

# MTA DOKTORI ÉRTEKEZÉS TÉZISEI

## A VIRÁLIS TRANZKRIPTOM REJTETT KOMPLEXITÁSA - INTEGRATÍV GENOMIKAI MEGKÖZELÍTÉSEK –



**TOMBÁ CZ DÓRA**

SZEGEDI TUDOMÁNYEGYETEM  
SZENT-GYÖRGYI ALBERT ORVOSTUDOMÁNYI KAR  
ORVOSI BIOLÓGIAI INTÉZET

SZEGED, 2024

<b>Bevezetés</b>	<b>3</b>
<b>Célkitűzések</b>	<b>6</b>
<b>Eredmények és értékelés</b>	<b>7</b>
<i>Technológiai újítások: „wetlab”, bioinformatika</i>	7
<i>Vírus genomok annotálása: DNS és RNS szekvenciákból</i>	8
<i>Vírus transzkriptomok</i>	9
A virális transzkriptomok strukturális analízise	9
A strukturális komplexitás „háttere”	10
Funkció a strukturális komplexitás mögött	12
Az alfaherpeszvírusok raRNS-ei	13
Dinamikus transzkriptom: a génextpresszió időbeli változásainak elemzése	15
A dinamikus PRV transzkriptom	15
A vírusgén domének időfüggő kifejeződése	16
<i>A gazdasejt transzkriptomok és a vírusfertőzés a gazdagének expressziójára kifejtett hatásának analízise</i>	18
<b>Irodalomjegyzék</b>	<b>25</b>
<b>Tudományometriai adatok</b>	<b>29</b>
<b>Az értekezés alapjául szolgáló közlemények listája</b>	<b>30</b>
<i>Az értekezés alapjául szolgáló saját, teljes terjedelmű közlemények</i>	30
<i>Saját, az értekezéshez kapcsolódó rövid közlemények</i>	32
<i>A dolgozat témájához kapcsolódó egyéb saját, teljes terjedelmű közlemények:</i>	33
PhD tézishez felhasznált, a témához kapcsolódó saját közlemény	34
<i>Egyéb, az értekezéstől független saját publikációk</i>	34
PhD tézishez felhasznált, a jelen dolgozattól független témában megjelent saját publikációk	35
<b>Köszönetnyilvánítás</b>	<b>36</b>
<b>Támogatók</b>	<b>37</b>

## BEVEZETÉS

A vírusok obligát paraziták, amelyek csak gazdasejten belül képesek szaporodni. Minden sejt életformával társulnak, a baktériumoktól az emberekig. A vírusok gazdasejtjeikre gyakorolt hatása változatos, jelentős károkat okozhatnak a humán egészségügyben, valamint a vadon élő és háziállatok között, és a mezőgazdaságban is. A vírusok globális hatását a COVID-19 pandémia világosan szemléltette. Tudományos szempontból a vírusok modellszervezetként szolgálnak egyszerű genomjuk és gyors reprodukciós rátájuk miatt; alkalmazásuk kiterjed génbeviteli<sup>1</sup> és génterápiás<sup>2</sup> technikákra, idegpálya jelölésére<sup>3</sup>, valamint vakcina<sup>4</sup>- és immunterápiás<sup>5</sup> fejlesztésekre. A bakteriofágok<sup>6</sup> használata az antibiotikum-rezisztens baktériumok elleni védekezésben és a bélrendszeri mikrobiom helyreállításában is fontos. A vírusok rendszertana változékony, dinamikus, és új vírusráfedezések, valamint szekvenálási technológiák eredményi következtében folyamatosan frissül. A Baltimore-rendszer alapján a vírusok osztályozása genetikai tulajdonságaik szerint történik, emellett az *International Committee on Taxonomy of Viruses* (ICTV) fogalmaz meg nevezéktani szabályokat. A vírusok osztályozása az ICTV alapján több birodalomra bontva történik<sup>7</sup>.

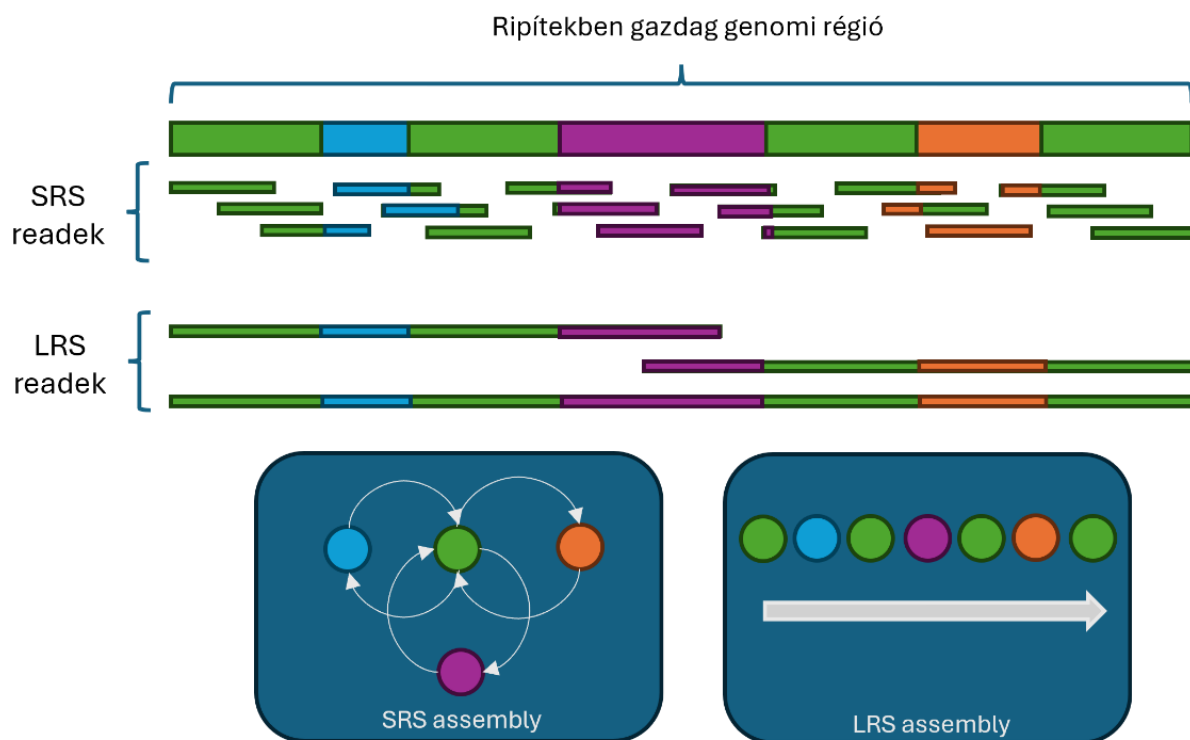
A vírusok számos különböző fertőzési mechanizmust fejlesztettek ki az évmilliók során, mint például a bakteriofágok DNS-injektálása, valamint az eukarióta vírusok endocitózisa és membránfúziója. A sejtmagba jutó DNS vírusok (néhány kivételtől eltekintve) a gazda DNS-szintetizáló mechanizmusait használják, míg az RNS vírusok a citoplazmában maradnak, ahol saját replikációs apparátust alkalmaznak. A génexpresszió szabályozása jelentős változatosságot mutat a különféle vírusszaládok között: a kisebb RNS vírusok örökítőanyaga gyakran mRNS-ként is funkcionál, míg a nagyobb genommal rendelkező DNS vírusok esetén komplex transzkripció mechanizmusok figyelhetők meg. Az egyes gének kinetikai osztályokba sorolhatók, attól függően, hogy a vírus életciklusának melyik stádiumában aktívak. A herpeszvírusoknál (HV) a legkorábban leíródozó (*immediate-early*; IE) gének rendszerint transzkripció faktorokat (TF) kódolnak, a korai (*early*; E) gének általában a DNS replikációhoz szükséges fehérjéket, míg a kései (*late*; L) gének a szerkezeti fehérjéket kódolják<sup>8</sup>.

