind III (Supplementary Figure S12). The nicks present in the rDNA units gave rise to discrete bands (Figure 4C and D, lanes 2 and 3; summarized in Figure 4E). Many restriction endonucleases may artefactually nick the DNA at sequences differing from the recognition site with only one base pair (19,20). To distinguish the genuine endogenous nicks from off-target nicking by restriction enzymes, the S1 nuclease and the restriction enzymes were applied in alternating sequence (for an illustration of the rationale see Supplementary Figure S13). Based on the comparison of the two cleavage patterns (exemplified by lanes 2 and 3 of Figure 4C and D), off-target nicks (labeled by an asterisk in Figure 4 and Supplementary Figure S12) created by restriction endonucleases were eliminated from the maps of endogenous nicks. This problem was avoided when Sfi I was used, since this enzyme generated no artefactual nicks within the rDNA units (see Figure 3B-C and compare lanes 2 and 3 in Supplementary Figure S14). The arrangement of nicks in each strand of the rDNA was also investigated by using strand-specific ss probes (Figure 4 and Supplementary Figures S12 and S14) following urea/heat-denaturation of rDNA units into dif-

Downloaded from https://academic.oup.com/nar/article/46/20/10649/5078800 by guest on 26 October 2021

dc_1977_21

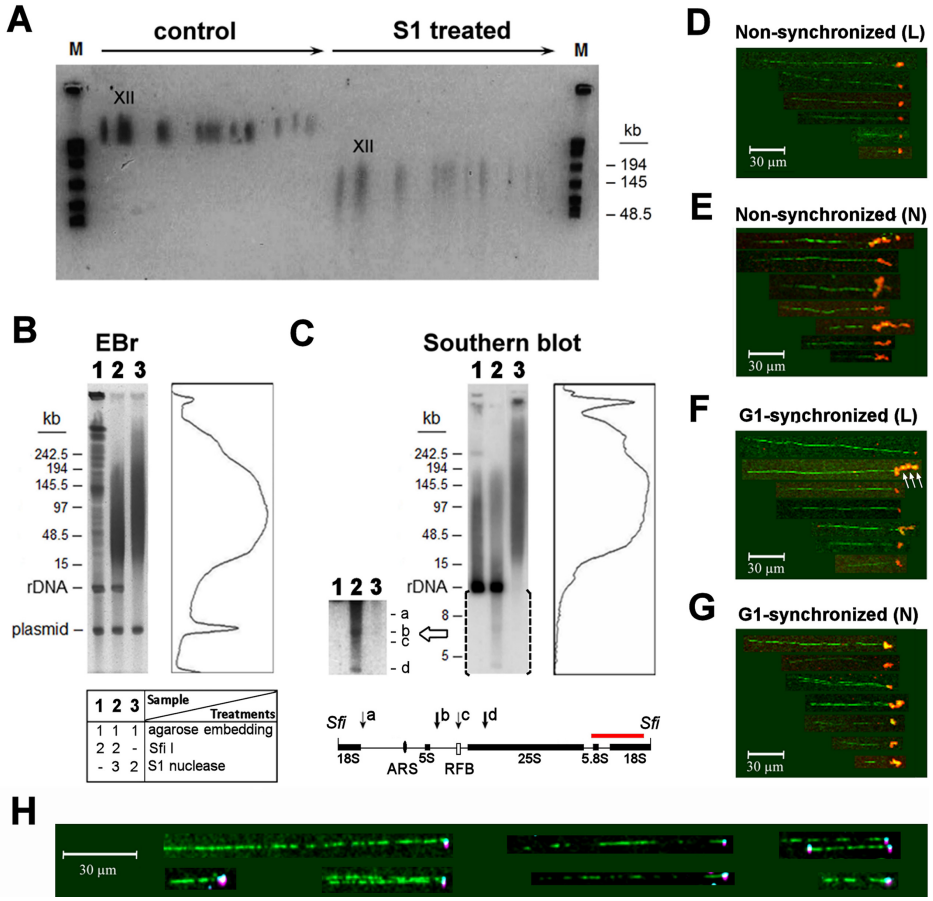


Figure 3. Endogenous nicks delimit loop-size intervals in chromosome XII of *S. cerevisiae.* (**A**) Two-dimensional gel electrophoretic analysis of S1 sensitive sites in *S. cerevisiae* chromosomes. Deproteinized chromosomal DNA molecules resolved by CHEF (first dimension) were digested with S1 nuclease (or treated with S1 buffer as a control), then further separated in the second dimension by FIGE. The arrows point from the large to the small chromosomes. The order of chromosomes from left to right is IV, XII (marked), VII/XV, XIII/XVI, II, XIV, X, XI, V/VIII, IX, III, VI, I. (B, C) CHEF and Southern blot analysis of rDNA specific S1 sensitive sites in the deproteinized DNA of agarose embedded WDHY199 spheroplasts. (**B**) EBr-stained CHEF gel after electrophoresis of the Sfi I and/or S1 digested samples in a representative experiment using the same strain. Lanes: 1. Sfi I digestion; 2. Sfi I and S1 digestion; 3. S1 digestion. The densitometric size distribution of gDNA fragments after S1 nuclease digestion (in lane 3) is shown right to the gel. The table underneath the agarose gel image shows the order of treatments. (**C**) Southern hybridization using an rDNA specific probe. The bracketed section left to the blot shows an enhanced exposure. The sketch below the blot shows positions of rDNA fragments a-d on the rDNA map. These fragments were derived by S1 digestion of the Sfi I digest (panel C, lane 2), as detected by the probe (red). The size distribution of rDNA after S1 nuclease digestion (panel C, lane 3) is shown right to the blot. The rDNA fragments appeared as a 9.1 kb band upon Sfi I digestion that cuts once in every unit (panels 3B and C, lanes 1–2). The 2-micron plasmid, a small multi-copy selfish DNA element of the strain ((134); labeled as 'plasmid') served as an internal negative control confirming that the hybridization signals were rDNA specific. (D–G) Molecular combing of nick-translated chr XII. Biotinylated nucleotides were incorporated by limiting (L), or standard/non-limiting (N) nick-translation into the gDNA of agarose embedded non-synchronized (panels D and E), or G1-synchronized (panels F and G) BY4741 cells, before separation of the chromosomes by CHEF. (**D**) chr XII from non-synchronized cells after limiting nick-translation; (**E**) chr XII from non-synchronized cells, after non-limiting/standard nick-translation; (**F**) chr XII from G1-synchronized cells, after limiting nick-translation (white arrows point to several juxtaposed labeled rDNA units); (**G**) chr XII from G1-synchronized cells after non-limiting, standard nick-translation. Biotin label was visualised using AlexaFluor 647-conjugated anti-biotin antibody (red). Note that the stretches of labeled DNA are folded to form large, bright spots. (**H**) Labeling of R-loops on combed *S. cerevisiae* chr XII DNA. The RNA:DNA-hybrids were detected by the S9.6 antibody (pink), and the displaced, ss DNA was stained using Cy3B-conjugated SSBP (cyan; prepared and labeled as described previously (75)). DNA molecules were stained with YOYO-1 (green). Images were assembled from the fields-of-view analyzed. For statistics see Supplementary Tables S1, S3 and S5.

ferentially migrating complementary strands (as in (57,76)). Nicks that were mapped by S1 nuclease (i.e. when restriction enzyme digestion was preceded by S1 treatment), and those that were also observed in the urea/heat-denaturing electrophoretic experiments, were considered genuine, preformed nicks, not caused by restriction enzyme digestion, or by locally denatured regions of the ds DNA. The positions of endogenous breaks are represented on the rDNA maps of Figures 4E and 6, and in Supplementary Figures S12 and S14. Some nicks were detected both on the sense and antisense strands by single strand specific probes, although ds breaks were not detected within the rDNA units without S1 digestion (Figure 4C and D, fragment 'f'; compare lane 1 with lanes 4 and 5). Such nicks may have occurred either in one or in the other strand in the different rDNA units, or, alternatively, arranged within the affected units close enough to be mapped to the same position using the ss probes, but distant enough so that the two strands can be kept together.

Taken together, these data show that endogenous nicks accumulate at discrete sites along the rDNA units, except

Downloaded from https://academic.oup.com/nar/article/46/20/10649/5078800 by guest on 26 October 2021

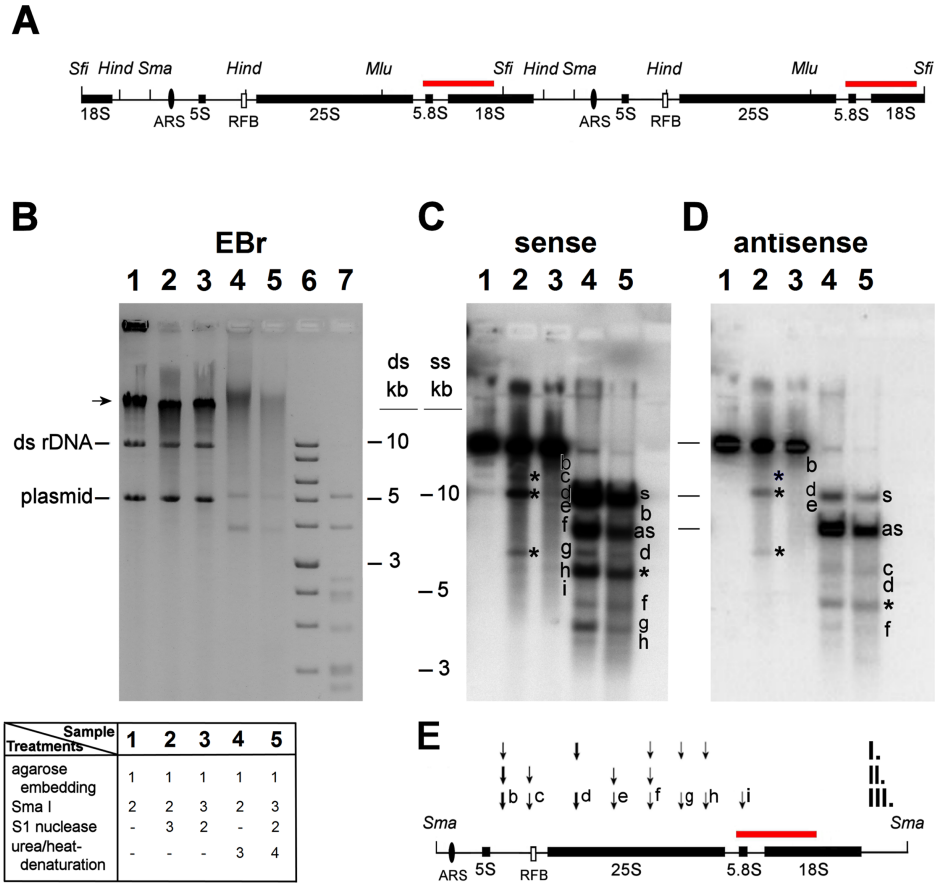


Figure 4. Nicks accumulate at discrete sites within the rDNA units. Deproteinized WDHY199 spheroplasts were analyzed by Southern blotting using strand specific rDNA probes. (**A**) Restriction map of two contiguous rDNA units. The black boxes indicate the 25S, 18S, 5.8S and 5S exons. ARS and RFB are illustrated by a black oval and a white rectangle, respectively. The hybridization position of the Southern probes is indicated by a red bar. Sfi: Sfi I; Hind: Hind III; Sma: Sma I; Mlu: Mlu I. Note that Sfi I, Sma I and Mlu I are rare cutting restriction endonucleases. (**B**) EBr stained gel containing gDNA after: Sma I cleavage (lane 1); Sma I cleavage followed by S1 nuclease digestion (2); S1 nuclease digestion followed by Sma I cleavage (3); Sma I cleavage followed by urea/heat-denaturation (4); S1 nuclease digestion, then Sma I cleavage, then urea/heat-denaturation (5). The order of treatments is summarized by numbers in the table underneath the agarose gel image. Lanes 6 and 7 contain non-denatured and denatured 1 kb marker, respectively. Arrow: compression zone; plasmid: 2-micron plasmid (134); ds rDNA: 9.1 kb non-denatured rDNA units. (C and D) Southern hybridizations with sense (**C**), and antisense (**D**) strand-specific probes. 's' and 'as' indicate the separately migrating 9.1 kb sense and antisense strands of denatured rDNA units, respectively. Note that the difference in electrophoretic migration between the two strands of same size is due to their different base composition which leads to different overall conformational characteristics even in the presence of the denaturing agent (see (57)). The signal intensities were consistently stronger in the case of the electrophoresed ss than ds DNA, possibly because hybridization of the ss probe to ss target sequences is more efficient than to ds target sequences. 'b-i': fragments derived from endogenous nicks revealed by S1 nuclease digestion or urea/heat-treatment; *: artefactual fragments generated by the nicking activity of Sma I. (See Supplementary Figure S13 for the consequences of 'off-target' nicking.) (**E**) The scheme summarizes the location of nicks at one rDNA unit. The arrows indicate the positions of nicks ('b-i') mapped in the rDNA. I.: nicks detected with the sense strand-specific probe in denatured, ss rDNA samples (lanes 4 and 5); II.: nicks detected with the antisense strand-specific probe in denatured, ss rDNA samples (lanes 4 and 5); III.: nicks converted to ds breaks with S1 nuclease, detected with either sense or antisense strand-specific probes in non-denatured, ds rDNA samples (lane 3). One representative blot of at least three independent experiments is shown.

for a ~2 kb region encompassing the ARS (see Figure 6). In line with the reduced incidence of nicks in the ARS-containing region, the position of the ss breaks was indistinguishable when α-factor-, or nocodazole-synchronized, G1 and G2/M phase arrested cells, respectively, were studied (Supplementary Figure S14), suggesting that the endogenous nicks occur independently from, and persist through DNA replication.

The arrangement of nicks was insensitive to RNase HI or RNase A treatment of the plugs (Supplementary Figure S12 and data not shown), indicating that cleavages by S1 nuclease at the ss DNA of the R-loops in rDNA (77,78) do not contribute to the cleavage patterns observed in Southern hybridisations. In support of this conclusion, the localization of R-loops and nicks along the rDNA unit diverged from each-other in IGS1 and IGS2 regions (see below).

In order to enhance sensitivity and also to distinguish the nicks revealed by S1 digestion or urea/heat-denaturation from the constitutive, Fob1-dependent nicks carrying Top1-blocked 3′ termini at the RFB (17), we mapped endogenous ss nicks in the rDNA units by incorporating biotinylated nucleotides into free 3′OHs of gel-embedded samples by nick-translation followed by rSW blotting ((58); see flow-chart in Supplementary Figure S15). The localization of the nicks was made possible by using limiting nick-translation conditions (see the optimization in Supplementary Figure S6).