A DNS szintézis a prokarióta genomokban általában egyetlen replikációs origóból (Ori) indul, míg az eukarióta DNS-ek több tízezer Ori-t tartalmaznak. A vírusok egy vagy néhány Ori-val rendelkeznek, például az 1-es típusú Herpesz simplex vírus (HSV-1) hárommal<sup>9</sup>. Az eukarióta DNS-replikáció az Ori felismerő komplex (*origin recognition complex*; ORC) Ori-hoz való kötődésével kezdődik, amely komplex segíti a repliszóma összeállítását, beleértve a DNS helikázt, polimerázt, és más fehérjéket.

Az úgynevezett replikációs RNS-ek fontos szerepet játszanak a replikáció szabályozásában, többek között az ORC Ori-hoz való verbuválásával és a replikációs fehérjék transzlációjának gátlásával. Bizonyos hosszú nem-kódoló (*long non-coding*; lnc) RNS-ek a herpeszvírusokban már ismertek voltak, de az újabb felfedezések azt sugallják, hogy a replikációs szabályozás összetettebb, mint korábban gondolták.

A DNS és RNS szekvenálás mára alapvető eszközévé vált a modern genomikai kutatásoknak. Az első generációs Sanger-szekvenálás magas költsége és időigényessége miatt helyét fokozatosan átvették a fejlettebb technológiák, úgymint a második generációs rövid-leolvasású (*short read-sequencing*; SRS) és a harmadik generációs hosszú-leolvasású (*long-read sequencing*; LRS) szekvenálási módszerek. Az SRS, különösen az Illumina technológiája, jelentősen felgyorsította a

genomok szekvenálását, bár a módszer nem optimális a komplex és ismétlődő régiók térképezéséhez. Ezzel szemben az LRS, melyet a Pacific Biosciences (PacBio) és az Oxford Nanopore Technologies (ONT) cégek fejlesztettek ki, lehetővé teszi hosszabb szekvenciák és ritka strukturális variációk pontosabb elemzését<sup>10</sup>. Az LRS előnyei közé tartozik a teljes genomok összeállítása, komplex ismétlődő régiók meghatározása (**1. ábra**), ugyanakkor kevésbé pontos és költségesebb, mint az SRS. Az optimális pontosság és lefedettség érdekében az LRS és SRS technológiák kombinálása javasolt a genomikában; a technológiák gyors fejlődése várhatóan tovább fogja javítani a szekvenálás hatékonyságát és költséghatékonyosságát.

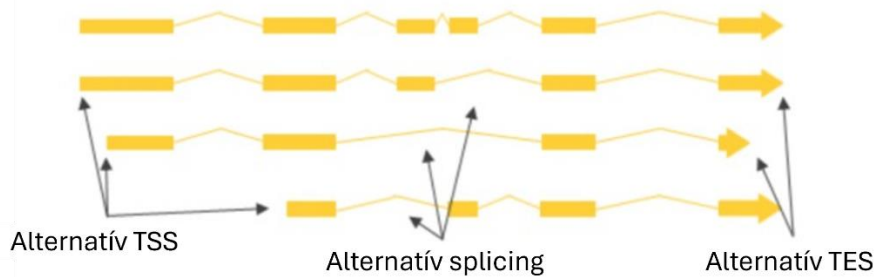


**1. ábra** Egy ismétlődésekben gazdag genomi régió sematikus ábrázolása: SRS (fent), és LRS (lent) módszerrel való szekvenálás. Mivel az SRS módszer által generált leolvasások rövidebbek, mint a repeat egység hossza, így az összeillesztés nem egyértelmű, a bioinformatika nem tudja a genomi régiók egymásutánosságát megmondani (SRS assembly), míg ez az LRS technikával egyértelmű. Zölddel jelöltem az ismétlődő szakaszokat, míg a többi szín egyedi régiókat ábrázol.

A genom szekvenáláson túl az LRS-nek számos előnye van a transzkriptomikában is: a módszer lehetővé teszi az olyan RNS variánsok azonosítását, amelyek a hagyományos SRS technológiával és más genomikai módszerekkel nehezen vagy egyáltalán nem detektálhatók. Az LRS technika alkalmas a transzkripció kezdő- és vég helyeinek (TSS és TES), valamint az alternatív splicing révén létrejövő izoformáknak a pontos azonosítására; azáltal, hogy képes a teljes RNS molekulák szekvenálására. Így az összetett spliceolt izoformák is azonosíthatók (**2. ábra**). Emellett a

technológia segítségével megkülönböztethetők a párhuzamosan átfedő transzkriptek, melyek mennyiségi meghatározása is lehetővé válik ezáltal.

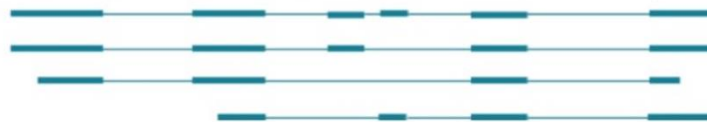
### Transzkript izoformák



### Rövid-read szekvenálás



### Hosszú-read szekvenálás



**2. ábra Az LRS képes az izoformák azonosítására**<sup>11</sup>. Az egyes TSS-ek, TES-ek és splice junctionok SRS-sel is meghatározhatók, azonban ezek kombinációját nehéz megkülönböztetni, amennyiben több izoforma található ugyanazon a lokuszon. Továbbá, az LRS képes teljes hosszúságú transzkriptek leolvasására, amelyek teljes kontig információt nyújtanak az egyes RNS-ekben szereplő exonokról.

A legújabb LRS módszerek lehetővé teszik a natív RNS közvetlen szekvenálását is (ONT direkt RNS szekvenálás)<sup>12</sup>, minimalizálva a könyvtárkészítés során keletkező torzításokat, és lehetővé teszi az epitranszkriptomikai módosulások kimutatását is.

A vírusgenomok annotálása hagyományosan az ORF-ek azonosításával történt, illetve a Northern blot, a DNS chip és a qRT-PCR<sup>13</sup> technikákat alkalmazták, mely módszerek bár specifikusak és érzékenyek, korlátozottan alkalmazhatók csak transzkripciós aktivitás vizsgálatára, mivel csak ismert szekvenciák kimutatására képesek. Az SRS technológiák megjelenésével új lehetőségek nyíltak a vírusok és gazdasejtjeik közötti genetikai kölcsönhatások részletes vizsgálatára, a transzkripció kezdő- és végpontjainak detektálására, s maguknak a transzkripteknek az annotálására<sup>14,15</sup>. Az LRS technológiák bevezetése még tovább gyorsította a vírus transzkriptomikai kutatásokat, lehetővé téve az ismeretlen mRNS-ek, nem-kódoló (nc)RNS-ek, poligén RNS-ek, valamint különböző transzkript izoformák (TI-k) azonosítását.

A világon elsőként alkalmaztuk mind a PacBio és ONT módszereit eukarióta vírusgenom<sup>16</sup> – és transzkriptom szekvenálásra<sup>15,17-21</sup>.

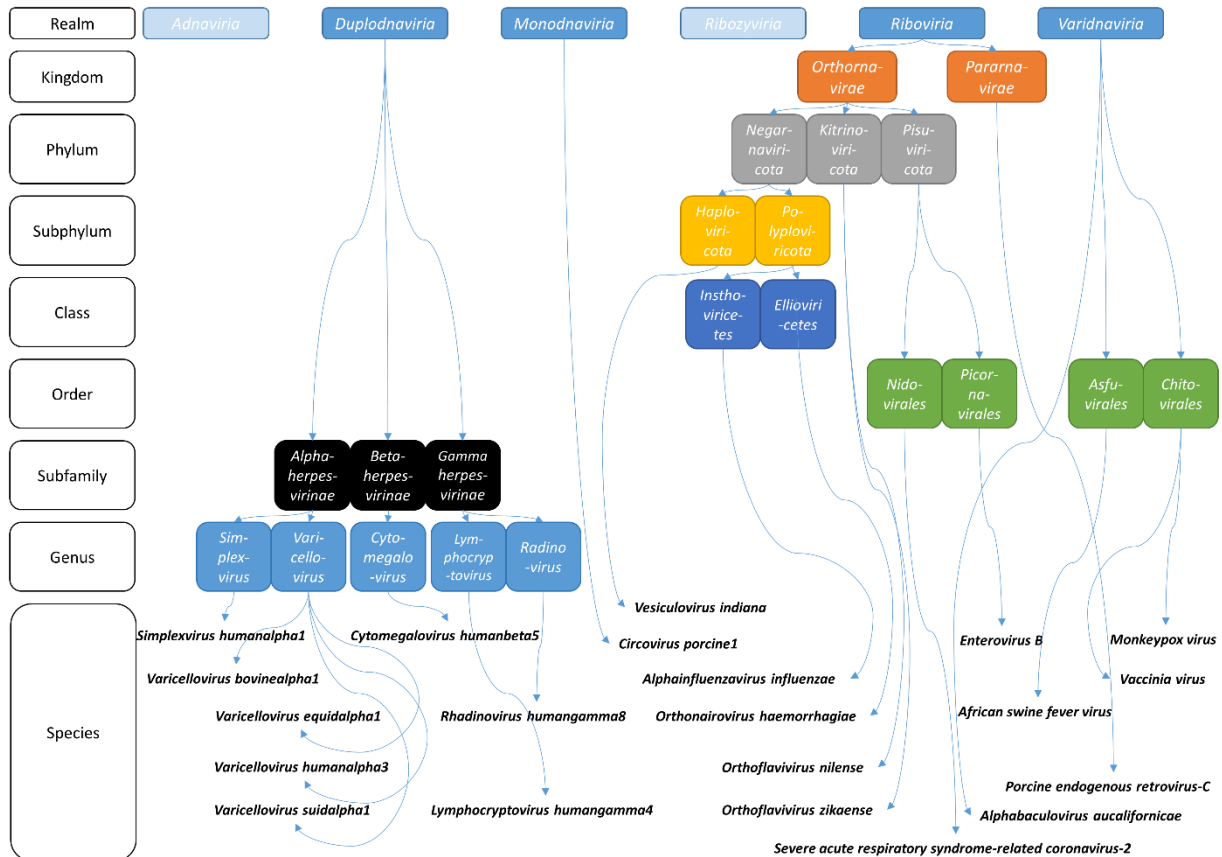
## CÉLKITŰZÉSEK

---

1. **Technológiai innováció a molekuláris virológiában:** a vírustranszkriptomok újradefiniálása, feltárása kombinált szevenálási módszerek alkalmazásával – integrálni a technológiai és bioinformatikai újításokat a projektjeink során
2. **Komplex (analitikai) megközelítések:** a vírusok szerkezetére vonatkozó ismeretek új dimenzióinak feltárása
  - a. **statikus transzkriptom meghatározása:** azonosítani, annotálni az mRNS-eket, ncRNS-eket, transzkript izoformákat
  - b. **dinamikus génextpresszió analízis:** meghatározni a vírusgének, illetve transzkriptek kinetikai osztályait, meghatározni a vírusok időben változó génextpressziós profilját
  - c. **RNS modifikációk detektálása:** azonosítani azokat a genomi régiókat, amelyek a fertőzés egyes időpontjaiban metiláltak, vagy hipereditáltak
3. **Funkcionális genomika**
  - a. olyan mutáns vírusok léterhozása, melyek transzkripció faktort kódoló, vagy egyéb, a vírusfertőzés szempontjából kulcsfontosságú génben deletáltak. Ezt azzal a céllal, hogy meghatározzuk e gének, a vírus globális expressziós profiljára gyakorolt hatását.
  - b. az általunk e projektek során leírt új RNS osztály, a replikáció-asszociált (ra)RNS-ek képviselőinek funkcionális analízise – időben változó expressziójuk, izoformáik azonosítása, validálásuk Northern blottal, qRT-PCR-rel
  - c. az új transzkript izoformák kódoló potenciáljának elemzése *in silico*
4. **Vírus-gazdasejt interakciók:** megérteni a vírusok patogenitását és ezáltal, hogy miként küzdhetjük le fertőzést.
  - a. a vírusfertőzés gazdasejtekre gyakorolt hatásának elemzése
  - b. a sejtvonal virális génextpresszióra gyakorolt hatásának vizsgálata

## EREDMÉNYEK ÉS ÉRTÉKELÉS

Céljaink megvalósításához számos vírus transzkriptom profilját és genom szekvenciáját határoztuk meg. E vírusokat az 3-as ábrán mutatom be.



3. ábra Az ICTV 2022-es taxonómiai besorolása és nevezéktana alapján az általunk analizált vírusfajok besorolása. Az egyszerűsítés érdekében minden esetben csak azokat a magasabb rendszertani kategóriákat jelöltem, melyekben különbség van a vizsgált fajok között. Az ábrán a birodalom (Realm), királyság (Kingdom), törzs (Phylum), altörzs (Subphylum), osztály (Class), rend (order), alcsalád (Subfamily), nemzetség (Genus) és faj (Species) neveket mind a 2022-es besorolás alapján jelöltem.