Downloaded from https://academic.oup.com/nar/article/46/20/10649/5078800 by guest on 26 October 2021

dc_1977_21

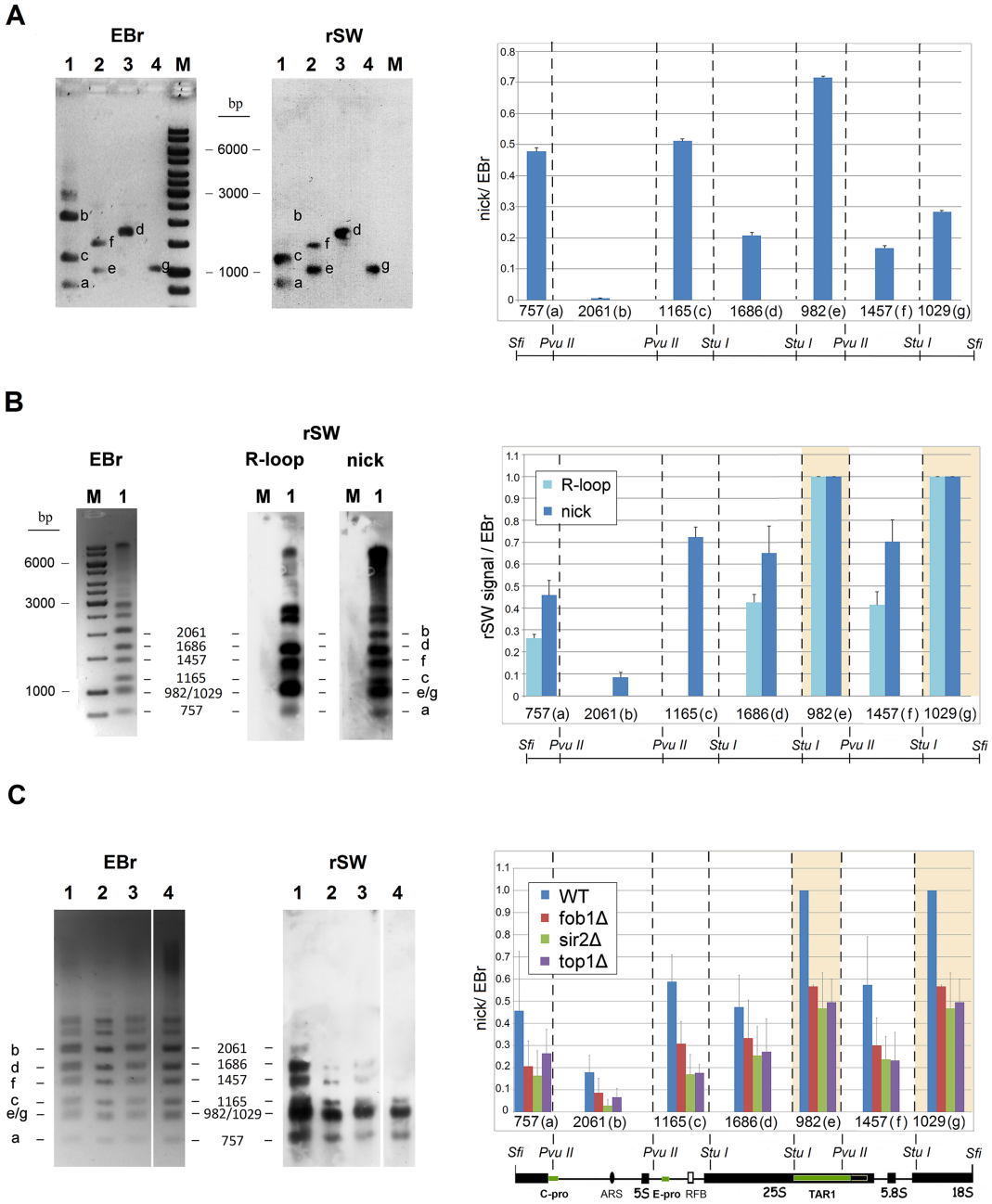


Figure 5. Nicks are unevenly distributed within the rDNA units, do not coincide with R-loops at the intergenic spacers and their incidence correlates well with rDNA silencing. (**A**) Reverse Southwestern blot (rSW) analyses of free 3′OHs in *S. cerevisiae* (WDHY199) rDNA units (see also Supplementary Figure S15). Agarose embedded, deproteinized gDNA was nick-translated in limiting conditions, digested with Sfi I and the fragments were separated by gel electrophoresis. The 9.1 kb rDNA units were cut out from the agarose gels and further digested with Stu I. After the next round of gel electrophoresis, the Stu I fragments (lane 1: 3983 bp; lane 2: 2419 bp; lane 3: 1686 bp; lane 4: 1029 bp) were cut out and further digested with Pvu II. M: 1 kb marker. Left sub-panel: EBr stained gel. Right: rSW anti-biotin staining of the blot. a-g: Pvu II generated bands of Stu I fragments. Histogram: fragment size (bp) and map position of the fragments are represented on the X-axis and on the map below, respectively, and labeling intensities (nick/EBr ratio: Biotin signal intensities of the fragments normalized to their corresponding EBr signal intensities) are shown on the Y-axis. (**B**) rSW analyses of free 3′ OHs and R-loops in *S. cerevisiae* (BY4741) rDNA units. Agarose-embedded and deproteinized gDNA samples were analyzed as in A. Immunostaining of blot was performed sequentially with the S9.6 antibody (sub-panel rSW, R-loop), then with an anti-biotin antibody (sub-panel rSW, nick). Lane 1, wild-type cells; M: 1 kb marker. EBr: EBr-stained gel. a–g: Pvu II and Stu I double-digest fragments of the rDNA units. For the description of X- and Y-axis see (A). S9.6 and biotin signal intensities of the fragments were each normalized relative to their corresponding EBr signal intensities. The band with the highest biotin/EBr or S9.6/EBr signal ratio on the gel was set to 1 on the relative scale. Note that the 982 bp and 1029 bp fragments migrated very closely together on the gel, so they were labeled as e/g and their cumulative intensities (marked by shaded areas) were determined. (**C**) rSW analyses of free 3′OHs in the rDNA of the *S. cerevisiae* WT (BY4741) and isogenic mutants *fob1*Δ, *top1*Δ, and *sir2*Δ. The agarose embedded, deproteinized DNA samples were analysed as in A. The size and position of restriction fragments (a–g) are labeled. Lane number, corresponding coloured bars in the histogram, and corresponding strain: 1, blue bars, WT; 2, red bars, *fob1*Δ; 3, green bars, *sir2*Δ; 4, purple bars, *top1*Δ. All the lanes shown are from the same gel. The X- and Y-axis are as in (A). For description of fragments e/g see (B). Bar charts in (A)–(C) represent mean values ± SEM obtained from three independent experiments. Positions of the fragments in (A)–(C) are mapped over the SfiI-rDNA unit (for a more precise mapping of fragment positions see Figure 6).

Downloaded from https://academic.oup.com/nar/article/46/20/10649/5078800 by guest on 26 October 2021

Note also that the method has a wide dynamic range and is highly sensitive (see Supplementary Figure S16). The distribution of nicks with free 3′OH ends along the rDNA units was not uniform (Figure 5A). Remarkably, biotin-dUTP incorporation was barely detected in the fragments carrying the ARS sequences of rDNA. This observation is consistent with the results of the Southern hybridization experiments (Figure 4) and excludes rDNA replication firing from the primary causes of nick formation. Moreover, we noticed that three of the rDNA fragments that showed the highest incidence of nicks (Figure 5A, fragments 'a': 757 bp, 'c': 1165 bp, 'e': 982 bp) contain RNAP II promoter sequences (i.e. *C-PRO*, *E-PRO* and the promoter of *TAR1* gene, respectively (79–83)). Detection of R-loops with S9.6 staining on the same blots (Figure 5B) revealed strong coincidence with the nicks in several segments of the rDNA units. However, R-loops were completely absent from fragments 'b' (2061 bp) and 'c' (1165 bp), in line with (78). Thus, the high level of nick labeling in the 1165 bp fragment cannot be attributed to free 3′OHs of RNA origin, i.e. the elongation of R-loops by the DNA polymerase used for nick-translation could not contribute to the nick signals. This conclusion is in line with the results of the combing experiments shown in Supplementary Figure S1.

Finally, to further investigate if nick formation could be modulated by mutations affecting transcription, we compared the incidence of nicks along the rDNA units in strains with constitutively defective rDNA silencing (mutants *top1Δ*, *sir2Δ* and *fob1Δ*). Importantly, malfunctioning of rDNA silencing (an rDNA specific phenomenon central to the cooperation of RNAP II and I (83,84)) led to a decrease of nicking in the rDNA units in all of the three mutant strains (Figure 5C).

DISCUSSION

Here we report that endogenous, non-random single-strand discontinuities/nicks are scattered along the *S. cerevisiae* genome. The nicks are distributed at ≤ 200 kb distance apart from each other on each strand, so as to yield ∼70–100 kb ds fragments upon further breakage at these predilection points. These findings extend our earlier reports on loop-size DNA fragmentation (25,36,37,85) as: (i) The nick-character of the breaks was corroborated by their independent distribution on the two DNA strands and in molecular combing experiments. (ii) By excluding superhelical tension as a factor eliciting the breaks, what further confirmed their endogenous nature. (iii) By determining the distribution of the nicks along both the whole gDNA and the rDNA in WT and *S. cerevisiae* mutants. The loop-size periodicity was recapitulated in the rDNA cluster, implicating sequence-unrelated factors in their generation/maintenance. The coincidence of the nicks with strategic sites of transcription regulation along the whole genome, including the rDNA cluster, is in line with the hypothesis that these nicks may be related to the promoter-proximal breaks of mammalian models (reviewed in ref. (39)). The analogy with the latter observations is supported by the association of nick incidence with active promoters, which is contrasted with the lack of nick accumulation in ARS regions. The demonstrated potential of gDNA for ds breakage at predilection points defined by nicks raises questions concerning the generally assumed ds character of the breaks detected in mammalian cells. The co-occurrence of nicks and R-loops in a fraction of gDNA and in certain regions but not in others of the rDNA units is pertinent to the role of these structures in genomic instability.

Evidence for the endogenous character of the nicks

Several possible mechanisms for the artefactual generation of strand discontinuities in our experimental circumstances have been excluded in earlier publications (25,26,33,36,85), and the role of superhelical tension has been ruled out in this paper (Supplementary Figure S5). Our microarray experiments that were conducted using fixed cells also support this notion. Further strong arguments are provided by the non-random localization of nicks found by several methods in this study. Mapping of the nicks within the rDNA units by Southern blot analyses of denatured chr XII DNA samples revealed several strand-specific breaks, at discrete sites. On the other hand, the IGS2-ARS-containing 2 kb region was poor in breaks (Figures 4-6). The relative lack of breaks in the ARS regions (as shown in gDNA by microarray experiments (Figure 2), and within the rDNA cluster by Southern hybridization and rSW (Figures 4 and 5, respectively) provides compelling evidence that some genomic regions are spared of nicking. rSW is a highly sensitive method that detects nicks carrying 3′OH in a particular region, so all the breaks within the restriction fragment are detected regardless of their exact position. Since this method was applied to detect nicks in a naturally amplified locus, the nearly complete absence of nicks from IGS2 in some experiments (e.g. Figure 5A) suggests that the level of random nicking in our setup is usually very low. The facts that the epigenetic state of both IGS1 and IGS2 is heterochromatic (86,87) and IGS1 harbors much more nicks than IGS2, argue that the site-specific nicks reported here cannot be attributed to nonspecific nucleases acting in open chromatin regions. The low level of nicks in the rARS region in spite of the fact that one fifth of the rDNA units are engaged in replication during a single cell cycle also suggests that accessibility is not the decisive factor in the generation of the nicks. The fact that the incidence of nicks is linked to transcriptional processes (see Figures 2E, 5C and Supplementary Figure S9) supports the notion that the nicks are of physiological function and origin. The *S. pombe* genome, distant from that of the bakers' yeast in evolution, also harbours endogenous nicks at loop-size intervals, in line with our hypothesis that the promoter-proximal, gene activity-dependent DNA breaks observed in mammalian cells and those described herein are related.

Collectively, our data suggest that the nicks detected in our assays exist before cell lysis, i.e. they are present in the gDNA, including the rDNA, of live cells, and become manifest as ds breaks only after lysis under conditions that allow unfolding and diffusive motility of chromosomal DNA.

The incidence of nicks is connected with transcriptional processes

Accumulation of the nick signal at RNAP II promoters and the finding that active gDNA promoters preferentially har-

Downloaded from https://academic.oup.com/nar/article/46/20/10649/5078800 by guest on 26 October 2021

dc_1977_21

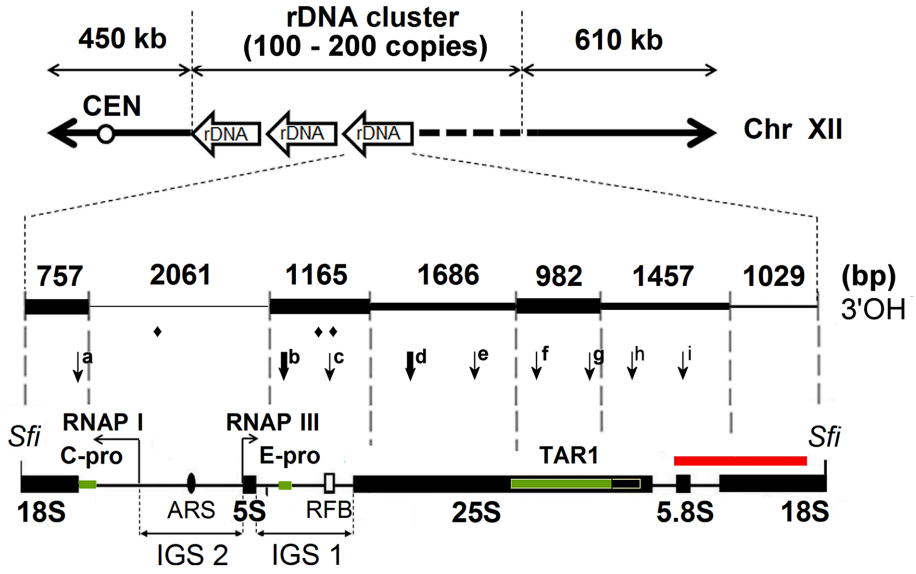


Figure 6. Overview of endogenous strand-breaks mapped within the rDNA unit. Black boxes indicate the 25S, 18S, 5.8S and 5S exons of the 9.1 kb rDNA unit. Black arrows show transcription from the RNAP I and III promoters. ARS and RFB are illustrated by a black oval and a white rectangle, respectively. CEN: centromere. Sfi: SfiI cleavage sites. The green bars show RNAP II promoters. Horizontal arrows show the direction of transcription from the RNAP I and III promoters. Arrows directed downward show the ss breaks detected in this work. Thick and thin arrows indicate the positions of strongly and weakly discernible breaks, respectively. The intervals of different thickness above the arrows indicate the 3′OH labeling intensity observed within the different restriction fragments in the rSW experiments (see Figure 5). Breaks labeled with rhombus (♦) were reported in (17,104,135). The position of the (strand-specific) Southern probe is indicated by a red horizontal bar. The numbers above the map show the size of the corresponding restriction fragments.

bor nicks (Figures 2, 6 and Supplementary Figure S7A) imply that the ss breaks appear in the context of a RNAP II-related transcriptional activity. According to Supplementary Figure S9, they are associated with initiation rather than elongation.

In the rDNA cluster, some of the DNA breaks that mapped at the locus coincide with the RNAP II promoters embedded in the rDNA units (see Figure 6). The finding that nicks with free 3′OH ends co-localize with RNAP II promoters also within the rDNA is surprising, because nucleoplasmic gDNA gene transcription and the RNAP II related nucleolar transcriptional processes are fundamentally different. Whereas the RNAP II controlled transcriptional processes at bulk gDNA lead mainly to the synthesis and concomitant processing of protein coding primary mRNA transcripts, RNAP II appears to transcribe mostly noncoding RNA in the nucleolus (except for *TAR1*; see below). The RNAP II promoters in the rDNA are *E-PRO, C-PRO* and *TAR1* (see Figure 6). *E-PRO*, a bidirectional promoter, resides in the rDNA units close to the RFB and has a role in the regulation of rDNA amplification via its noncoding RNA products (54,79). Another bidirectional RNAP II promoter, *C-PRO* (a member of the Cryptic Unstable Transcripts; CUTs), lies in the IGS2 region and its biological role is presently unknown. CUTs represent 200–500 b noncoding RNAs that are rapidly degraded in an Nrd1-dependent pathway and are the major products of pervasive transcription (88–90). *TAR1* (Transcript Antisense to Ribosomal RNA), is embedded into the 25S exon on the antisense strand, is expressed at very low levels and encodes a protein which is localized in the inner membrane of mitochondria (80,91). The levels of transcripts originating from the IGS promoters is low in WT cells, due to rDNA silencing. In this process, the highly active transcriptional activity of RNAP I at the 35 S rDNA leads to the suppression of RNAP II activity at the IGSs (83,84,92–95). Silencing appears to be the direct consequence of histone deacetylation by Sir2. The Sir2 enzyme-containing protein complex is recruited to Fob1 (in a Top1-dependent manner (17,84)) and also to the Pol I promoter (96).

In view of the fundamental differences between RNAP II functions in rDNA and in the rest of the genome, the role and regulation of nick generation may be very different in the two locations. Indeed, opposite to their prevalence at active promoters in the bulk of gDNA, the nicks accumulate at the sluggish RNAP II promoters in rDNA and avoid the much more active RNAP I and RNAP III promoters. The strikingly similar behaviour of three mutants (*top1Δ, sir2Δ, fob1Δ*) affecting the same pathway of regulation of rDNA silencing (84,96,97) strongly argues for a scenario where nicks play a role in transcriptional regulation also in the rDNA. However, the molecular interpretation of the links between nick incidence and transcriptional regulation awaits a better understanding of the intertwined and still mysterious regulation of RNAP I and RNAP II activities at the rDNA locus (82,90,93,94), involving the antagonistic activity of RNAP I and II (94) on the one hand, and the contribution of RNAP I transcription to the maintenance of RNAP II silencing (96) on the other.