### TECHNOLÓGIAI ÚJÍTÁSOK: „WETLAB”, BIOINFORMATIKA

Kombinált szekvenálási módszerek alkalmazásával újradefiniáltuk számos vírus (3. ábra, 1. táblázat) transzkriptom profilját. Ehhez integráltuk az összes szekvenálási technológiát, mind a PacBio, mind az ONT módszert elsőként alkalmaztuk továbbá mind a labor, mind a bioinformatika terén újításokat vezettünk be.

Mivel sem a PacBio, sem az ONT egyik módszere sem problémamentes, mindegyik megközelítés okoz valamilyen mértékű torzítást az eredményekben, ezért kombináltuk a technikákat, valamint számos kiegészítő módszert alkalmaztunk: LRS [amplifikált és nem amplifikált (ún. direkt cDNS) szekvenálás, direkt RNS szekvenálás], SRS, CAGE-Seq, cap-szelekció, Terminátor-kezelés, méretszelekció, más, független adatszettek alkalmazása, stb. E kombinált megközelítést révén annotáltuk a vizsgálatba bevont vírusok TSS, TES, valamint a splice helyeit, intronokat, illetve magukat az RNS molekulákat is. A statikus transzkriptomok megállapítása mellett a vírusgének

kinetikai elemzését is megoldottuk az alkalmazott módszerek révén. Mindehhez a laboratóriumi technikák fejlesztésén túl, számos az elemzéshez elengedhetetlen bioinformatikai scriptet, programcsomagot is kidolgoztunk<sup>22</sup>, továbbá matematikai megközelítéseket, az RT<sup>2</sup>-PCR adatok elemzéséhez<sup>23</sup>.

ICTV elnevezés	Közismert (magyar) név, rövidítés	ICTV elnevezés	Közismert (magyar) név, rövidítés
<i>Simplexvirus humanalpha1</i>	Herpesz szimplex vírus 1-es típus, HSV-1	<i>Orthonairovirus haemorrhagiae</i>	Krími-kongói vérzések láz vírusa, CCHFV
<i>Varicellovirus humanalpha3</i>	Varicella Zoster vírus, VZV	<i>Orthoflavivirus nilense</i>	Nyugat-nílusi vírus, WNV
<i>Varicellovirus bovinealpha1</i>	Bovine alphaherpesvirus 1, BoHV-1	<i>Orthoflavivirus zikaense</i>	Zika vírus, ZIKV
<i>Varicellovirus equidalpha1</i>	Equid alphaherpesvirus 1, EHV-1	<i>Severe acute respiratory syndrome-related coronavirus-2</i>	Súlyos akut légzőszervi szindróma-koronavírus 2, SARS-CoV-2
<i>Varicellovirus suidalpha1</i>	Pseudorabies vírus (Aujeszky-féle vírus), PRV	<i>Enterovirus B</i>	Coxsackie-vírus 5B, CVB5
<i>Cytomegalovirus humanbeta5</i>	Human citomegalovírus, HCMV	<i>Porcine endogenous retrovirus</i>	Sertés endogén retrovírus, PERV
<i>Rhadinovirus humangamma8</i>	Kaposi-szarkómával társított herpeszvírus, KSHV	<i>African swine fever virus</i>	Afrikai sertéspestis vírus, ASFV
<i>Lymphocryptovirus humangamma4</i>	Epstein-Barr vírus, EBV	<i>Alphabaculovirus aocalifornicae</i>	Autographa californica multiple nucleopolyhedrovirus, AcMNPV
<i>Vesiculovirus indiana</i>	Vesicular stomatitis Indiana vírus, VSIV	<i>Monkeypox virus</i>	Majomhimlő vírus, MPXV
<i>Circovirus porcine1</i>	Porcine circovirus 1, PCV-1	<i>Vaccinia virus</i>	Vakcína vírus, VACV
<i>Alphainfluenzavirus influenzae</i>	Influenza A vírus, IVA	<i>Varicellovirus cercopithecinealpha9</i>	<i>Simian varicella virus (SVV) *</i>

1. táblázat Összefoglaló táblázat a vizsgált vírusfajokról. \* Az SVV transzkriptomikai elemzéséhez kizárólag mások által szekvenált adatokat használtunk fel.

## VÍRUS GENOMOK ANNOTÁLÁSA: DNS ÉS RNS SZEKVENCIÁKBÓL

Az RNS molekulák azonosításhoz szükség van referencia genomokra. Célszerű azt a törzset használni a térképezéshez, aminek a transzkriptomával dolgozunk. Mi a PacBio platformot használtuk a PRV genom vad típusú Kaplan (Ka) törzsének szekvenálására, s így elsőként szekvenáltunk meg e módszerrel eukarióta vírus genomot<sup>16</sup> (GenBank: KJ717942.1). A Ka törzs mellett egy általunk generált gendeletált törzset<sup>24</sup>, illetve egy szintén általunk, újonnan izolált PRV-t<sup>25</sup>, valamint a PCV-1<sup>26</sup> genomot is megszekvenáltunk. A PERV<sup>27</sup> és a VACV<sup>28</sup> genomot pedig a transzkriptok PacBio RSII, Sequel és ONT szekvenálásból nyert adataiból állítottuk össze. Továbbá,



a SARS-CoV-2 izolátum genom szekvenciáját is meghatároztuk (első magyar SARS-CoV-2 teljes genom szekvencia) a long-read RNS-Seq eredményezte leolvasásokból<sup>29</sup>.