Thus the nicks detected in our assays are associated with the transcriptional regulation involving RNAP II also at the rDNA locus. On the other hand, the nicks appear not to be related to DNA replication, since they occur in G1- and G2/M synchronized cells alike (see Supplementary Figure S14) and do not accumulate in the ARS-containing regions (see Figures 2, 4, 5 and 6).

Downloaded from https://academic.oup.com/nar/article/46/20/10649/5078800 by guest on 26 October 2021

Possible molecular mechanisms of nick generation

Detection of breaks having free 3′OH end formally rules out the role of Top1 in their generation, since this enzymatic reaction would yield covalent enzyme-DNA complexes, i.e. Top1ccs (98,99).

In yeast, Top1 and, often interchangeably, Top2 play important roles both in initiation and elongation of transcription (100–103). In addition to its general role in resolving superhelical stress during transcription that would give rise to random nicks, Top1 is responsible for the generation of two site- and strand-specific, persistent nicks (coinciding with 'c' in Figure 6) at the rRFB (17,104). The presence of these known, Top1- and further, possibly Top3-related nicks that are also localised at the RFB (17,104) (see ♦, ♦♦ in Figure 6), may be mainly responsible for the observation that few combed chr XII molecules were detected carrying label on both ends, as opposed to combed gDNA (see Supplementary Table S5). The nicks detected via their free 3′OH in the rSW experiments must be Top1-unrelated. Indeed, analysis of the three rDNA silencing deficient strains revealed nicks at the same localization, albeit at reduced incidence (see Figure 5C). In view of the 5′-tyrosine linkage in the cleavage complex formed by these enzymes, it is much more likely that Top2 (or perhaps Top3) is responsible for the generation and maintenance of the breaks at the rDNA (detected both by Southern hybridizations and the rSW procedure) in sites 'a-b' and 'd-i' of Figure 6. We note, however, that the involvement of Top1 cannot be ruled out unambiguously, since this enzyme may be released from the Top1cc state by e.g. Tdp1 (Tyrosyl-DNA Phosphodiesterase 1; (105)) which would generate free 3′OHs available for labeling. Free 3′OH ends can also be generated in the course of base excision repair. However, neither the preferential generation of such lesions during RNAP II elongation (106), nor their faster repair at nucleosome depleted regions (107) correlate with our observations. On the other hand, in normal mammalian cells and conditions without exposure to mutagens or any other forms of stress, repair factors appear to function at the promoter-proximal, Top2-induced transient lesions that accompany transcriptional activation (39,108). A scenario where the DNA breaks observed herein could be generated by Top2 would be highly attractive: Top2ccs with a single nicked DNA strand (such as those mapped in human colon cancer cells (9)) could tether the bases of loops in concert with topological relaxation with its concomitant nucleosome destabilizing effect (67). Addressing the possible role of Top2 (or Top3) directly would not be trivial since these enzymes are essential for growth in budding yeast (109,110), have multiple roles in transcriptional regulation (15), and any break made by Top2 before the transfer of a conditional mutant to restrictive conditions is expected to persist even after the enzymes are down-regulated, inhibited, subjected to heat-reversal, or upon inactivation (111–114). The findings that in the bulk of gDNA Top2 contributes to the assembly of the RNAP II preinitiation complex (102,103), whereas the depletion of Top2 appears to unslilence RNAP II related ncRNA transcription in rDNA (77), fit the correlation between nick incidence and RNAP II promoter activity in gDNA on the one hand, and between nicking in rDNA and rDNA silencing on the other.

The nicks may be transient

The ∼70–100 kb average size of the ds fragments bordered by endogenous nicks was measured by different approaches: (i) electrophoretic sizing of large fragments (also by CHEF, to avoid the focusing artefact of FIGE (115)), and (ii) by microscopic observation of the DNA fragments generated via stretching-induced breaking (116) at nicks in the molecular combing experiments. On the other hand, the distribution of 3′OH-labeled nicks mapped along the gDNA on microarrays coincided with the promoters of active protein coding genes in general. Since these comprise the majority of yeast genes (117), the TSSs potentially harboring nicks may be ∼2 kb apart from each-other, assuming that sampling was random. It follows that any particular endogenous nick detected in the microarray experiments along the genome must be present only in a minor subpopulation of cells at the time of fixation and nick-labeling, in line with Supplementary Figure S8D. The size of this subpopulation is determined by the sensitivity of nick detection and by biological reasons related to the mechanism of nick generation.

The promoter-proximal association of Top2 has been documented in (102,103,118). If Top2 was responsible for the generation of the nicks observed in our assays, stochasticity could arise in two ways: (i) the active promoters transiently interact with a limited pool of enzymes in a competitive manner, or (ii) each active promoter is permanently Top2-associated, but the enzyme at any particular promoter spends a small part of its catalytic cycle in the Top2cc state, when the relative duration of this time window would determine the average distance of nicks present in the DNA at any given time point. In summary, we propose that the generation of nicks may be a random process preferentially involving active RNAP II promoters.

Higher-order organization of the chromatin and the distribution of nicks

The generation of nicks occurs only in every ∼10th unit of the rDNA on the average (Figure 3B and C), excluding sequence related factors from the primary determinants of nick generation on the one hand, and demanding topological interpretation on the other. Chromatin loops ranging from 20–300 kb have been described in *S. cerevisiae* (119–126). The differences in the size estimates may be due to the differences of the experimental systems, the varied experimental criteria of loop definition and the dynamic features of chromatin determining the loop subpopulations sampled by the particular method. The distribution of nicks, as revealed by independent methods, is in the middle of this wide range. Hence we asked if higher-order organizational features of the chromatin, a question of special interest in the case of the repetitive rDNA locus, would coincide with nick distribution. In view of the obvious demand for superhelical stress dissipation and means of disentanglement in the case of looped domains (127), their anchorage, be it a transient event, e.g. involving topoisomerases, would stand to reason.

Downloaded from https://academic.oup.com/nar/article/46/20/10649/5078800 by guest on 26 October 2021

dc_1977_21

Only ∼50% fraction of the 100–150 rDNA units is transcriptionally active (83,95), with extensive variations depending on the actual metabolic conditions of the cells (86). The loop-size defined by the constitutive Top1ccs involved in rDNA silencing (17) has not yet been reported. The silenced units are thought to be packaged into regular nucleosomes while the transcriptionally inactive IGS1 and IGS2 regions of active rDNA units exhibit complex features (86). Since the distribution of silenced and active units appears to be random (95), the periodicity of nick accumulation in every 10th unit, based on independent assays (see Figure 3 and Supplementary Table S3), cannot be readily accounted for by the 1:1 dichotomy of active and silenced units. The incidence of active replication origins (1:5) more closely corresponds to the average incidence of nicks, even without considering ARS clustering (48). However, in view of the lack of accumulation of nicks in the ARS regions and their presence at the promoters also in G1 cells (data not shown), this imperfect correlation may not be meaningful. Known manifestations of the higher-order organization of the locus include IGS1-IGS2 looping via interaction between RFB and ARS involving RNAP II (128), and the formation of mutually exclusive promoter-terminator loops by the rDNA genes controlled by the different RNA polymerases (49). Whether these interactions occur within a single unit or involve inter-repeat looping, as pointed out in (127), is an open question.

Thus, the incidence of nicks detected in the rDNA cluster does not seem to exactly coincide with any of the periodicities observed so far. Determination of these relationships will perhaps be made possible by superresolution or electronmicroscopic localization of nicks relative to the other known structural and functional elements of the rDNA units.

Implications for genomic instability

Our electrophoretic and molecular combing data both suggest that endogenous nicks are present on the two strands of gDNA in a non-apposed manner, i.e. in an arrangement what is expected to maintain ds chromosomal continuity when the DNA is kept and handled within agarose plugs. By analogy to observations made in the combing experiments that involve considerable stretching of the DNA molecules (116), endogenous ss breaks likely serve as predilection points for the loop-size ds fragmentation of the gDNA observed upon cell lysis in the absence of embedding, due to shearing forces (e.g. see Supplementary Figure S2A, lane 8). This interpretation is in line with early observations made on shear-induced ds breakages of nicked T5 phage DNA molecules (129). Thus, detection of DSBs in experimental conditions devoid of embedding may in fact reveal endogenous ss breaks.

The mechanical vulnerability of the DNA at nicks demonstrated herein may reflect a similar susceptibility to mechanical breakage in living cells. Furthermore, demonstration of the endogenous character of these breaks entails the presence of highly recombinogenic free ends (including free 3′OH termini) *in vivo*. In view of the possibility that the breaks are caused by Top2, it is worth mentioning that in mammalian cells, promoter proximal Top2β, in addition to being implicated in transcriptional regulation (39), has been incriminated as a crucial player in chromosomal translocations leading to cancer (39,130–132). A further intriguing connection with our findings is represented by recent observations implicating Top2 in the aging of *S. cerevisiae* (109).

Co-localization of nicks and R-loops could be observed both in the bulk of gDNA and in rDNA. According to our rSW experiments, 7 kb out of the 9.1 kb rDNA unit harbored both entities. The fact that the IGS1 and IGS2 regions are devoid of R-loops is in line with the transcriptional silencing of these regions and correlate with earlier S9.6-ChIP-seq data (78). In agreement with the rSW results, R-loops were frequently detected at the combed chr XII ends (see Figure 3H), similarly to the endogenous nicks. In view of the fact that R-loops contain a ss region (demonstrated by their SSBP staining, see Figure 3H), these structures may be also prone to mechanical breakage upon combing. These are potentially important observations in view of the genomic instability accompanying R-loop misregulation (21,22,70,133).

DATA AVAILABILITY

Microarray datasets generated and analyzed in this study have been submitted to Gene Expression Omnibus (GEO), with the following accession number: GSE110218.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

The authors thank the following colleagues for contributing strains and reagents: Wolf-Dietrich Heyer (UC Davis, USA; yeast strain), Martin Reijns (The University of Edinburgh, UK; RNase H2) and David Tollervey (The University of Edinburgh, UK; reagents). We thank Haiqing Fu (NIH) for helping with the combing experiments. Thanks are due to Alastair Kerr (The University of Edinburgh, UK) for his advice with bioinformatic analyses. Thanks are due to Adél Nagy Vezendiné, Ágota Szántó Kelemenné and Andrea Farkas Tankáné (University of Debrecen, Hungary) for technical assistance.

Author Contributions: E.H. did most of the electrophoretic analyses including the rSW studies, evaluated them as well as the Southern blot experiments, first pointed out the nick - RNAP II association within the rDNA cluster and contributed to writing the manuscript. E.K. performed the Southern blot experiments. The majority of combing experiments were conducted by P.N., who also took part in some of the rSW analyses. L.S. did the microarray experiments, the bioinformatic analysis of the data together with L.H. and contributed to writing of the manuscript. The chip-on-beads experiments were performed by L.I. A.E.H. contributed to the conceptualization of the results, helped with advice on the R-loop aspects of the work and revised the manuscript. V.D. supervised the Southern blot experiments and contributed to the conceptualization of the results as well as revised the manuscript. R.J. helped with the combing experiments made by G.Sz. during his Fulbright fellow-

Downloaded from https://academic.oup.com/nar/article/46/20/10649/5078800 by guest on 26 October 2021

ship. Y.P. contributed to the interpretation of the topoisomerase aspects of the work. Zs.A. and L.I. contributed to the CHEF studies and strain validation. M.R.W. prepared and characterized dye-conjugated SSBP. G.Sz. devised the experiments and wrote the manuscript.

FUNDING

Hungarian National Science and Research Foundation (OTKA) [K72762, NK101337, K128770]; TÁMOP [TÁMOP 4.2.2-08/1-2008-0015, TÁMOP 4.2.1/B-09/1/KONV-2010-0007, TÁMOP 4.2.2.A-11/1/KONV-2012-0023 'VÉD-ELEM', TÁMOP 4.2.4. A/2-11-1-2012-0001]; National Excellence Program, National Research, Development and Innovation Office, Hungary, GINOP [GINOP-2.3.2-15-2016-00044]. G.S. was also supported by a Fulbright fellowship. É.H. was also supported by GINOP-2.3.2-15-2016-00026. L.S. was supported by the Hungarian Academy of Sciences [LP2015-9/2015], NKFIH-ERC-HU-117670 and GINOP-2.3.2-15-2016-00024. M.R.W. was supported by the Francis Crick Institute, which receives its core funding from Cancer Research UK [FC001211], the UK Medical Research Council [FC001211] and the Wellcome Trust [FC001211]. R.J. and Y.P. were supported by the Intramural Program of the National Cancer Institute, Center for Cancer Research, NIH (BC 006161). Funding for open access charge: Charity Open Access Fund (COAF) (including the Wellcome Trust and Cancer Research UK).
Conflict of interest statement. None declared.

REFERENCES

1. Madabhushi,R., Gao,F., Pfenning,A.R., Pan,L., Yamakawa,S., Seo,J., Rueda,R., Phan,T.X., Yamakawa,H., Pao,P.C. *et al.* (2015) Activity-induced DNA breaks govern the expression of neuronal Early-Response genes. *Cell*, **161**, 1592–1605.
2. Miyaji,M., Furuta,R., Sano,K., Tsutsui,K.M. and Tsutsui,K. (2015) Genomic regions targeted by DNA topoisomerase IIbeta frequently interact with a nuclear scaffold/matrix protein hnRNP U/SAF-A/SP120. *J. Cell. Biochem.*, **116**, 677–685.
3. Ju,B.G., Lunyak,V.V., Perissi,V., Garcia-Bassets,I., Rose,D.W., Glass,C.K. and Rosenfeld,M.G. (2006) A topoisomerase IIbeta-mediated dsDNA break required for regulated transcription. *Science*, **312**, 1798–1802.
4. Cholewa-Waclaw,J., Bird,A., von Schimmelmann,M., Schaefer,A., Yu,H., Song,H., Madabhushi,R. and Tsai,L.H. (2016) The role of epigenetic mechanisms in the regulation of gene expression in the nervous system. *J. Neurosci.*, **36**, 11427–11434.
5. Guha,M., Saare,M., Maslovskaja,J., Kisand,K., Liiv,I., Haljasorg,U., Tasa,T., Metspalu,A., Milani,L. and Peterson,P. (2017) DNA breaks and chromatin structural changes enhance the transcription of autoimmune regulator target genes. *J. Biol. Chem.*, **292**, 6542–6554.
6. Uuskula-Reimand,L., Hou,H., Samavarchi-Tehrani,P., Rudan,M.V., Liang,M., Medina-Rivera,A., Mohammed,H., Schmidt,D., Schwalie,P. *et al.* (2016) Topoisomerase II beta interacts with cohesin and CTCF at topological domain borders. *Genome Biol.*, **17**, 182.
7. Erener,S., Hesse,M., Kostadinova,R. and Hottiger,M.O. (2012) Poly(ADP-ribose)polymerase-1 (PARP1) controls adipogenic gene expression and adipocyte function. *Mol. Endocrinol.*, **26**, 79–86.
8. Manville,C.M., Smith,K., Sondka,Z., Rance,H., Cockell,S., Cowell,I.G., Lee,K.C., Morris,N.J., Padget,K., Jackson,G.H. *et al.* (2015) Genome-wide ChIP-seq analysis of human TOP2B occupancy in MCF7 breast cancer epithelial cells. *Biol. Open*, **4**, 1436–1447.
9. Baranello,L., Kouzine,F., Wojtowicz,D., Cui,K., Przytycka,T.M., Zhao,K. and Levens,D. (2014) DNA break mapping reveals topoisomerase II activity genome-wide. *Int. J. Mol. Sci.*, **15**, 13111–13122.
10. Yang,F., Kemp,C.J. and Henikoff,S. (2015) Anthracyclines induce double-strand DNA breaks at active gene promoters. *Mutat. Res.*, **773**, 9–15.
11. Gilbert,N. and Allan,J. (2014) Supercoiling in DNA and chromatin. *Curr. Opin. Genet. Dev.*, **25**, 15–21.
12. Ma,J. and Wang,M.D. (2016) DNA supercoiling during transcription. *Biophys. Rev.*, **8**, 75–87.
13. Ma,J. and Wang,M. (2014) Interplay between DNA supercoiling and transcription elongation. *Transcription*, **5**, e28636.
14. Schultz,M.C., Brill,S.J., Ju,Q., Sternglanz,R. and Reeder,R.H. (1992) Topoisomerases and yeast rRNA transcription: negative supercoiling stimulates initiation and topoisomerase activity is required for elongation. *Genes Dev.*, **6**, 1332–1341.
15. Joshi,R.S., Nikolaou,C. and Roca,J. (2018) Structure and chromosomal organization of yeast genes regulated by Topoisomerase II. *Int. J. Mol. Sci.*, **19**, E134.
16. Yan,H., Tammaro,M. and Liao,S. (2016) Collision of trapped topoisomerase 2 with transcription and replication: generation and repair of DNA Double-Strand breaks with 5′ Adducts. *Genes (Basel)*, **7**, E32.
17. Krawczyk,C., Dion,V., Schär,P. and Fritsch,O. (2014) Reversible Top1 cleavage complexes are stabilized strand-specifically at the ribosomal replication fork barrier and contribute to ribosomal DNA stability. *Nucleic Acids Res.*, **42**, 4985–4995.
18. Kumala,S., Hadj-Sahraoui,Y., Rzeszowska-Wolny,J. and Hancock,R. (2012) DNA of a circular minichromosome linearized by restriction enzymes or other reagents is resistant to further cleavage: an influence of chromatin topology on the accessibility of DNA. *Nucleic Acids Res.*, **40**, 9417–9428.
19. Taylor,J.D. and Halford,S.E. (1989) Discrimination between DNA sequences by the EcoRV restriction endonuclease. *Biochemistry*, **28**, 6198–6207.
20. Pingoud,A., Wilson,G.G. and Wende,W. (2014) Type II restriction endonucleases–a historical perspective and more. *Nucleic Acids Res.*, **42**, 7489–7527.
21. Chedin,F. (2016) Nascent Connections: R-Loops and chromatin patterning. *Trends Genet.*, **32**, 828–838.
22. Gaillard,H. and Aguilera,A. (2016) Transcription as a threat to genome integrity. *Annu. Rev. Biochem.*, **85**, 291–317.
23. Rosemeyer,V., Laubrock,A. and Seibl,R. (1995) Nonradioactive 3′-end-labeling of RNA molecules of different lengths by terminal deoxynucleotidyltransferase. *Anal. Biochem.*, **224**, 446–449.
24. Huang,Z. and Szostak,J.W. (1996) A simple method for 3′-labeling of RNA. *Nucleic Acids Res.*, **24**, 4360–4361.
25. Szabo,G. Jr, Boldog,F. and Wikonkal,N. (1990) Disassembly of chromatin into approximately equal to 50 kb units by detergent. *Biochem. Biophys. Res. Commun.*, **169**, 706–712.
26. Szabo,G. Jr and Bacso,Z. (1996) Chromatin isolated from viable human PBLs contains DNA fragmented to $>/=$ 50 kb. *Cell Death Differ.*, **3**, 237–241.
27. Kantidze,O.L. and Razin,S.V. (2009) Chromatin loops, illegitimate recombination, and genome evolution. *Bioessays*, **31**, 278–286.
28. Berezney,R., Dubey,D.D. and Huberman,J.A. (2000) Heterogeneity of eukaryotic replicons, replicon clusters, and replication foci. *Chromosoma*, **108**, 471–484.
29. Bode,J., Benham,C., Ernst,E., Knopp,A., Marschalek,R., Strick,R. and Strissel,P. (2000) Fatal connections: when DNA ends meet on the nuclear matrix. *J. Cell. Biochem. Suppl.*, **35**, 3–22.
30. Bode,J., Goetze,S., Heng,H., Krawetz,S.A. and Benham,C. (2003) From DNA structure to gene expression: mediators of nuclear compartmentalization and dynamics. *Chromosome Res.*, **11**, 435–445.
31. Lemaitre,J.M., Danis,E., Pasero,P., Vassetzky,Y. and Méchali,M. (2005) Mitotic remodeling of the replicon and chromosome structure. *Cell*, **123**, 787–801.
32. Demeret,C., Vassetzky,Y. and Mechali,M. (2001) Chromatin remodelling and DNA replication: from nucleosomes to loop domains. *Oncogene*, **20**, 3086–3093.
33. Szekvolgyi,L., Hegedüs,E., Molnár,M., Bacsó,Z., Szarka,K., Beck,Z., Dombrádi,V., Austin,C. and Szabó,G. (2006) Nick-forming

Downloaded from https://academic.oup.com/nar/article/46/20/10649/5078800 by guest on 26 October 2021

dc_1977_21

sequences may be involved in the organization of eukaryotic chromatin into approximately 50 kbp loops. *Histochem. Cell Biol.*, **125**, 63–73.
34. Desai,N.A. and Shankar,V. (2003) Single-strand-specific nucleases. *FEMS Microbiol. Rev.*, **26**, 457–491.
35. Hershey,H.V. and Werner,D. (1976) Evidence for non-deoxynucleotide linkers in Ehrlich ascites tumour cell DNA. *Nature*, **262**, 148–150.
36. Varga,T., Szilagyi,I. and Szabo,G. Jr (1999) Single-strand breaks in agarose-embedded chromatin of nonapoptotic cells. *Biochem. Biophys. Res. Commun.*, **264**, 388–394.
37. Szekvolgyi,L., Rákosy,Z., Bálint,B.L., Kókai,E., Imre,L., Vereb,G., Bacsó,Z., Goda,K., Varga,S., Balázs,M. *et al.* (2007) Ribonucleoprotein-masked nicks at 50-kbp intervals in the eukaryotic genomic DNA. *Proc. Natl. Acad. Sci. U.S.A.*, **104**, 14964–14969.
38. Gal,I., Varga,T., Szilágyi,I., Balázs,M., Schlammadinger,J. and Szabó,G. Jr (2000) Protease-elicited TUNEL positivity of non-apoptotic fixed cells. *J. Histochem. Cytochem.*, **48**, 963–970.
39. Puc,J., Aggarwal,A.K. and Rosenfeld,M.G. (2017) Physiological functions of programmed DNA breaks in signal-induced transcription. *Nat. Rev. Mol. Cell Biol.*, **18**, 471–476.
40. Kim,Y.H., Ishikawa,D., Ha,H.P., Sugiyama,M., Kaneko,Y. and Harashima,S. (2006) Chromosome XII context is important for rDNA function in yeast. *Nucleic Acids Res.*, **34**, 2914–2924.
41. Woolford,J.L. Jr and Baserga,S.J. (2013) Ribosome biogenesis in the yeast Saccharomyces cerevisiae. *Genetics*, **195**, 643–681.
42. Mekhail,K., Seebacher,J., Gygi,S.P. and Moazed,D. (2008) Role for perinuclear chromosome tethering in maintenance of genome stability. *Nature*, **456**, 667–670.
43. Takeuchi,Y., Horiuchi,T. and Kobayashi,T. (2003) Transcription-dependent recombination and the role of fork collision in yeast rDNA.. *Genes Dev.*, **17**, 1497–1506.
44. Kobayashi,T. (2008) A new role of the rDNA and nucleolus in the nucleus–rDNA instability maintains genome integrity. *Bioessays*, **30**, 267–272.
45. Kobayashi,T. (2011) Regulation of ribosomal RNA gene copy number and its role in modulating genome integrity and evolutionary adaptability in yeast. *Cell. Mol. Life Sci.*, **68**, 1395–1403.
46. Kobayashi,T. (2014) Ribosomal RNA gene repeats, their stability and cellular senescence. *Proc. Jpn. Acad. Ser. B Phys. Biol. Sci.*, **90**, 119–129.
47. French,S.L., Osheim,Y.N., Cioci,F., Nomura,M. and Beyer,A.L. (2003) In exponentially growing Saccharomyces cerevisiae cells, rRNA synthesis is determined by the summed RNA polymerase I loading rate rather than by the number of active genes. *Mol. Cell. Biol.*, **23**, 1558–1568.
48. Pasero,P., Bensimon,A. and Schwob,E. (2002) Single-molecule analysis reveals clustering and epigenetic regulation of replication origins at the yeast rDNA locus. *Genes Dev.*, **16**, 2479–2484.
49. Mayan,M. and Aragon,L. (2010) Cis-interactions between non-coding ribosomal spacers dependent on RNAP-II separate RNAP-I and RNAP-III transcription domains. *Cell Cycle*, **9**, 4328–4337.
50. Seiler,J.A., Conti,C., Syed,A., Aladjem,M.I. and Pommier,Y. (2007) The intra-S-phase checkpoint affects both DNA replication initiation and elongation: single-cell and -DNA fiber analyses. *Mol. Cell. Biol.*, **27**, 5806–5818.
51. Benoukraf,T., Cauchy,P., Fenouil,R., Jeanniard,A., Koch,F., Jaeger,S., Thieffry,D., Imbert,J., Andrau,J.C., Spicuglia,S. *et al.* (2009) CoCAS: a ChIP-on-chip analysis suite. *Bioinformatics*, **25**, 954–955.
52. Halasz,L., Karányi,Z., Boros-Oláh,B., Kuik-Rózsa,T., Sipos,É., Nagy,É., Mosolygó-L,Á., Mázló,A., Rajnavölgyi,É., Halmos,G. *et al.* (2017) RNA-DNA hybrid (R-loop) immunoprecipitation mapping: an analytical workflow to evaluate inherent biases. *Genome Res.*, **27**, 1063–1073.
53. Quinlan,A.R. and Hall,I.M. (2010) BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics*, **26**, 841–842.
54. Steinmetz,E.J., Warren,C.L., Kuehner,J.N., Panbehi,B., Ansari,A.Z. and Brow,D.A. (2006) Genome-wide distribution of yeast RNA polymerase II and its control by Sen1 helicase. *Mol. Cell*, **24**, 735–746.
55. Ramirez,F., Dündar,F., Diehl,S., Grüning,B.A. and Manke,T. (2014) deepTools: a flexible platform for exploring deep-sequencing data. *Nucleic Acids Res.*, **42**, W187–W191.
56. Mulla,W.A., Seidel,C.W., Zhu,J., Tsai,H.J., Smith,S.E., Singh,P., Bradford,W.D., McCroskey,S., Nelliat,A.R., Conkright,J. *et al.* (2017) Aneuploidy as a cause of impaired chromatin silencing and mating-type specification in budding yeast. *Elife*, **6**, e27991.
57. Hegedus,E., Kókai,E., Kotlyar,A., Dombrádi,V. and Szabó,G. (2009) Separation of 1-23-kb complementary DNA strands by urea-agarose gel electrophoresis. *Nucleic Acids Res.*, **37**, e112.
58. Gebhard,C., Schwarzfischer,L., Pham,T.H., Schilling,E., Klug,M., Andreesen,R. and Rehli,M. (2006) Genome-wide profiling of CpG methylation identifies novel targets of aberrant hypermethylation in myeloid leukemia. *Cancer Res.*, **66**, 6118–6128.
59. Rosemeyer,V., Seibl,R. and Laubrock,A. (1996) 3′-RNA labelling with terminal transferase. Google Patents.
60. Cerritelli,S.M. and Crouch,R.J. (2009) Ribonuclease H: the enzymes in eukaryotes. *FEBS J.*, **276**, 1494–1505.
61. Loomis,E.W., Sanz,L.A., Chédin,F. and Hagerman,P.J. (2014) Transcription-associated R-loop formation across the human FMR1 CGG-repeat region. *PLos Genet.*, **10**, e1004294.
62. Sanz,L.A., Hartono,S.R., Lim,Y.W., Steyaert,S., Rajpurkar,A., Ginno,P.A., Xu,X. and Chédin,F. (2016) Prevalent, dynamic, and conserved R-Loop structures associate with specific epigenomic signatures in mammals. *Mol. Cell*, **63**, 167–178.
63. Sparks,J.L., Chon,H., Cerritelli,S.M., Kunkel,T.A., Johansson,E., Crouch,R.J. and Burgers,P.M. (2012) RNase H2-initiated ribonucleotide excision repair. *Mol. Cell*, **47**, 980–986.
64. Clausen,A.R., Williams,J.S. and Kunkel,T.A. (2015) Measuring ribonucleotide incorporation into DNA in vitro and in vivo. *Methods Mol. Biol.*, **1300**, 123–139.
65. Tannous,E., Kanaya,E. and Kanaya,S. (2015) Role of RNase H1 in DNA repair: removal of single ribonucleotide misincorporated into DNA in collaboration with RNase H2. *Sci. Rep.*, **5**, 9969.
66. Reijns,M.A.M., Kemp,H., Ding,J., de Procé,S.M., Jackson,A.P. and Taylor,M.S. (2015) Lagging-strand replication shapes the mutational landscape of the genome. *Nature*, **518**, 502–506.
67. Imre,L., Simándi,Z., Horváth,A., Fenyőfalvi,G., Nánási,P., Niaki,E.F., Hegedüs,É., Bacsó,Z., Weyemi,U., Mauser,R. *et al.* (2017) Nucleosome stability measured in situ by automated quantitative imaging. *Sci. Rep.*, **7**, 12734.
68. Vogelstein,B., Pardoll,D.M. and Coffey,D.S. (1980) Supercoiled loops and eucaryotic DNA replicaton. *Cell*, **22**, 79–85.
69. Dobrovolskaia,I.V., Kenward,M. and Arya,G. (2010) Twist propagation in dinucleosome arrays. *Biophys. J.*, **99**, 3355–3364.
70. Santos-Pereira,J.M. and Aguilera,A. (2015) R loops: new modulators of genome dynamics and function. *Nat. Rev. Genet.*, **16**, 583–597.
71. Skourti-Stathaki,K., Kamieniarz-Gdula,K. and Proudfoot,N.J. (2014) R-loops induce repressive chromatin marks over mammalian gene terminators. *Nature*, **516**, 436–439.
72. Szekvolgyi,L., Imre,L., Minh,D.X., Hegedus,E., Bacso,Z. and Szabo,G. (2009) Flow cytometric and laser scanning microscopic approaches in epigenetics research. *Methods Mol. Biol.*, **567**, 99–111.
73. Szekvolgyi,L., Bálint,B.L., Imre,L., Goda,K., Szabó,M., Nagy,L. and Szabó,G. (2006) Chip-on-beads: flow-cytometric evaluation of chromatin immunoprecipitation. *Cytometry A*, **69**, 1086–1091.
74. Egloff,S. and Murphy,S. (2008) Cracking the RNA polymerase II CTD code. *Trends Genet.*, **24**, 280–288.
75. Fili,N., Mashanov,G.I., Toseland,C.P., Batters,C., Wallace,M.I., Yeeles,J.T., Dillingham,M.S., Webb,M.R. and Molloy,J.E. (2010) Visualizing helicases unwinding DNA at the single molecule level. *Nucleic Acids Res.*, **38**, 4448–4457.
76. Materna,T., Weber,S., Kofler-Mongold,V. and Phares,W. (1998) Electrophoretic separation of both single- and double-stranded nucleic acids in the same urea-containing agarose gel. *Anal. Biochem.*, **255**, 161–163.
77. El Hage,A., French,S.L., Beyer,A.L. and Tollervey,D. (2010) Loss of Topoisomerase I leads to R-loop-mediated transcriptional blocks during ribosomal RNA synthesis. *Genes Dev.*, **24**, 1546–1558.
78. El Hage,A., Webb,S., Kerr,A. and Tollervey,D. (2014) Genome-wide distribution of RNA-DNA hybrids identifies RNase H targets in