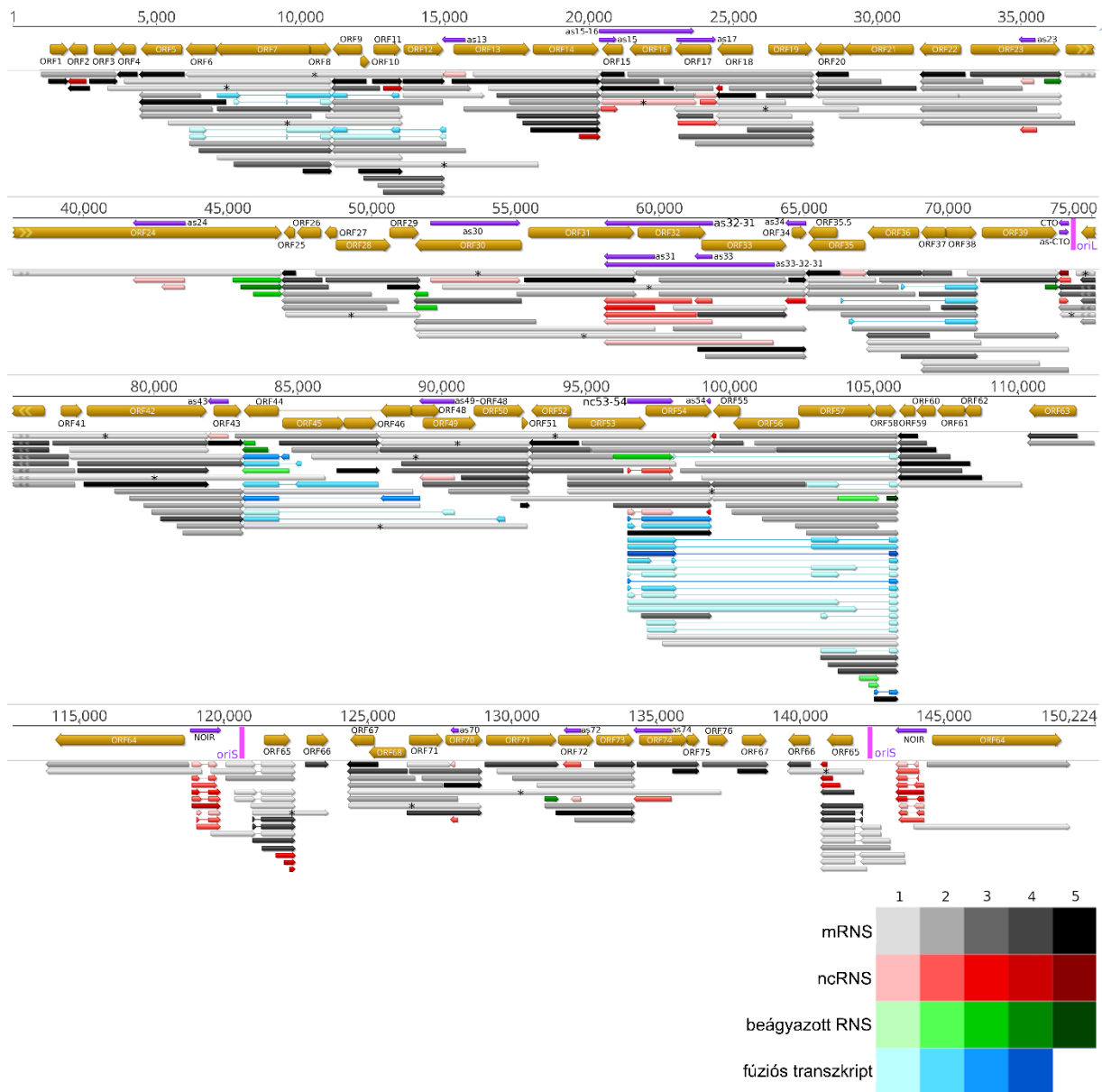
## VÍRUS TRANZKRIPTOMOK

Munkánk során az LRS módszerek közül elsőként a PacBio technológiáját (RS II készülék) vezettük be a vírusok transzkriptomának átfogó jellemzésére. Először modellvírusunk, a PRV Ka törzsének RNS profilját határoztuk meg<sup>15,17</sup>. Ezt követte számos további vírus, úgy, mint a HSV-1<sup>30</sup>, HCMV<sup>31,32</sup>, VACV<sup>33</sup>, PCV-1<sup>34</sup> és a PERV<sup>35</sup>, ugyanezen módszerrel. Majd az újabb technikák megjelenésével alkalmazni kezdtük a PacBio Sequel II-t [PRV, HSV-1, HCMV<sup>36</sup>, VACV<sup>21</sup>, AcMNPV<sup>37</sup>] és az ONT MinION [PRV<sup>18</sup>, VZV<sup>38</sup>, EBV<sup>39</sup>, HSV-1<sup>40</sup>, AcMNPV<sup>20</sup>, VACV<sup>41</sup>, ASFV<sup>42</sup>, HCMV<sup>36</sup>, BoHV-1<sup>43</sup>, PERV<sup>35</sup>, IVA, CVB5, VSIV, WNV, CCHFV, ZIKV<sup>44</sup>, SARS-CoV-2<sup>29</sup>, EHV-1<sup>45</sup>, MPXV<sup>46</sup>, KSHV<sup>47</sup>] készülékét is. Az időközben megjelenő technológiai újítások (új könyvtárkészítő kitek, enzimek, basecallerek, stb.) folyamatosan bevezetésre kerültek kutatásaink során, így ugyanazon vírusról esetenként több frissített elemzést is közöltünk. Továbbá, ha elérhető volt, mások adatait is bevontuk az analízisbe<sup>39,47,48</sup>. Sok esetben pedig, főként, azon vírusok esetében, ahol korábban még SRS platformot sem használtak transzkriptomikai céllal, ott mi ezt is elvégeztük (PRV<sup>14,17</sup>, EHV-1<sup>45</sup>, ASFV<sup>42</sup>, PCV<sup>26</sup>), és bevezettünk egy SLRS módszert is<sup>43</sup>. A 'statikus' elemzés mellett, amikor az egyes vírusok RNS variánsait vizsgáljuk, dinamikus elemzéseket is végeztünk, azaz meghatároztuk a vírusgénnek kinetikai osztályait, azt, hogy a vírusfertőzés mely időpontjában fejeződik ki, illetve, hogy az adott génről leíró izoformák közt figyelhető-e meg időbeli eltérés.

### A virális transzkriptomok strukturális analízise

A projektjeink fő célkitűzése a virális komplexitás feltérképezése volt, melyet előbb NGS, majd TGS módszerekkel vizsgáltunk.

Az elemzéseink feltárták, hogy a vírusok RNS profilja sokkal komplexebb, mint azt gondoltuk, s hogy ez általános érvényű. A virális transzkriptom korábban fel nem ismert összetettségét mutattuk ki, beleértve új fehérjét kódoló és nem kódoló gének felfedezését, új mono- és poligén transzkripciós egységeket, valamint kiterjedt transzkripciós átfedéseket mind a szomszédos, mind pedig a távoli gének között<sup>17,20,30,38</sup>. A HV-ok replikációs origóit átfedő, nem kódoló transzkripteket azonosítottunk, amelyek szerepet játszhatnak a DNS-szintézis szabályozásában<sup>15,17,39,45</sup>, továbbá a transzkripciós faktorokat kódoló gének közt is azonosítottunk átfedéseket, mely arra utalhat, hogy ezek a gének nemcsak fehérje, hanem RNS szinten is befolyásolhatják a génexpressziót. Meghatároztuk mostanra számos humán- illetve állati patogén, így öt  $\alpha$ -, egy  $\beta$ - és két  $\gamma$ HV, továbbá két, a poxvírusok családjába tartozó vírus teljes transzkriptom profilját, foglalkoztunk az afrikai sertéspestis vírusával, egy rovarvírussal, továbbá számos kis, főként RNS genomú vírus RNS profiljával is. A **4. ábrán** az EHV-1 transzkriptom látható.