Downloaded from https://academic.oup.com/nar/article/46/20/10649/5078800 by guest on 26 October 2021

tRNA genes, retrotransposons and mitochondria. *PLos Genet.*, **10**, e1004716.
79. Saka,K., Ide,S., Ganley,A.R. and Kobayashi,T. (2013) Cellular senescence in yeast is regulated by rDNA noncoding transcription. *Curr. Biol.*, **23**, 1794–1798.
80. Galopier,A. and Hermann-Le Denmat,S. (2011) Mitochondria of the yeasts Saccharomyces cerevisiae and Kluyveromyces lactis contain nuclear rDNA-encoded proteins. *PLoS One*, **6**, e16325.
81. Cesarini,E., Mariotti,F.R., Cioci,F. and Camilloni,G. (2010) RNA polymerase I transcription silences noncoding RNAs at the ribosomal DNA locus in Saccharomyces cerevisiae. *Eukaryot. Cell*, **9**, 325–335.
82. Kobayashi,T. and Ganley,A.R. (2005) Recombination regulation by transcription-induced cohesin dissociation in rDNA repeats. *Science*, **309**, 1581–1584.
83. Srivastava,R., Srivastava,R. and Ahn,S.H. (2016) The epigenetic pathways to ribosomal DNA silencing. *Microbiol. Mol. Biol. Rev.*, **80**, 545–563.
84. D'Alfonso,A., Di Felice,F., Carlini,V., Wright,C.M., Hertz,M.I., Bjornsti,M.A. and Camilloni,G. (2016) Molecular mechanism of DNA topoisomerase I-Dependent rDNA Silencing: Sir2p recruitment at ribosomal genes. *J. Mol. Biol.*, **428**, 4905–4916.
85. Szilagyi,I., Varga,T., Székvölgyi,L., Hegedüs,E., Goda,K., Kaczur,V., Bacsó,Z., Nakayama,Y., Pósafi,J., Pongor,S. *et al.* (2003) Non-random features of loop-size chromatin fragmentation. *J. Cell. Biochem.*, **89**, 1193–1205.
86. Dammann,R., Lucchini,R., Koller,T. and Sogo,J.M. (1993) Chromatin structures and transcription of rDNA in yeast Saccharomyces cerevisiae. *Nucleic Acids Res.*, **21**, 2331–2338.
87. Hamperl,S., Wittner,M., Babl,V., Perez-Fernandez,J., Tschochner,H. and Griesenbeck,J. (2013) Chromatin states at ribosomal DNA loci. *Biochim. Biophys. Acta*, **1829**, 405–417.
88. Xu,Z., Wei,W., Gagneur,J., Perocchi,F., Clauder-Münster,S., Camblong,J., Guffanti,E., Stutz,F., Huber,W. and Steinmetz,L.M. (2009) Bidirectional promoters generate pervasive transcription in yeast. *Nature*, **457**, 1033–1037.
89. Neil,H., Malabat,C., d'Aubenton-Carafa,Y., Xu,Z., Steinmetz,L.M. and Jacquier,A. (2009) Widespread bidirectional promoters are the major source of cryptic transcripts in yeast. *Nature*, **457**, 1038–1042.
90. Colin,J., Libri,D. and Porrua,O. (2011) Cryptic transcription and early termination in the control of gene expression. *Genet. Res. Int.*, **2011**, 653494.
91. Poole,A.M., Kobayashi,T. and Ganley,A.R. (2012) A positive role for yeast extrachromosomal rDNA circles? Extrachromosomal ribosomal DNA circle accumulation during the retrograde response may suppress mitochondrial cheats in yeast through the action of TAR1. *Bioessays*, **34**, 725–729.
92. Nomura,M., Nogi,Y. and Oakes,M. (2000–2013) Transcription of rDNA in the Yeast Saccharomyces cerevisiae. *Madame Curie Bioscience Database*. Landes Bioscience, Georgetown.
93. Huang,Y. (2002) Transcriptional silencing in Saccharomyces cerevisiae and Schizosaccharomyces pombe. *Nucleic Acids Res.*, **30**, 1465–1482.
94. Cioci,F., Vu,L., Eliason,K., Oakes,M., Siddiqi,I.N. and Nomura,M. (2003) Silencing in yeast rDNA chromatin: reciprocal relationship in gene expression between RNA polymerase I and II. *Mol. Cell*, **12**, 135–145.
95. Santoro,R. (2005) The silence of the ribosomal RNA genes. *Cell. Mol. Life Sci.*, **62**, 2067–2079.
96. Buck,S.W., Maqani,N., Matecic,M., Hontz,R.D., Fine,R.D., Li,M. and Smith,J.S. (2016) RNA Polymerase I and Fob1 contributions to transcriptional silencing at the yeast rDNA locus. *Nucleic Acids Res.*, **44**, 6173–6184.
97. Cioci,F., Vogelauer,M. and Camilloni,G. (2002) Acetylation and accessibility of rDNA chromatin in Saccharomyces cerevisiae in (Delta)top1 and (Delta)sir2 mutants. *J. Mol. Biol.*, **322**, 41–52.
98. Wang,J.C. (2002) Cellular roles of DNA topoisomerases: a molecular perspective. *Nat. Rev. Mol. Cell Biol.*, **3**, 430–440.
99. Xu,Y. and Her,C. (2015) Inhibition of Topoisomerase (DNA) I (TOP1): DNA damage repair and anticancer therapy. *Biomolecules*, **5**, 1652–1670.
100. Joshi,R.S., Pina,B. and Roca,J. (2012) Topoisomerase II is required for the production of long Pol II gene transcripts in yeast. *Nucleic Acids Res.*, **40**, 7907–7915.
101. French,S.L., Sikes,M.L., Hontz,R.D., Osheim,Y.N., Lambert,T.E., El Hage,A., Smith,M.M., Tollervey,D., Smith,J.S. and Beyer,A.L. (2011) Distinguishing the roles of Topoisomerases I and II in relief of transcription-induced torsional stress in yeast rRNA genes. *Mol. Cell. Biol.*, **31**, 482–494.
102. Roedgaard,M., Fredsoe,J., Pedersen,J.M., Bjergbaek,L. and Andersen,A.H. (2015) DNA topoisomerases are required for preinitiation complex assembly during GAL gene activation. *PLoS One*, **10**, e0132739.
103. Uthe,H., Vanselow,J.T. and Schlosser,A. (2017) Proteomic analysis of the mediator complex interactome in Saccharomyces cerevisiae. *Sci. Rep.*, **7**, 43584.
104. Burkhalter,M.D. and Sogo,J.M. (2004) rDNA enhancer affects replication initiation and mitotic recombination: Fob1 mediates nucleolytic processing independently of replication. *Mol. Cell*, **15**, 409–421.
105. Interthal,H., Pouliot,J.J. and Champoux,J.J. (2001) The tyrosyl-DNA phosphodiesterase Tdp1 is a member of the phospholipase D superfamily. *Proc. Natl. Acad. Sci. U.S.A.*, **98**, 12009–12014.
106. Wang,L., Zhou,Y., Xu,L., Xiao,R., Lu,X., Chen,L., Chong,J., Li,H., He,C., Fu,X.D. *et al.* (2015) Molecular basis for 5-carboxycytosine recognition by RNA polymerase II elongation complex. *Nature*, **523**, 621–625.
107. Mao,P., Brown,A.J., Malc,E.P., Mieczkowski,P.A., Smerdon,M.J., Roberts,S.A. and Wyrick,J.J. (2017) Genome-wide maps of alkylation damage, repair, and mutagenesis in yeast reveal mechanisms of mutational heterogeneity. *Genome Res.*, **27**, 1674–1684.
108. Gondor,A. and Ohlsson,R. (2018) Enhancer functions in three dimensions: beyond the flat world perspective. *F1000Res.*, **7**, doi:10.12688/f1000research.13842.1.
109. Tombline,G., Millen,J.I., Polevoda,B., Rapaport,M., Baxter,B., Van Meter,M., Gilbertson,M., Madrey,J., Piazza,G.A., Rasmussen,L. *et al.* (2017) Effects of an unusual poison identify a lifespan role for Topoisomerase 2 in Saccharomyces cerevisiae. *Aging (Albany, NY)*, **9**, 68–97.
110. Wagner,M., Price,G. and Rothstein,R. (2006) The absence of Top3 reveals an interaction between the Sgs1 and Pif1 DNA helicases in Saccharomyces cerevisiae. *Genetics*, **174**, 555–573.
111. Pedersen,J.M., Fredsoe,J., Roedgaard,M., Andreasen,L., Mundbjerg,K., Kruhøffer,M., Brinch,M., Schierup,M.H., Bjergbaek,L. and Andersen,A.H. (2012) DNA Topoisomerases maintain promoters in a state competent for transcriptional activation in Saccharomyces cerevisiae. *PLos Genet.*, **8**, e1003128.
112. Caldecott,K., Banks,G. and Jeggo,P. (1993) The induction and reversal of topoisomerase II cleavable complexes formed by nuclear extract from the CHO DNA repair mutant, xrs1. *Mutat. Res.*, **293**, 259–267.
113. Warters,R.L. and Barrows,L.R. (1994) Heat sensitivity of HeLa S3 cell DNA topoisomerase II. *J. Cell. Physiol.*, **159**, 468–674.
114. Kampinga,H.H., van den Kruk,G. and Konings,A.W. (1989) Reduced DNA break formation and cytotoxicity of the topoisomerase II drug 4′-(9′-acridinylamino)methanesulfon-m-anisidide when combined with hyperthermia in human and rodent cell lines. *Cancer Res.*, **49**, 1712–1717.
115. Pulse Field Gel Electrophoresis, Commentary. In: Frederick,RB, Ausubel,M, Kingston,RE, Moore,DD, Seidman,JG, Smith,JA and Struhl,K (eds). *Current Protocols in Molecular Biology*. John Wiley and Sons, Inc., New Jersey, 1994–1998.
116. Zimmermann,R.M. and Cox,E.C. (1994) DNA stretching on functionalized gold surfaces. *Nucleic Acids Res.*, **22**, 492–497.
117. Pelechano,V., Chavez,S. and Perez-Ortin,J.E. (2010) A complete set of nascent transcription rates for yeast genes. *PLoS One*, **5**, e15442.
118. Sperling,A.S., Jeong,K.S., Kitada,T. and Grunstein,M. (2011) Topoisomerase II binds nucleosome-free DNA and acts redundantly with topoisomerase I to enhance recruitment of RNA Pol II in budding yeast. *Proc. Natl. Acad. Sci. U.S.A.*, **108**, 12693–12698.
119. Mayan,M.D. and Aragon,L. (2014) Chromosome conformation capture (3C) of tandem arrays in yeast. *Methods Mol. Biol.*, **1205**, 219–229.
120. Singh,B.N., Ansari,A. and Hampsey,M. (2009) Detection of gene loops by 3C in yeast. *Methods*, **48**, 361–367.

Downloaded from https://academic.oup.com/nar/article/46/20/10649/5078800 by guest on 26 October 2021

dc_1977_21

121. Sun,X., Huang,L., Markowitz,T.E., Blitzblau,H.G., Chen,D., Klein,F. and Hochwagen,A. (2015) Transcription dynamically patterns the meiotic chromosome-axis interface. *Elife*, **4**, e07424.
122. Uzunova,K., Georgieva,M. and Miloshev,G. (2013) Saccharomyces cerevisiae linker histone-Hho1p maintains chromatin loop organization during ageing. *Oxid. Med. Cell Longev.*, **2013**, 437146.
123. Acquaviva,L., Székvölgyi,L., Dichtl,B., Dichtl,B.S., de La Roche Saint André,C., Nicolas,A. and Géli,V. (2013) The COMPASS subunit Spp1 links histone methylation to initiation of meiotic recombination. *Science*, **339**, 215–218.
124. Sommermeyer,V., Béneut,C., Chaplais,E., Serrentino,M.E. and Borde,V. (2013) Spp1, a member of the Set1 Complex, promotes meiotic DSB formation in promoters by tethering histone H3K4 methylation sites to chromosome axes. *Mol. Cell*, **49**, 43–54.
125. Heng,H.H., Chamberlain,J.W., Shi,X.M., Spyropoulos,B., Tsui,L.C. and Moens,P.B. (1996) Regulation of meiotic chromatin loop size by chromosomal position. *Proc. Natl. Acad. Sci. U.S.A.*, **93**, 2795–2800.
126. Moens,P.B. and Pearlman,R.E. (1988) Chromatin organization at meiosis. *Bioessays*, **9**, 151–153.
127. O'Sullivan,J.M. (2010) Nucleolar structure: it's all in a tangle. *Cell Cycle*, **9**, 4609–4610.
128. Mayan,M.D. (2013) RNAP-II molecules participate in the anchoring of the ORC to rDNA replication origins. *PLoS One*, **8**, e53405.
129. Hayward,G.S. (1974) Unique double-stranded fragments of bacteriophage T5 DNA resulting from preferential shear-induced breakage at nicks. *Proc. Natl. Acad. Sci. U.S.A.*, **71**, 2108–2112.
130. Canela,A., Maman,Y., Jung,S., Wong,N., Callen,E., Day,A., Kieffer-Kwon,K.R., Pekowska,A., Zhang,H. *et al.* (2017) Genome organization drives chromosome fragility. *Cell*, **170**, 507–521.
131. Schwer,B., Wei,P.C., Chang,A.N., Kao,J., Du,Z., Meyers,R.M. and Alt,F.W. (2016) Transcription-associated processes cause DNA double-strand breaks and translocations in neural stem/progenitor cells. *Proc. Natl. Acad. Sci. U.S.A.*, **113**, 2258–2263.
132. Stults,D.M., Killen,M.W., Williamson,E.P., Hourigan,J.S., Vargas,H.D., Arnold,S.M., Moscow,J.A. and Pierce,A.J. (2009) Human rRNA gene clusters are recombinational hotspots in cancer. *Cancer Res.*, **69**, 9096–9104.
133. Richard,P. and Manley,J.L. (2016) R loops and links to human disease. *J. Mol. Biol.*, **429**, 3168–3180.
134. Chan,K.M., Liu,Y.T., Ma,C.H., Jayaram,M. and Sau,S. (2013) The 2 micron plasmid of Saccharomyces cerevisiae: a miniaturized selfish genome with optimized functional competence. *Plasmid*, **70**, 2–17.
135. Vogelauer,M. and Camilloni,G. (1999) Site-specific in vivo cleavages by DNA topoisomerase I in the regulatory regions of the 35 S rRNA in Saccharomyces cerevisiae are transcription independent. *J. Mol. Biol.*, **293**, 19–28.

Downloaded from https://academic.oup.com/nar/article/46/20/10649/5078800 by guest on 26 October 2021

frontiers in Immunology

dc_1977_21

MINI REVIEW
published: 31 January 2018
doi: 10.3389/fimmu.2018.00151

The Role of Indoleamine-2,3-Dioxygenase in Cancer Development, Diagnostics, and Therapy

*Lilla Hornyák[1†], Nikoletta Dobos[2†], Gábor Koncz[3], Zsolt Karányi[1,4], Dénes Páll[4], Zoltán Szabó[5], Gábor Halmos[2] and Lóránt Székvölgyi[1]**

[1] *MTA-DE Momentum Genome Architecture and Recombination Research Group, Department of Biochemistry and Molecular Biology, Faculty of Medicine, University of Debrecen, Debrecen, Hungary,* [2] *Department of Biopharmacy, Faculty of Pharmacology, University of Debrecen, Debrecen, Hungary,* [3] *Department of Immunology, Faculty of Medicine, University of Debrecen, Debrecen, Hungary,* [4] *Department of Internal Medicine, Faculty of Medicine, University of Debrecen, Debrecen, Hungary,* [5] *Department of Emergency Medicine, Faculty of Medicine, University of Debrecen, Debrecen, Hungary*

Tumors are composed of abnormally transformed cell types and tissues that differ from normal tissues in their genetic and epigenetic makeup, metabolism, and immunology. Molecular compounds that modulate the immune response against neoplasms offer promising new strategies to combat cancer. Inhibitors targeting the indoleamine-2,3-dioxygenase 1 enzyme (IDO1) represent one of the most potent therapeutic opportunities to inhibit tumor growth. Herein, we assess the biochemical role of IDO1 in tumor metabolism and immune surveillance, and review current diagnostic and therapeutic approaches that are intended to increase the effectiveness of immunotherapies against highly aggressive and difficult-to-treat IDO-expressing cancers.

Keywords: indoleamine-2,3-dioxygenase, gene expression, metabolism, immunotherapy, immune surveillance, cancer diagnostics, clinical trial

OPEN ACCESS

Edited by:
Suhendan Ekmekcioglu, University of Texas MD Anderson Cancer Center, United States

Reviewed by:
Paolo Puccetti, University of Perugia, Italy
Abhishek D. Garg, KU Leuven, Belgium
Maria Laura Belladonna, University of Perugia, Italy

****Correspondence:***
Lóránt Székvölgyi
lorantsz@med.unideb.hu

[†]These authors have contributed equally to this work.

Specialty section:
This article was submitted to Cancer Immunity and Immunotherapy, a section of the journal Frontiers in Immunology

Received: *12 December 2017*
Accepted: *17 January 2018*
Published: *31 January 2018*

Citation:
Hornyák L, Dobos N, Koncz G, Karányi Z, Páll D, Szabó Z, Halmos G and Székvölgyi L (2018) The Role of Indoleamine-2,3-Dioxygenase in Cancer Development, Diagnostics, and Therapy. Front. Immunol. 9:151. doi: 10.3389/fimmu.2018.00151

INTRODUCTION

Biochemistry and Regulation of Indoleamine-2,3-Dioxygenase 1 (IDO1) Activity

Indolamine-2,3-dioxygenase 1 is a cytosolic enzyme with a heme (Fe^{2+}) prosthetic group that catalyzes the first and rate-limiting step of tryptophan (Trp) catabolism (**Figure 1A**). IDO1 converts the essential amino acid Trp to kynurenine (Kyn) by cleaving the 2,3-double bond of the indole ring while a molecular oxygen (O_2) merges into the unsealed molecule. The product is *N*-formylkynurenine that becomes rapidly and spontaneously transformed into Kyn (1). In the next steps, Kyn is further converted to other active metabolites, such as hydroxykynurenine, anthranilic acid, kynurenic acid, 3-hydroxyanthranilic acid, quinolinic acid, and picolinic acid (**Figure 1A**). The two end-products of the pathway are NAD^+ and ATP that both fuel cellular metabolism (2).

In humans, IDO1 has an evolutionary paralog (indolamine-2,3-dioxygenase 2; IDO2) and a functional ortholog (tryptophan-2,3-dioxygenase; TDO) that catalyze the same biochemical reaction; however, IDO2 and TDO show high tissue specificity and much lower expression level than IDO1 that significantly restrict their activity (2, 4). TDO is transcribed only in the liver[1] [protein-level expression is not established (5)] and its major function is to control the free Trp concentration of the blood (4). IDO2 mRNA is expressed at low levels in the placenta and liver (protein expression is not known[2]), while IDO1 shows a high protein expression in the peripheral lymph organs (lymph nodes, spleen and tonsils[3]).

[1] https://www.proteinatlas.org/ENSG00000151790-TDO2/tissue.
[2] https://www.proteinatlas.org/ENSG00000188676-IDO2/tissue.
[3] https://www.proteinatlas.org/ENSG00000131203-IDO1/tissue.

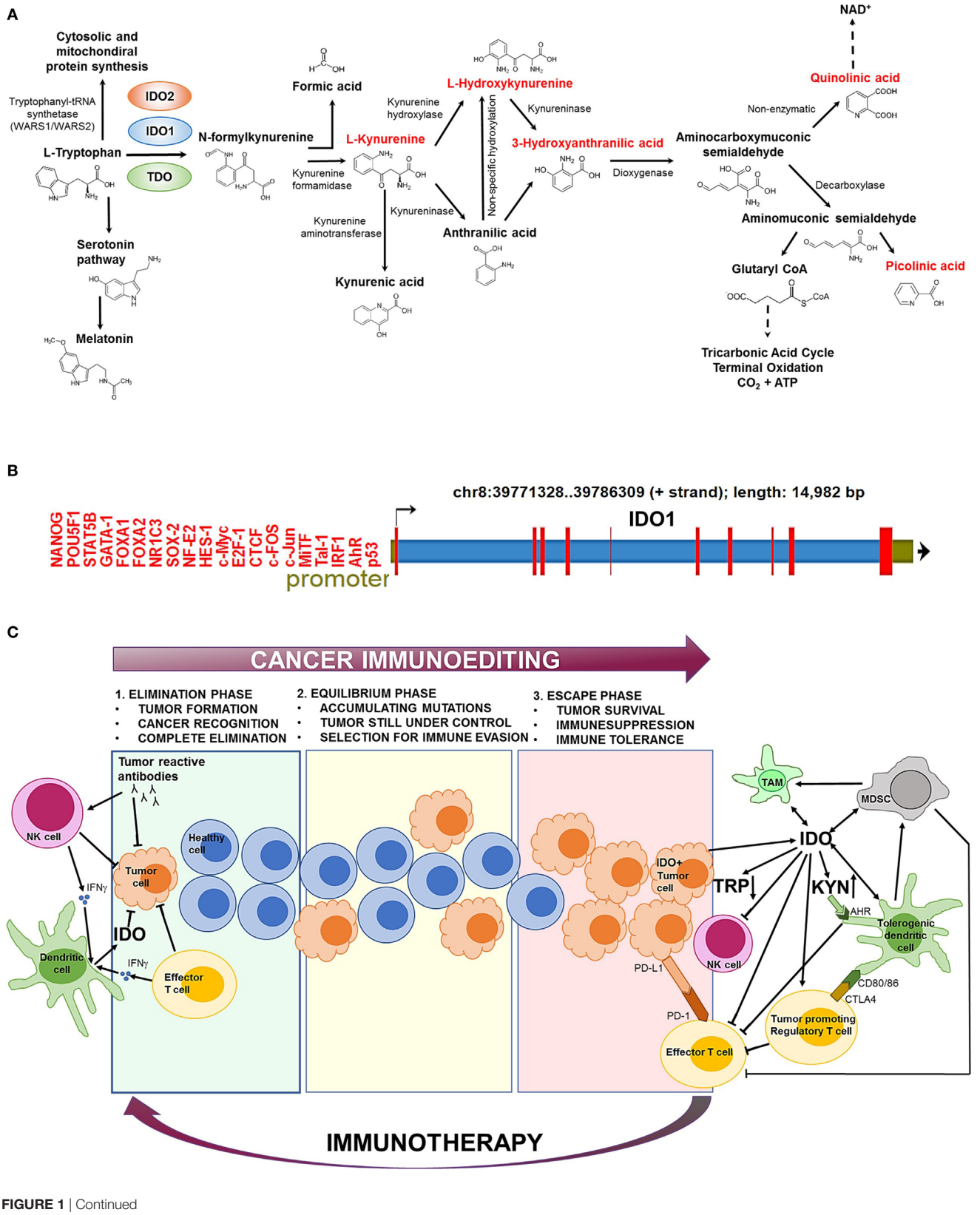


FIGURE 1 | Continued

dc_1977_21

FIGURE 1 | The biochemical function and regulation of indolamine-2,3-dioxygenase 1 (IDO1). **(A)** The kynurenine (Kyn) pathway of tryptophan (Trp) catabolism. L-Trp is metabolized in three separate biochemical pathways (indicated by arrows). In the Kyn pathway, IDO1/IDO2 and tryptophan-2,3-dioxygenase (TDO) catalyze the first and rate-limiting step of Trp degradation that gives rise to *N*-formylkynurenine. *N*-formylkynurenine is then transformed into L-Kyn and formic acid by kynurenine formamidase. L-Kyn is converted to anthranilic acid by kynureninase or L-hydroxykynurenine by kynurenine hydroxylase. Non-specific hydroxylation of anthranilic acid results in L-hydroxykynurenine. Kynureninase converts L-hydroxykynurenine to 3-hydroxyanthranilic acid that is further metabolized by hydroxyanthranilate dioxygenase to aminocarboxymuconic semialdehyde. The semialdehyde spontaneously forms quinolinic acid that is a precursor of NAD^+ synthesis, or a decarboxylase enzyme converts it to aminomuconic semialdehyde. Aminomuconic semialdehyde is then converted to picolinic acid or glutaryl-CoA that is metabolized in the tricarbonic acid cycle and terminal oxidation. Metabolites that are highlighted in red have been directly implicated in immunosuppressive mechanisms and cancer development. **(B)** The structure of the IDO1 gene. IDO1 is located on chromosome 8 [39771328–39786309 forward (+) strand; 14,982 base pairs] comprising 10 exonic region (red bars). The promoter region (green section upstream the transcription start site) contains several transcription factor-binding sites that have been identified by ChIP sequencing. ChIP peaks were collected from the GTRD database of transcription binding sites (3). Only normal (non-transformed) cell types were considered. **(C)** The role of IDO1 in cancer immunoediting. In the first phase of immunoediting ("elimination"), sporadically arising transformed cells are destroyed by the innate and adaptive immune systems. Activated B cells produce tumor reactive antibodies to eradicate most transformed cells. Natural killer (NK) cells and effector T cells release inflammatory cytokines, such as IFN-γ, which activate dendritic cells (DCs) that secrete low levels of IDO1. IDO1 depletes the essential amino acid Trp from the tumor microenvironment that inhibits tumor growth. In the "equilibrium" phase, surviving tumor cells are still controlled by the immune system; however, they rapidly accumulate mutations. When the immune system can no longer block the abnormal and autonomous growth of "edited" cells, the tumor becomes clinically manifested ("escape"). The escape phase is associated with high IDO1 level that is primarily produced by tumor cells and tolerogenic immune cells [e.g., tolerogenic DCs, myeloid-derived suppressor cells (MDSCs), tumor-associated macrophages (TAMs)]. Trp depletion and Kyn accumulation lead to immunosuppression and tolerogenicity by inhibiting effector T cell and NK cell functions and stimulating regulatory T cells. IDO1 also promotes the expansion and activation of MDSCs and induces polarization of macrophages to a tolerogenic phenotype. Increased Kyn levels activate the aryl hydrocarbon receptor (AhR) that switch the activity of DCs from immunogenic to tolerogenic. Elevated CTLA4 expression of regulatory T cells results in further increase of IDO1 secretion by DCs. IDO1-induced expansion and activation of regulatory T cells, tolerogenic DCs, and MDSCs suppress the activity of antitumor effector T cells. Other immunosuppressive agents (e.g., PD-L1/PD-1, CTLA4) also inhibit effector T cell functions. Oncological immunotherapy aims to reverse immunoediting (backward arrow) by inhibiting and activating local immunosuppressive and tumor eradication mechanisms, respectively.

The activity of IDO1 is mainly regulated at the transcriptional level (**Figure 1B**). *Bona fide* transcription factor binding sites have been detected by ChIP-seq for a large catalog of human transcription factors (3). The identified gene regulatory proteins involve (i) NF- κB, which allows IDO1 mRNA expression regulation through the interferon pathway (6), (ii) the aryl hydrocarbon receptor (AhR) that binds to putative dendritic-cell responsive elements and promotes the L-Kyn-dependent induction of IDO1 (7, 8), and (iii) CTCF that mediates IDO1 expression *via* long-range chromatin looping interactions between the promoter and distal enhancer regions (9).

Superimposed on the transcriptional control of IDO1, specific posttranslational mechanisms also operate that affect the activity and half-life of the enzyme. For instance, the diffusible messenger nitrogen monoxide (NO) reacts with the heme cofactor of IDO1 generating ferric (Fe^{3+}) heme and nitrate (NO_3), which leads to the dose-dependent and reversible inhibition of enzymatic activity (10, 11). It has been also shown that endogenous NO production accelerates the proteasomal degradation of IDO1 (12). Other antioxidants like the anti-inflammatory agent pyrrolidine dithiocarbamate (13) restrict the availability of heme and thereby block holoenzyme assembly (14). In inflammatory conditions, NO and superoxide anions (O_2^-) are simultaneously produced in large amounts, which rapidly generate the highly reactive peroxynitrite anion. Peroxynitrite inhibits IDO1 *via* the nitration of critical tyrosine residues (Tyr15, Tyr345, and Tyr353), without affecting IDO1 protein level (15, 16).

Hypoxia also leads to reduced IDO1 expression and, therefore, reduced Kyn production (17). Low IDO1 activity during the hypoxic phase promotes the activation of immune cells (18); however, hypoxic conditions also augment the secretion of effector CD4(+) T-cell cytokines, especially IFN-gamma, which in turn upregulates IDO1 mRNA expression (19).

At the protein level, IDO1 is mainly regulated by proteasomal degradation in response to immunogenic stimuli. Suppressor of cytokine signaling 3 (SOCS3), for instance, binds to IDO1 under inflammatory conditions and targets the IDO1/SOCS3 complex for polyubiquitination and proteasomal digestion. IDO1 degradation converts tolerogenic dendritic cells (DCs) into immunogenic cells and, therefore, a prerequisite for normal dendritic-cell function (20). Activated AhR is another component of the ubiquitin/proteasome system that contributes to the regulatory proteolysis of IDO1 and other proteins (8) that affect IDO1's half-life.

Indoleamine-2,3-dioxygenase 1 contains two phosphorylatable tyrosine residues (Y115 and Y253) that modulate its enzymatic activity (21). Phosphorylation of any of these residues results in conformational changes in IDO1 and blocks the catalytic activity of IDO1 (22). In addition to the regulation of catalytic activity, these motifs act as docking sites for various molecular partners that either prolong or shorten IDO1's half-life and maintain its immunoregulatory effects or stimulate inflammatory responses, respectively (8). For example, IL-6 triggers the phosphorylation of the Y253 residue of IDO1 that recruits the ECS (Elongin-Cullin-SOCS) E3 ligase complex, targeting IDO/SOCS3 for proteasomal degradation (23). On the contrary, phosphorylation of the Y115 residue allows the binding of SH2 domain tyrosine phosphatases SHP1 and SHP2, while SOCS3 becomes excluded. Hence, the TGF-β/Fyn/SHP axis activates the non-canonical NF-κB pathway that upregulates IDO1 production. Recently, ligand-bound AhR and Arginase 1 have been also shown to promote IDO1 phosphorylation through Src kinases, activating the signaling function of IDO1 that leads to *de novo* synthesis of the enzyme *via* TGF-β production (8, 24, 25).

The Physiological and Pathological Function of IDO1

The IDO1 pathway was originally described as an innate immune mechanism that defended the host organism against infections (26). The immunoprotective role of IDO1 was directly

supported by the anti-pathogen effect of Trp metabolites (L-Kyn, L-hydroxykynurenine, 3-hydroxyanthranilic acid, quinolinic acid, picolinic acid) that prevented the proliferation and spread of intracellular pathogens (4, 27). Subsequent studies, however, identified tissue macrophages producing high levels of IDO1 upon IFN-γ stimulation that strongly inhibited the proliferation of effector T cells (28). It was also shown that accumulating Trp metabolites induced the differentiation of regulatory T cells and apoptosis of effector T cells that gave rise to immunosuppression (4, 26). L-Kyn is particularly toxic to lymphocytes (2) and induces the differentiation of regulatory T cells *via* AhR binding (29). L-hydroxykynurenine aids the suppression of CD4+ T cells and promotes the action of regulatory T cells (30). 3-hydroxyanthranilic acid modulates the immune functions of monocytes and lymphocytes, induces the apoptosis of effector T cells, and promotes the proliferation of regulatory T cells. Quinolinic acid stimulates the cell cycle of cancer cells and contributes to the acquisition of multidrug resistance against chemotherapeutic agents (29). Picolinic acid inhibits effector T cell proliferation (31). Later, it has become clear that the balance between the immunoprotective and immunosuppressive roles of IDO1 and Trp metabolites is tightly controlled by the stoichiometry of available local factors (e.g., IL-6, IL-12, CD40, IFN-γ, CTLA4, Foxo3a, IL-10, and PD-1) (26, 32). The resultant effect of these local activities modulates IDO1 expression and helps maintain global immune homeostasis and peripheral immune tolerance.

There are many pathologic diseases that are associated with increased IDO1 activity, including atherosclerosis, obesity, autoimmunity, major infections (e.g., community-acquired pneumonia, tuberculosis, listeriosis, influenza, HBV, HCV, HIV, sepsis), rejection of organ transplants, and cancer (2, 27). Originally, IDO1 has been considered as an anti-cancer molecule, proposing that increased IDO1 activity of antigen-presenting cells depletes the essential amino acid Trp from tumor cells and inhibits their growth. However, with the discovery of IDO1-mediated immunosuppressive functions, the pro-cancer activity of the enzyme has been recognized (33). IDO1 is overexpressed in more than 50% of tumors (34) that utilize IDO1-associated immunosuppressive mechanisms to promote their spread and survival (35). In cervical cancer, for instance, IDO1 shows a significantly higher mRNA transcription and protein expression level than in normal cervix, and also in comparison to other cancers (36). The extent of IDO overexpression also depends on the type of tumor and risk factors that reach patients in their life. For instance, oral squamous cell carcinoma (OSCC) was compared in never-smokers and never-drinkers (NSND) with smoker and drinker (SD) patients. In NSND patients suffering from OSCC, expression of IDO1 was significantly higher than in SDs (37).

Indoleamine-2,3-dioxygenase1 production is elevated upon (i) IFN-γ production of effector T cells (2), (ii) inflammatory cytokine production of innate immune cells (6, 38), (iii) IL-10 and IL-27 stimulation (39), (iv) CTLA4 expression on Treg cells causing increased IDO1 secretion by dendritic cells (DCs) (40), and (v) TGF-β, IL-10 and adenosine production of Treg and other immunosuppressive cells (40–42), (vi) cyclooxygenase-2 (COX-2) and prostaglandin E2 (PGE2) stimuli that are mediated through the PKC, PI3K, and MAPK pathways (several types of tumors carry PI3K or MAPK oncogenic mutations leading to constitutive IDO1 expression.) (43).