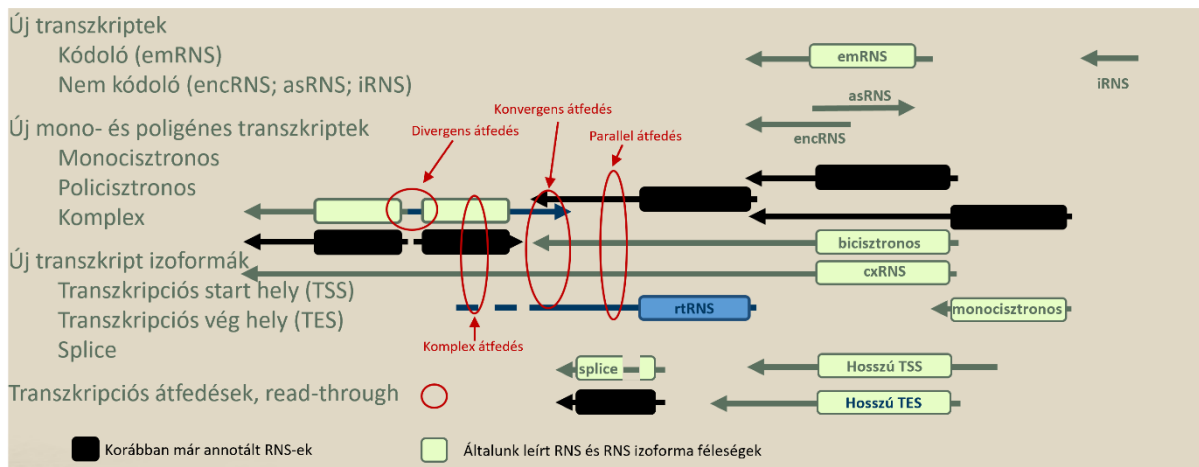


4. ábra *EHV-1* transzkriptom<sup>45</sup>. Az ONT MinION szekvenátorát alkalmaztuk az *EHV-1* genomról leíró RNS-ek komplexitásának feltárásához. Az ábrán a LoRTIA program által azonosított transzkriptok kerültek feltüntetésre. Csak azok a nagyon hosszú RNS molekulák kerültek ábrázolásra, amelyeket a LoRTIA programcsomaggal azonosítottunk a dcDNS szekvenálás adatait felhasználva. A ritka transzkriptok esetében, ahol a TSS megállapításra nem volt lehetőség, csillaggal jeleztük, ha a TSS csak becült pozíció. Az ábra a Geneious programmal készült.

### A strukturális komplexitás „hátere”

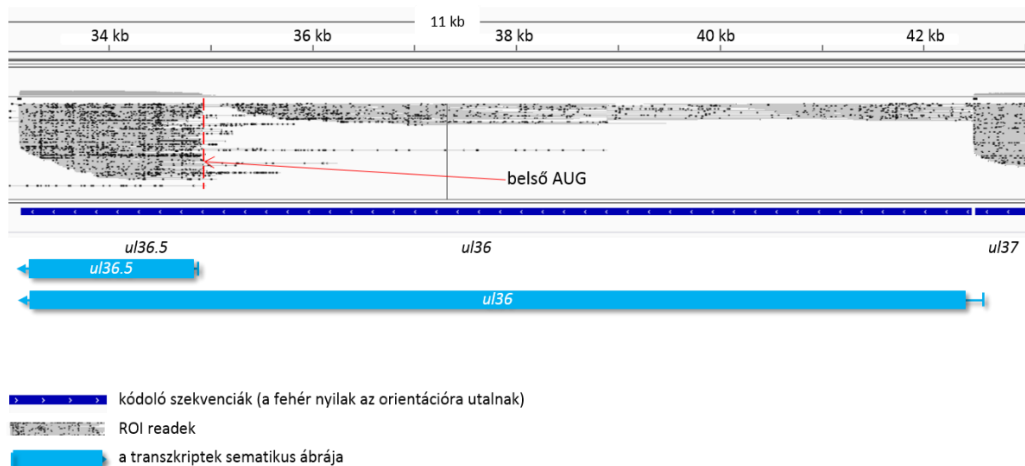
Kimutattuk tehát, hogy az egyes vírusok RNS profilja rendkívül komplex, hogy egy-egy génről nem csupán egy, hanem esetenként tucatnyi RNS-féleség íródik le. A legtöbb újonnan felfedezett transzkript olyan kategóriákba tartozik, amelyeket az SRS és más korábbi technikák segítségével nehéz tanulmányozni, mint például a beágyazott (*embedded*) mRNS-ek (emRNS-ek), a poligén transzkriptok, az átfedő transzkriptok és RNS-izoformák, beleértve a splice, TSS és TES variánsokat. Az LRS révén számos ncRNS-t [antiszensz (as)RNS, intergenikus (iRNS) és beágyazott nem kódoló (enc)RNS] is azonosítottunk. Szintén leírtunk egy speciális transzkript osztályt, melyet replikációs origó-asszociált (ra)RNS-eknek<sup>49,50</sup> nevezünk. Leírtunk számos új transzkripció átfedést mind a

szomszédos (lehet paralel, konvergens, divergens és komplex), illetve távoli gének között egyaránt. A komplex átfedéseket komplex (cx) RNS-ek alakítják ki. Mind a konvergens, mind a divergens átfedés lehet puha (*soft*), illetve kemény (*hard*): az **5. ábrán** láthatóak.

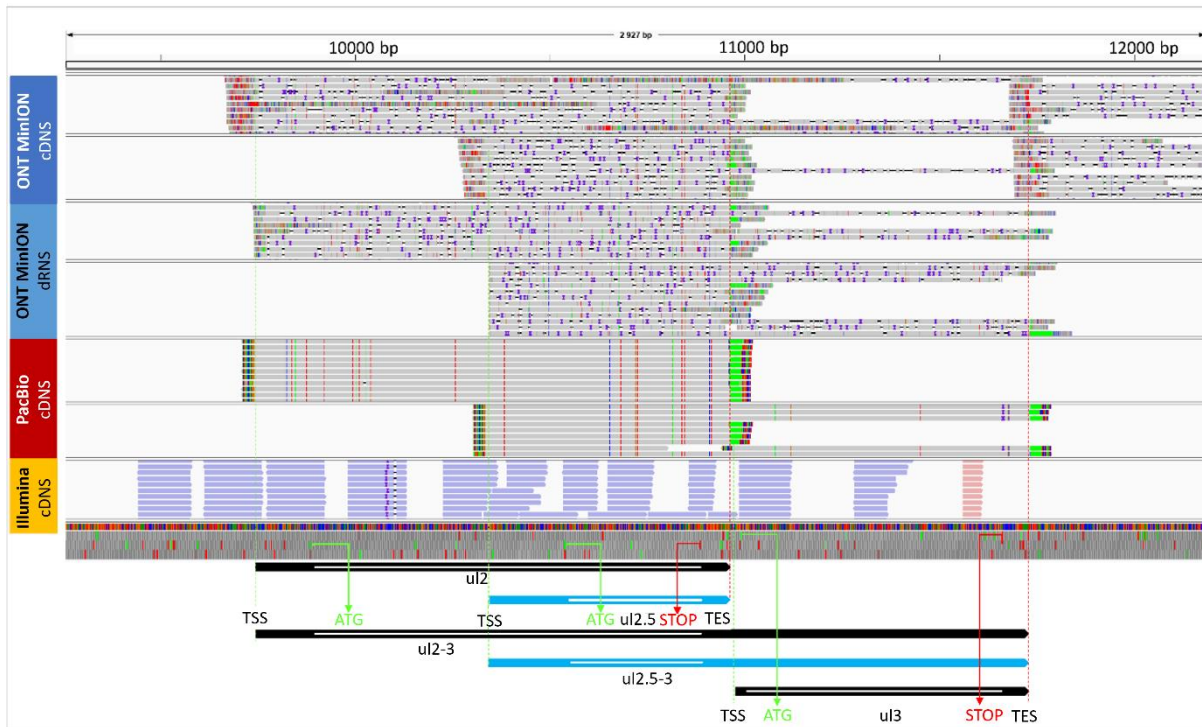


**5. ábra** Az általunk leírt transzkript féleségek és a kanonikus RNS-ek sematikus ábrája. A feketével jelölt, már korábban annotált RNS-ekhez képest számos új transzkript és transzkript izoforma variánst írtunk le, ezeket zölddel ábrázoltam. Az ábra az egyes transzkript hosszvariánsok leíródásának következtében kialakult (potenciális) átfedéseket is mutatja.