The mechanism of IDO1-elicited immunosuppression is not fully understood; however, increased IDO1 and Kyn levels are known to inhibit natural killer (NK) cell function (44, 45), prevent the activation of effector T cells, stimulate the activation of Treg cells (35, 46) and the differentiation of tolerogenic DCs (47), and promote the expansion and activation of myeloid-derived suppressor cells (48). Furthermore, Trp depletion inhibits mechanistic target of rapamycin complex 1 that stimulates T cell apoptosis and antigen-presenting cell-mediated inflammation (1, 49).

Paradoxically, the adaptive and innate immune systems that primarily protect against cancer development drive the formation of the highly aggressive and difficult-to-treat IDO1+ tumors. The genetic and biochemical characteristics of the emerging tumor is established by the process of "immunoediting" that prevents and promotes tumor formation at the same time, involving three consecutive stages called "elimination,", "equilibrium," and "escape" (50) (**Figure 1C**). In the first phase ("elimination"), most transformed cells are efficiently recognized and destroyed by the action of effector cells as NK and T cells (50). At this stage, IDO1 is produced at low levels within the tumor microenvironment and inhibits tumor proliferation (33). In the "equilibrium" phase, surviving tumor cells become "edited" by the continuous attack of the immune system and accumulate mutations (51). In the "escape" phase, IDO1 is produced in large quantities by tumor cells and tolerogenic immune cells that are recruited to the tumor microenvironment (52, 53). Increased IDO1 activity leads to elevated Kyn production that prevents effector T cell (35, 46) and NK cell functions (44, 45). In parallel, IDO1 induces the activation and expansion of Treg cells (26), DCs (47), and MDCSs (48) that further suppress the function of antitumor T cells. These mechanisms collectively establish an immunosuppressive tumor microenvironment that supports tumor growth. IDO1 positivity is strongly associated with multidrug resistance of tumors and inversely correlates with patient survival (54). Therefore, timely diagnosis and therapeutic correction of IDO+ tumors are of crucial importance to prevent clinical manifestation.

IDO1 in Cancer Diagnostics and Therapy

Indoleamine-2,3-dioxygenase 1 overexpression increases the relative concentration of Kyn compared to Trp, hence Kyn/Trp ratio can be used as a prognostic clinico-pathological marker to monitor cancer invasiveness and progression. Accordingly, increased systemic Kyn/Trp ratio and elevated IDO1 activity have been associated with poor prognosis and low survival of patients diagnosed with cervical cancer and glioblastoma multiforme (55, 56). For the sensitive detection of Trp metabolites in IDO1$^+$ tumor tissues, a wealth of Trp-based radiotracers has been developed for positron emission tomography imaging. Radioactive Trp analogs as α-[^{11}C]methyl-L-tryptophan (AMT) and L- and D-1-[^{18}F]fluoroethyl-tryptophan provide useful information about response to immunotherapy, but they are also crucial for the preclinical and clinical validation of novel IDO1 inhibitors (57, 58).

Protein expression of IDO1 was found to be high in a number of tumor samples (36, 56); therefore, IDO1 may be a relevant therapeutic target to abrogate immune suppression (59).

dc_1977_21

TABLE 1 | Clinical trials of indoleamine-2,3-dioxygenase (IDO) inhibitors in cancer therapy.

Agent	Tumor type	NCT number	Study phase	Recruitment status	(Estimated) primary completion date
IDO peptide vaccine	• Non-small cell lung cancer (NSCLC)	NCT01219348	Phase 1	Completed	August 2012
	• Metastatic malignant melanoma	NCT02077114	Phase 1	Completed	September 2014
		NCT01543464	Phase 2	Terminated	September 2016
		NCT03047928	Phase 1 Phase 2	Not yet recruiting	1 April, 2019
Indoximod (1-methyl-D-tryptophan; D-1MT; NSC-721782)	• Unspecified adult solid tumors	NCT00567931	Phase 1	Completed	July 2012
	• Breast cancer • Lung cancer • Melanoma • Pancreatic cancer • Solid tumors	NCT00739609	Phase 1	Terminated	October 2012
	• Metastatic breast cancer	NCT01302821		Withdrawn	December 2014
		NCT01792050	Phase 2	Active, not recruiting	December 2016
	• Metastatic melanoma	NCT02073123	Phase 1 Phase 2	Recruiting	December 2016
	• Glioblastoma multiforme • Glioma • Gliosarcoma • Malignant brain tumor	NCT02052648	Phase 1 Phase 2	Recruiting	December 2016
		NCT02502708	Phase 1	Recruiting	July 2017
	• Non-small cell lung cancer (NSCLC)	NCT02460367	Phase 1 Phase 2	Recruiting	June 2017
	• Metastatic pancreatic adenocarcinoma • Metastatic pancreatic cancer	NCT02077881	Phase 1 Phase 2	Recruiting	July 2017
	• Acute myeloid leukemia	NCT02835729	Phase 1 Phase 2	Recruiting	July 2018
Epacadostat (INCB024360, 4-amino-1,2,5-oxadizaole-3-carboximidamide)	• Advanced malignancies	NCT01195311	Phase 1	Completed	May 2013
	• Myelodysplastic syndromes (MDS)	NCT01822691	Phase 2	Completed	January 2015
	• Epithelial ovarian cancer • Fallopian tube cancer • Primary peritoneal cancer	NCT01685255	Phase 2	Terminated	23 October, 2014
		NCT02118285	Phase 1	Completed	12 November, 2015
		NCT02042430		Active, not recruiting	31 March, 2016
		NCT01982487	Phase 1 Phase 2	Withdrawn	September 2017
		NCT02166905	Phase 1 Phase 2	Recruiting	12 February, 2018
		NCT02785250	Phase 1	Recruiting	May 2018
		NCT02575807	Phase 1 Phase 2	Recruiting	December 2018
	• Mucosal melanoma • Skin melanoma • Uveal melanoma	NCT01961115	Phase 2	Active, not recruiting	31 October, 2016
		NCT01604889	Phase 1 Phase 2	Terminated	27 December, 2016
	• Gastrointestinal stromal tumors	NCT03291054	Phase 2	Not yet recruiting	September 2019
	• Recurrent/metastatic endometrial carcinoma	NCT03310567	Phase 2	Not yet recruiting	30 January, 2020
	• Squamous cell carcinoma of the head and neck	NCT03325465	Phase 2	Not yet recruiting	June 2020
	• Advanced solid tumors	NCT02559492	Phase 1	Active, not recruiting	December 2017
		NCT03085914	Phase 1 Phase 2	Recruiting	April 2021
		NCT02959437	Phase 1 Phase 2	Recruiting	September 2021
	• Metastatic pancreatic adenocarcinoma	NCT03006302	Phase 2	Not yet recruiting	February 2021
	• Metastatic non-small cell lung cancer (NSCLC)	NCT03322540	Phase 3	Not yet recruiting	17 June, 2022
		NCT03322566	Phase 3	Not yet recruiting	26 October, 2022
	• Renal cell carcinoma	NCT03260894	Phase 3	Not yet recruiting	May 2023

(*Continued*)

dc_1977_21

TABLE 1 | Continued

Agent	Tumor type	NCT number	Study phase	Recruitment status	(Estimated) primary completion date
GDC-0919	• Advanced solid tumors	NCT02048709	Phase 1	Completed	February 2016
HTI-1090 (SHR9146)	• Advanced solid tumors	NCT03208959	Phase 1	Not yet recruiting	1 April, 2018
PF-06840003	• Oligodendroglioma • Astrocytoma • Malignant glioma	NCT02764151	Phase 1	Recruiting	30 April, 2018
NLG802	• Advanced solid tumors	NCT03164603	Phase 1	Recruiting	May 2018
BMS-986205	• Advanced cancer	NCT03335540	Phase 1	Not yet recruiting	14 March, 2021

Clinical trials were identified on the website: https://clinicaltrials.gov/ct2/results?cond=&term=IDO&cntry1=&state1=&recrs= as of 9 November, 2017.

Currently, several IDO inhibitors undergo clinical evaluation (60) and many of them are now in Phase II clinical trials (**Table 1**). Most inhibitors were designed to block the Kyn pathway (e.g., epacadostat, indoximod, GDC-0919, and an IDO1 peptide-based vaccine; **Figure 1A**) that suspends immunosuppression (1). Some of the tested compounds are used alone, or in combination with immunotherapy (CTLA4, PD-1 blockade), chemotherapy, adoptive transfer, COX-2 inhibitors (e.g., celecoxib), membrane-associated PGE2 synthase inhibitors (e.g., MF63), PGE2 receptor (EP4) competitive antagonists (e.g., GW627368X), and PI3K inhibitors (43, 59). The latter combinative therapies are intended to improve the inhibition of local immunosuppression around tumor tissues and to enhance tumor eradication (61).

Epacadostat (INCB024360) and indoximod (NLG8189 or 1-methyl-D-tryptophan) are the most common IDO inhibitors that are well tolerated and show obvious beneficial effects in cancer therapies; however, both have some major side effects. Epacadostat showed grade 3 and 4 adverse effects in patients with advanced malignancies, most frequently abdominal pain, hypokalemia, fatigue, and some minor effects involving nausea, decreased appetite, vomiting, constipation, diarrhea, dyspnea, back pain, and cough (62). Indoximod treatment also showed some major toxicities in a dose-escalation study, involving grade 1 fatigue and grade 2 hypophysitis (63). In combination with the microtubular poison docetaxel, the most common adverse effects of indoximod were fatigue, anemia, hyperglycemia, infection, and nausea (62, 63).

Combination of IFN-γ treatment with IDO1 inhibitors is a promising new cancer immunotherapeutic strategy that effectively enhances antitumor immunity and eliminates TRCs (i.e., stem cell-like cancer cells that are self-renewing, highly tumorigenic, and can repopulate tumors). In clinical practice, administration of IFN-γ with IDO1 inhibitors is the only oncolytic therapy that significantly disrupts TRCs. IFN-γ induces the entry of TRCs into dormancy, while IDO1 inhibitor-elicited immunosuppression allows effector T cells and NK cells to attack dormant TRCs (64). Monotherapies with IDO inhibitors or other combinative treatments usually terminate with failure because of immune evasion of TRCs, which leads to metastasis formation, tumor recurrence and multidrug resistance (64).

Mutational load/neoantigen-burden, basal level of tumor infiltrating T cells (TILs), differential expression of immune-checkpoint genes within the tumor tissue are important biomarkers that help predict the tumor's predisposition toward immune-checkpoint inhibitors (ICIs) targeting IDO1, CTLA4, or PD-1 and increase the clinical success of immunotherapies. Most ICI-responsive cancers (e.g., lung and bladder cancers, melanoma) were associated with intrinsically high TIL numbers and high mutational load/neoantigen-burden, while other cancers (e.g., glioblastoma) were predicted to be ICI-resistant based on their biomarker profiles (37, 65).

Beyond the application of biomarkers and chemical inhibitors, IDO1 can be genetically targeted by genome editing tools that offer new therapeutic opportunities for cancer patients. In animal studies, genetic inhibition of IDO1 expression reactivated the antitumor immune response against IDO^+ cancer cells and inhibited tumor growth (63). The shIDO-ST treatment, for instance, is based on a *Salmonella typhimurium* (ST) vector that codes for a small hairpin RNA targeting IDO1 (shIDO) (66). Intravenously injected shIDO-ST accumulated in the tumor tissues of mice, causing IDO1 silencing and concomitant infiltration and activation of polymorphonuclear neutrophil granulocytes (PMNs). PMNs produced reactive oxygen species that established a highly toxic microenvironment for tumor cell growth (67). A recent genomic sequencing study that combined large-scale tumor exome analysis with MHC I class prediction revealed a strong positive correlation between IDO1 expression, mutational burden, and neoantigen load in cervical cancers (36). The above studies collectively identify IDO1 as an attractive target to increase the effectiveness of cancer immunotherapies.

Conclusion and Outlook

The mechanism of "cancer immunoediting" is the direct consequence of a T cell-dependent immunoselection process that drives the formation of IDO1+ tumors. The action of a competent immune system is, therefore, determinative for the acquisition of cancer immunogenicity. Important posttranslational control mechanisms affect the activity and half-life of IDO1 (e.g., NO, hypoxia, proteasomal degradation, phosphorylation) that should be considered in terms of cancer therapy. For instance, IDO1 inhibitors could be administered as co-therapeutic agents in the presence of redox regulators, IFN-γ, or anti-IL-6. Combining IDO1 drugs with the inhibition of specific transcription factors regulating IDO1 activity (e.g., AhR) may also improve the effectiveness and specificity of chemotherapies. Current genome

dc_1977_21

editing and exome sequencing technologies offer promising new strategies to identify novel tumor-specific mutational antigens and thus expand the repertoire of tumor-specific immunotherapies.

AUTHOR CONTRIBUTIONS

All authors participated in the writing of this manuscript and agreed to be accountable for the content of the work.

FUNDING

LS received grants from the Hungarian Academy of Sciences (Lendület programme, LP2015-9/2015), NKFIH-ERC-HU-117670, and GINOP-2.3.2-15-2016-00024. GH was supported by the GINOP-2.3.2-15-2016-00043 project. GK was supported by NKFIH-K-125224. DP and ZS were supported by GINOP-2.3.2-15-2016-00062.