Az LRS technikák rendkívül hatékonyak azon RNS-ek azonosításában, amelyek egy hosszabb gazdagén TSS-ével vagy TES-ével közösnek osztoznak. Azok a beágyazott (e)RNS-ek, amelyek tartalmaznak egy rövidebb ORF-et [in-frame (if)ORF] a hosszabb gazdagén kódoló régióján belül, potenciálisan egy N-terminálisában csonkított polipeptidet kódolhatnak. Az ilyen emRNS-ek ismertek voltak a herpeszvírusoknál (pl. a PRV esetén az *ul3.5* gén), azonban eredményeink megsokszorozták ezeknek a potenciális mRNS-eknek a számát minden herpeszvírus-alcsaládban<sup>17,18,30</sup>, valamint egy bakulovírusban is<sup>20</sup>. A teljes EHV-1 transzkriptomot bemutató **5. ábrán** zöld színnel jelöltük az eRNS-eket, míg a **6. ábra** a PRV, a **7. ábra** pedig a HSV-1 egy-egy újonnan felfedezett eRNS-ét mutatja.



**6. ábra** Az *ul36.5* – új, potenciális fehérje kódoló gén<sup>17</sup>. Az ábrán a PRV egy 11 kbp-os genomi régiója és az ide térképeződő PacBio RSII readek IGV ábrázolása látható.



7. ábra A HSV-1 ul2.5 és az ul2.5-3 eRNS-ek<sup>48</sup>. Egy példa az ifORF-et tartalmazó transzkriptekre, melyet számos technikával kimutattunk. Az ábrán az UL2 és az egyik csonkolt transzkript variánsa (UL2.5) leolvasásai láthatóak, az IGV-vel vizualizálva. A readok LRS és SRS adatszetekből származnak. Látható, hogy a dRNS-Seq és a kétféle LRS cDNS technika ugyanazt a TSS-t detektálja (azzal a megjegyzéssel, hogy a dRNS-Seq rövidebb 5'-UTR-eket produkál, átlagosan 23 bp hiányzik). Az is látható, hogy az SRS módszer specifikus könyvtárkészítési metódus (pl. CAGE) nélkül nem hatékony az 5' végek megállapítására.

### Funkció a strukturális komplexitás mögött

Az LRS technológiák megjelenése jelentősen felgyorsította az új vírus transzkriptok és TI-ik, úgy, mint a splice, TSS és TES változatok felfedezését. Számos lncRNS-t azonosítottunk az  $\alpha$ HV-ok OriS és OriL régiói közelében<sup>17,33,50</sup>. Ezek funkciója még nem ismert, csak feltételezzük a replikációban betöltött szerepüket (elhelyezkedésük, illetve a HCMV és EBV Ori régióiban leírt és funkcionálisan vizsgált RNS-ei alapján). Ezt az új RNS osztályt replikáció origó-asszociált RNS-eknek (raRNS) nevezzük, s azok az RNS-ek alkotják, melyek a vírusok Ori-jai közvetlen közelében íródnak le, vagy azokkal át is fednek<sup>40</sup>. Továbbá olyan RNS-ek, melyek az Ori-k körüli, a vírus transzkripció, illetve replikáció szabályozásában résztvevő génekkel fednek át<sup>50</sup>.

A HV-ok mindhárom alcsaládjában ( $\alpha$ ,  $\beta$  és  $\gamma$ ) azonosítottunk raRNS-eket, melyek a vírus genom Ori régiói köré térképeződnek<sup>15,17,31</sup>, továbbá a bakulo<sup>20,37</sup>- és cirkovírusokban<sup>34</sup> is találtuk az Ori régiókba térképeződő transzkripteket. Egy általános, a herpeszvírusok mindhárom alcsaládját érintő analízist végeztünk, melybe saját és mások SRS, LRS és RAMPAGE-Seq [RNA annotation and mapping of promoters for analysis of gene expression sequencing] adatait is bevontuk, illetve konkrétan e régiók vizsgálata céljából végeztünk további LRS és CAGE-Seq kísérleteket, qPCR validálást<sup>49</sup>.

Speciális elhelyezkedésük és az ebből (is) következő feltételezett szerepük miatt e tézisben hangsúlyozottan tárgyalom az raRNS-eket, az alphaherpeszvírusok OriS régiójára fókuszáltn.

### **Az alfaherpeszvírusok raRNS-ei**

Az raRNA-ek közé azok az RNS-ek tartoznak ide, melyek átfedik az Ori-k valamelyikét, vagy azok közelébe térképeződnek. Vannak köztük lncRNS-ek és mRNS-ek hosszú, 5'-, illetve 3'-UTR izoformái egyaránt. Az OriS elhelyezkedése az  $\alpha$ -HV-ok esetén konzervatív, mindig az *us1* génhez képest upstream helyezkedik el. Az alcsalád általunk szekvenált öt (PRV, HSV-1, VZV, EHV-1 és BoHV-1) tagja, illetve a csak adatok szintjén feldolgozott SVV mindegyike esetén azonosítottunk új, az OriS körül leíró, illetve azzal átfedő transzkripteket és új RNS izoformákat.

A NOIR-1 transzkript családba tartozó RNS-ek mind az IR régióban íródnak le és közös 3' koterminálisban végződnek egymással és az antiszensz transzkripttel (AST) és a hosszú látencia transzkripttel (LLT) is<sup>6</sup>. Az általunk leírt NOIR-2 szintén az IR régióból íródik le, a *noir-2* gén az *ie180* géntől downstream helyezkedik el, azzal parallel orientációban, s konvergens állásban a *noir-1*-hez képest. A NOIR-2 rendkívül alacsony abundanciájú RNS molekula. Az ORF-ek hiánya arra utal, hogy az összes *noir* gén lncRNS. Egyik leolvasási keretben sem figyelhető meg GC-preferencia, ez is erősíti azt, hogy nem kódoló gének.

Az AZURE (*antisense transcripts in the IR-US overlapping region*) névre keresztelt transzkriptet is azonosítottuk, mely az *us1*-gyel ellentétes orientációjú. Az AZURE egy lncRNS, mely részben átfed az *us3* génnel, illetve az IR régió egy szegmensével. Az AZURE hosszabb, TSS variánsa részben az *us4* génnel is átfed, antiszensz állásban.