REFERENCES

1. Zhai L, Spranger S, Binder DC, Gritsina G, Lauing KL, Giles FJ, et al. Molecular pathways: targeting IDO1 and Other tryptophan dioxygenases for cancer immunotherapy. *Clin Cancer Res* (2015) 21:5427–33. doi:10.1158/1078-0432.CCR-15-0420
2. Bilir C, Sarisozen C. Indoleamine 2,3-dioxygenase (IDO): only an enzyme or a checkpoint controller? *J Oncol Sci* (2017) 3:52–6. doi:10.1016/j.jons.2017.04.001
3. Yevshin I, Sharipov R, Valeev T, Kel A, Kolpakov FGTRD. A database of transcription factor binding sites identified by ChIP-seq experiments. *Nucleic Acids Res* (2017) 45:D61–7. doi:10.1093/nar/gkw951
4. van Baren N, Van den Eynde BJ. Tryptophan-degrading enzymes in tumoral immune resistance. *Front Immunol* (2015) 6:34. doi:10.3389/fimmu.2015.00034
5. Thul PJ, Akesson L, Wiking M, Mahdessian D, Geladaki A, Ait Blal H, et al. A subcellular map of the human proteome. *Science* (2017) 356:eaal3321. doi:10.1126/science.aal3321
6. Zulfiqar B, Mahroo A, Nasir K, Farooq RK, Jalal N, Rashid MU, et al. Nanomedicine and cancer immunotherapy: focus on indoleamine 2,3-dioxygenase inhibitors. *Onco Targets Ther* (2017) 10:463–76. doi:10.2147/OTT.S119362
7. Vogel CFA, Wu D, Goth SR, Baek J, Lollies A, Domhardt R, et al. Aryl hydrocarbon receptor signaling regulates NF-κB RelB activation during dendritic-cell differentiation. *Immunol Cell Biol* (2013) 91:568–75. doi:10.1038/icb.2013.43
8. Pallotta MT, Fallarino F, Matino D, Macchiarulo A, Orabona C. AhR-mediated, non-genomic modulation of IDO1 function. *Front Immunol* (2014) 5:497. doi:10.3389/fimmu.2014.00497
9. Dixon JR, Jung I, Selvaraj S, Shen Y, Antosiewicz-Bourget JE, Lee AY, et al. Chromatin architecture reorganization during stem cell differentiation. *Nature* (2015) 518:331–6. doi:10.1038/nature14222
10. Thomas SR, Mohr D, Stocker R. Nitric oxide inhibits indoleamine 2, 3-dioxygenase activity in interferon-gamma primed mononuclear phagocytes. *J Biol Chem* (1994) 269:14457–64.
11. Thomas SR, Terentis AC, Cai H, Takikawa O, Levina A, Lay PA, et al. Post-translational regulation of human indoleamine 2,3-dioxygenase activity by nitric oxide. *J Biol Chem* (2007) 282:23778–87. doi:10.1074/jbc.M700669200
12. Hucke C, MacKenzie CR, Adjogble KDZ, Takikawa O, Däubener W. Nitric oxide-mediated regulation of gamma interferon-induced bacteriostasis: inhibition and degradation of human indoleamine 2,3-dioxygenase. *Infect Immun* (2004) 72:2723–30. doi:10.1128/IAI.72.5.2723-2730.2004
13. Genes P, Liu SF, Ye X, Malik AB. Inhibition of NF-kB activation by pyrrolidine dithiocarbamate prevents in vivo expression of. *Circulation* (1999) 100:1330–7. doi:10.1161/01.CIR.100.12.1330
14. Thomas SR, Salahifar H, Mashima R, Hunt NH, Richardson DR, Stocker R. Antioxidants inhibit indoleamine 2,3-dioxygenase in IFN-gamma-activated human macrophages: posttranslational regulation by pyrrolidine dithiocarbamate. *J Immunol* (2001) 166:6332–40. doi:10.4049/jimmunol.166.10.6332
15. Fujigaki H, Saito K, Lin F, Fujigaki S, Takahashi K, Martin BM, et al. Nitration and inactivation of IDO by peroxynitrite. *J Immunol* (2006) 176:372–9. doi:10.4049/jimmunol.176.1.372
16. Grohmann U, Fallarino F, Bianchi R, Orabona C, Vacca C, Fioretti MC, et al. A defect in tryptophan catabolism impairs tolerance in nonobese diabetic mice. *J Exp Med* (2003) 198:153–60. doi:10.1084/jem.20030633
17. Schmidt SK, Ebel S, Keil E, Woite C, Ernst JF, Benzin AE, et al. Regulation of IDO activity by oxygen supply: inhibitory effects on antimicrobial and immunoregulatory functions. *PLoS One* (2013) 8:e63301. doi:10.1371/journal.pone.0063301
18. Herbert A, Ng H, Jessup W, Kockx M, Cartland S, Thomas SR, et al. Hypoxia regulates the production and activity of glucose transporter-1 and indoleamine 2,3-dioxygenase in monocyte-derived endothelial-like cells: possible relevance to infantile haemangioma pathogenesis. *Br J Dermatol* (2011) 164:308–15. doi:10.1111/j.1365-2133.2010.10086.x
19. Roman J, Rangasamy T, Guo J, Sugunan S, Meednu N, Packirisamy G, et al. T-cell activation under hypoxic conditions enhances IFN-γ secretion. *Am J Respir Cell Mol Biol* (2010) 42:123–8. doi:10.1165/rcmb.2008-0139OC
20. Orabona C, Pallotta MT, Volpi C, Fallarino F, Vacca C, Bianchi R, et al. SOCS3 drives proteasomal degradation of indoleamine 2, 3-dioxygenase (IDO) and antagonizes IDO-dependent tolerogenesis. *Proc Natl Acad Sci U S A* (2008) 105:20828–33. doi:10.1073/pnas.0810278105
21. Maria P, CIriana O, Claudia V, Ursula G, Paolo P, Francesca F. Proteasomal degradation of indoleamine 2,3-dioxygenase in CD8+ dendritic cells is mediated by suppressor of cytokine signaling 3 (SOCS3). *Int J Tryptophan Res* (2010) 3:91–7. doi:10.4137/IJTR.S3971
22. Albini E, Rosini V, Gargaro M, Mondanelli G, Belladonna ML, Pallotta MT, et al. Distinct roles of immunoreceptor tyrosine-based motifs in immunosuppressive indoleamine 2,3-dioxygenase 1. *J Cell Mol Med* (2017) 21:165–76. doi:10.1111/jcmm.12954
23. Orabona C, Pallotta M, Grohmann U. Different partners, opposite outcomes: a new perspective of the immunobiology of indoleamine 2,3-dioxygenase. *Mol Med* (2012) 18:834–42. doi:10.2119/molmed.2012.00029
24. Mondanelli G, Bianchi R, Pallotta MT, Orabona C, Albini E, Iacono A, et al. A relay pathway between arginine and tryptophan metabolism confers immunosuppressive properties on dendritic cells. *Immunity* (2017) 46:233–44. doi:10.1016/j.immuni.2017.01.005
25. Bessede A, Gargaro M, Pallotta MT, Matino D, Brunacci C, Bicciato S, et al. Aryl hydrocarbon receptor control of a disease tolerance defense pathway. *Nature* (2015) 511:184–90. doi:10.1038/nature13323.Aryl
26. Munn DH, Mellor AL. Indoleamine 2,3 dioxygenase and metabolic control of immune responses. *Trends Immunol* (2013) 34:137–43. doi:10.1016/j.it.2012.10.001
27. Schmidt SV, Schultze JL. New insights into IDO biology in bacterial and viral infections. *Front Immunol* (2014) 5:384. doi:10.3389/fimmu.2014.00384
28. Munn DH, Zhou M, Attwood JT, Bondarev I, Conway SJ, Marshall B, et al. Prevention of allogeneic fetal rejection by tryptophan catabolism. *Science* (1998) 281:1191–3. doi:10.1126/science.281.5380.1191
29. Heng B, Lim CK, Lovejoy DB, Bessede A, Gluch L, Guillemin GJ. Understanding the role of the kynurenine pathway in human breast cancer immunobiology. *Oncotarget* (2016) 7:6506–20. doi:10.18632/oncotarget.6467
30. Zaher SS, Germain C, Fu H, Larkin DFP, George AJT. 3-hydroxykynurenine suppresses CD4+ T-cell proliferation, induces T-regulatory-cell development, and prolongs corneal allograft survival. *Invest Ophthalmol Vis Sci* (2011) 52:2640–8. doi:10.1167/iovs.10-5793
31. Prodinger J, Loacker LJ, Schmidt RLJ, Ratzinger F, Greiner G, Witzeneder N, et al. The tryptophan metabolite picolinic acid suppresses proliferation and metabolic activity of CD4+ T cells and inhibits c-Myc activation. *J Leukoc Biol* (2016) 99:583–94. doi:10.1189/jlb.3A0315-135R
32. Grohmann U, Orabona C, Fallarino F, Vacca C, Calcinaro F, Falorni A, et al. CTLA-4-Ig regulates tryptophan catabolism in vivo. *Nat Immunol* (2002) 3:1097–101. doi:10.1038/ni846

dc_1977_21

33. Katz JB, Muller AJ, Prendergast GC. Indoleamine 2,3-dioxygenase in T-cell tolerance and tumoral immune escape. *Immunol Rev* (2008) 222:206–21. doi:10.1111/j.1600-065X.2008.00610.x
34. Löb S, Königsrainer A, Zieker D, Brücher BLDM, Rammensee HG, Opelz G, et al. IDO1 and IDO2 are expressed in human tumors: levo- but not dextro-1-methyl tryptophan inhibits tryptophan catabolism. *Cancer Immunol Immunother* (2009) 58:153–7. doi:10.1007/s00262-008-0513-6
35. Godin-Ethier J, Hanafi L-A, Piccirillo CA, Lapointe R. Indoleamine 2,3-dioxygenase expression in human cancers: clinical and immunologic perspectives. *Clin Cancer Res* (2011) 17:6985–91. doi:10.1158/1078-0432.CCR-11-1331
36. Qin Y, Ekmekcioglu S, Forget M-A, Szekvolgyi L, Hwu P, Grimm EA, et al. Cervical cancer neoantigen landscape and immune activity is associated with human papillomavirus master regulators. *Front Immunol* (2017) 8:689. doi:10.3389/fimmu.2017.00689
37. Foy JP, Bertolus C, Michallet MC, Deneuve S, Incitti R, Bendriss-Vermare N, et al. The immune microenvironment of HPV-negative oral squamous cell carcinoma from never-smokers and never-drinkers patients suggests higher clinical benefit of IDO1 and PD1/PD-L1 blockade. *Ann Oncol* (2017) 28:1934–41. doi:10.1093/annonc/mdx210
38. Rovira Gonzalez YI, Lynch PJ, Thompson EE, Stultz BG, Hursh DA. *In vitro* cytokine licensing induces persistent permissive chromatin at the Indoleamine 2,3-dioxygenase promoter. *Cytotherapy* (2016) 18:1114–28. doi:10.1016/j.jcyt.2016.05.017
39. Carbotti G, Barisione G, Airoldi I, Mezzanzanica D, Bagnoli M, Ferrero S, et al. IL-27 induces the expression of IDO and PD-L1 in human cancer cells. *Oncotarget* (2015) 6:43267–80. doi:10.18632/oncotarget.6530
40. Belladonna ML, Orabona C, Grohmann U, Puccetti P. TGF-β and kynurenines as the key to infectious tolerance. *Trends Mol Med* (2009) 15:41–9. doi:10.1016/j.molmed.2008.11.006
41. Yanagawa Y, Iwabuchi K, Onoé K. Co-operative action of interleukin-10 and interferon-γ to regulate dendritic cell functions. *Immunology* (2009) 127:345–53. doi:10.1111/j.1365-2567.2008.02986.x
42. Novitskiy SV, Ryzhov S, Zaynagetdinov R, Goldstein AE, Huang Y, Tikhomirov OY, et al. Adenosine receptors in regulation of dendritic cell differentiation and function. *Blood* (2008) 112:1822–31. doi:10.1182/blood-2008-02-136325
43. Hennequart M, Pilotte L, Cane S, Hoffmann D, Stroobant V, De Plaen E, et al. Constitutive IDO1 expression in human tumors is driven by cyclooxygenase-2 and mediates intrinsic immune resistance. *Cancer Immunol Res* (2017) 5:695–709. doi:10.1158/2326-6066.CIR-16-0400
44. Pietra G, Vitale M, Moretta L, Mingari MC. How melanoma cells inactivate NK cells. *Oncoimmunology* (2012) 1:974–5. doi:10.4161/onci.20405
45. Wang D, Saga Y, Mizukami H, Sato N, Nonaka H, Fujiwara H, et al. Indoleamine-2,3-dioxygenase, an immunosuppressive enzyme that inhibits natural killer cell function, as a useful target for ovarian cancer therapy. *Int J Oncol* (2012) 40:929–34. doi:10.3892/ijo.2011.1295
46. Mellor AL, Keskin DB, Johnson T, Chandler P, Munn DH. Cells expressing indoleamine 2,3-dioxygenase inhibit T cell responses. *J Immunol* (2002) 168:3771–6. doi:10.4049/jimmunol.168.8.3771
47. Li Q, Harden JL, Anderson CD, Egilmez NK. Tolerogenic phenotype of IFN-γ-induced IDO$^+$ dendritic cells is maintained via an autocrine IDO–kynurenine/AhR–IDO loop. *J Immunol* (2016) 197:962–70. doi:10.4049/jimmunol.1502615
48. Holmgaard RB, Zamarin D, Li Y, Gasmi B, Munn DH, Allison JP, et al. Tumor-expressed IDO recruits and activates MDSCs in a treg-dependent manner. *Cell Rep* (2015) 13:412–24. doi:10.1016/j.celrep.2015.08.077
49. Powell JJD, Pollizzi KNK, Heikamp EB, Horton MR. Regulation of immune responses by mTOR. *Annu Rev immunol* (2012) 30:39–68. doi:10.1146/annurev-immunol-020711-075024.Regulation
50. Kim R, Emi M, Tanabe K. Cancer immunoediting: from immune surveillance to immune escape. *Immunology* (2007) 121:1–17. doi:10.1111/j.1365-2567.2007.02587.x
51. Dunn GP, Old LJ, Schreiber RD. The three Es of cancer immunoediting. *Annu Rev Immunol* (2004) 22:329–60. doi:10.1146/annurev.immunol.22.012703.104803
52. Shou D, Liang W, Song Z, Yin J, Sun Q, Gong W. Suppressive role of myeloid-derived suppressor cells (MDSCs) in the microenvironment of breast cancer and targeted immunotherapies. *Oncotarget* (2016) 7:64505–11. doi:10.18632/oncotarget.11352
53. Zhao Q, Kuang D-M, Wu Y, Xiao X, Li X-F, Li T-J, et al. Activated CD69+ T cells foster immune privilege by regulating IDO expression in tumor-associated macrophages. *J Immunol* (2012) 188:1117–24. doi:10.4049/jimmunol.1100164
54. Okamoto A, Nikaido T, Ochiai K, Takakura S, Saito M, Aoki Y, et al. Indoleamine 2,3-dioxygenase serves as a marker of poor prognosis in gene expression profiles of serous ovarian cancer cells. *Clin Cancer Res* (2005) 11:6030–9. doi:10.1158/1078-0432.CCR-04-2671
55. Hascitha J, Priya R, Jayavelu S, Dhandapani H, Selvaluxmy G, Sunder Singh S, et al. Analysis of kynurenine/tryptophan ratio and expression of IDO1 and 2 mRNA in tumour tissue of cervical cancer patients. *Clin Biochem* (2016) 49:919–24. doi:10.1016/j.clinbiochem.2016.04.008
56. Zhai L, Ladomersky E, Lauing KL, Wu M, Genet M, Gritsina G, et al. Infiltrating T cells increase IDO1 expression in glioblastoma and contribute to decreased patient survival. *Clin Cancer Res* (2017) 23:6650–60. doi:10.1158/1078-0432.CCR-17-0120
57. Xin Y, Cai H. Improved radiosynthesis and biological evaluations of L- and D-1-[18F]fluoroethyl-tryptophan for PET imaging of IDO-mediated kynurenine pathway of tryptophan metabolism. *Mol Imaging Biol* (2017) 19:589–98. doi:10.1007/s11307-016-1024-z
58. Henrottin J, Lemaire C, Egrise D, Zervosen A, van Den Eynde B, Plenevaux A, et al. Fully automated radiosynthesis of N1-[18F]fluoroethyl-tryptophan and study of its biological activity as a new potential substrate for indoleamine 2,3-dioxygenase PET imaging. *Nucl Med Biol* (2016) 43:379–89. doi:10.1016/j.nucmedbio.2016.03.001
59. Munn DH, Mellor AL. IDO in the tumor microenvironment: inflammation, counter-regulation, and tolerance. *Trends Immunol* (2016) 37:193–207. doi:10.1016/j.it.2016.01.002
60. Routy JP, Routy B, Graziani GM, Mehraj V. The kynurenine pathway is a double-edged sword in immune-privileged sites and in cancer: implications for immunotherapy. *Int J Tryptophan Res* (2016) 9:67–77. doi:10.4137/IJTR.S38355
61. Kozłowska A, Mackiewicz J, Mackiewicz A. Therapeutic gene modified cell based cancer vaccines. *Gene* (2013) 525:200–7. doi:10.1016/j.gene.2013.03.056
62. Brochez L, Chevolet I, Kruse V. The rationale of indoleamine 2,3-dioxygenase inhibition for cancer therapy. *Eur J Cancer* (2017) 76:167–82. doi:10.1016/j.ejca.2017.01.011
63. Vacchelli E, Aranda F, Eggermont A, Sautès-Fridman C, Tartour E, Kennedy EP, et al. Trial watch: IDO inhibitors in cancer therapy. *Oncoimmunology* (2014) 3:e957994. doi:10.4161/21624011.2014.957994
64. Liu Y, Liang X, Yin X, Lv J, Tang K, Ma J, et al. Blockade of IDO-kynurenine-AhR metabolic circuitry abrogates IFN-γ-induced immunologic dormancy of tumor-repopulating cells. *Nat Commun* (2017) 8:15207. doi:10.1038/ncomms15207
65. Garg AD, Vandenberk L, Van Woensel M, Belmans J, Schaaf M, Boon L, et al. Preclinical efficacy of immune-checkpoint monotherapy does not recapitulate corresponding biomarkers-based clinical predictions in glioblastoma. *Oncoimmunology* (2017) 6:e1295903. doi:10.1080/2162402X.2017.1295903
66. Blache CA, Manuel ER, Kaltcheva TI, Wong AN, Ellenhorn JDI, Blazar BR, et al. Systemic delivery of *Salmonella typhimurium* transformed with IDO shRNA enhances intratumoral vector colonization and suppresses tumor growth. *Cancer Res* (2012) 72:6447–56. doi:10.1158/0008-5472.CAN-12-0193
67. Manuel ER, Diamond DJ. A road less traveled paved by IDO silencing: harnessing the antitumor activity of neutrophils. *Oncoimmunology* (2013) 2:e23322. doi:10.4161/onci.23322

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2018 Hornyák, Dobos, Koncz, Karányi, Páll, Szabó, Halmos and Székvölgyi. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.