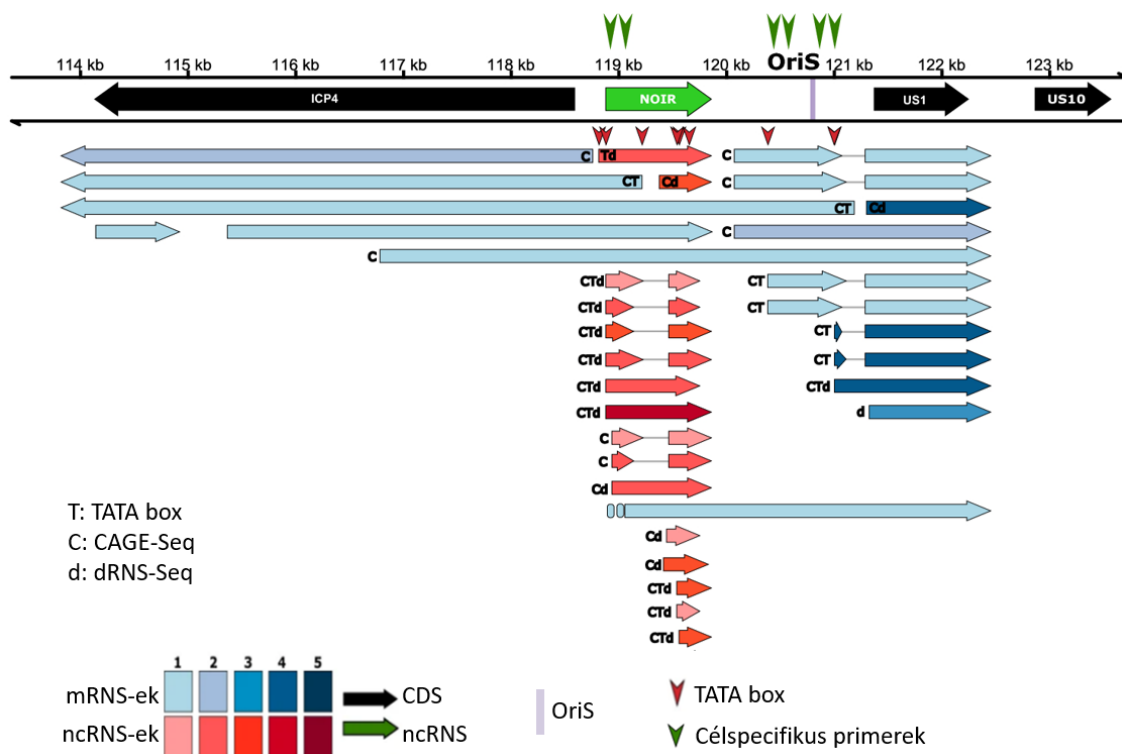
Az EHV-1 azonosított raRNS-ei közül az ORF65 (az *us1*, azaz az *icp22* homológja) génjéről leíró hosszú 5'-UTR izoforma átfedi az OriS-t<sup>45</sup>. Azonosítottunk egy TATA-boxot az OriS-en belül, és azonosítottuk azt a transzkriptet, amely valószínűleg ennek a promóternek a szabályozása alatt áll. Kimutattuk a PRV NOIR-1-gyel homológ NOIR-t. Az EHV-1 NOIR-1 transzkriptnek két splice izoformája van, illetve azonosítottunk spliceolatlan verziót is (**8. ábra**).

A BoHV-1 OriS-eit két egymással ellentétes orientációjú IE transzkript fedi át. Egyike ezen raRNS-eknek a bICP4 nagyon hosszú TSS transzkript izoformája, a másik pedig a bICP22 nagyon hosszú TSS variánsa. Ez a két TSS variáns az OriS mellett egymással is átfed<sup>51</sup>. Abundanciájukat (lévén nagyon hosszúak) az LRS alábecsüli. A bICP22-nek több izoformáját is azonosítottuk. A TAATGARAT és a TAATGARAT-szerű szekvenciák a herpeszvírusok IE génjei aktivitásához szükségesek. A *bicp22* gén promótere az IE transzkripció unit 1 (IEtu1) tartalmaz TAATGARAT-szerű szekvenciát (TAATCGAGA) és az ehhez asszociálható TSS átfed az OriS-sel. Ez egyedülálló az alfaherpeszvírusok között. Továbbá kimutattuk az ORIS-RNS1 és ORIS-RNS2 raRNS-eket is. A BoHV-1 raRNS-ek alacsony abundanciájúak, s csupán a fertőzés későbbi időpontjaiban jelennek meg. A HSV-1-ben ugyanezt a szerveződést figyeltük meg, míg a VZV-ben és a PRV-ben csak a bICP22 homológjának hosszú 5' UTR variánsát detektáltuk. Számos raRNS-t azonosítottunk a VZV OriS-sel átfedésben is. A VZV-ben a csoportunk által kimutatott<sup>38</sup> raRNS-eknek egy csoportja a kanonikus ORF62-ben (*icp4* gén homológja) helyezkedik el, részben azzal ellentétes orientációban (asRNS-ek). Továbbá, az *us1* homológ ORF63 hosszú TSS variánsai is átfedik az OriS-t.

Ezeknek az RNS-eknek a funkciója ismeretlen, lehetséges, hogy a NOIR transzkripteknek is közvetlen vagy közvetett szerepük van a replikáció szabályozásában.

Az  $\alpha$ HV-ok OriS közeli régiójában tehát az alábbi lncRNS-eket, illetve TI-kat azonosítottuk: a BoHV-1 esetén az OriS-RNS-t, a HSV-1 esetén OriS-RNS1-et, a PRV-nél, az EHV-1-nél, az SVV-nél a NOIR-1 transzkriptet, míg a PRV esetén a NOIR-2 RNS-t is. Kimutattuk, hogy a BoHV-1, EHV-1, HSV-1 és

SVV vírusokban a transzkripciót szabályozó gének (mint az *us1* és az *icp4*) nagyon hosszú 5' TI-jei átfednek az OriS-sel. Az  $\alpha$ HV-okban az US1 RNS esetén egy rendkívül komplex splice mintázatot azonosítottunk. Azonosítottunk egy a HSV-1 OriS-RNS1-ével antiszensz állású lncRNS-t. Az általunk leírt NOIR-1 géncsalád tagjai divergens állásúak az *icp4* génhez képest. Ezeknek a NOIR-1 RNS-eknek a kanonikus tagjai nincsenek átfedésben az *icp4*-gyel, azonban a hosszabb izoformák részben átfedik ezt a legfontosabb virális TF-t kódoló gént. Az EHV-1, VZV és SVV hosszú US1 transzkript variánsai a *noir-1* gén promóteréből indulnak. A PRV-nél csak feltételezzük ezt. Az SVV transzkriptom analízise során azonosítottuk a NOIR-1 egy olyan TSS variánsát, mely átfed az ICP4 RNS kanonikus változatával. A NOIR-1-nek mind a kanonikus, mind a fenti TSS izoformája átfed az OriS-sel. A VZV ugyanezen génszegmensében öt lncRNS-t azonosítottunk (NOIR-1A, -1B, 1C, -1D, and -1E). Mind a hat  $\alpha$ HV esetében olyan TSS-ekkel rendelkező TI-eket találtunk, amelyek nagyon közel helyezkednek el az OriS-en belüli TATA-boxokhoz, ami arra utal, hogy ezek a promóter elemek funkcionálisak lehetnek.



**8. ábra Az EHV-1 OriS-közeli transzkriptjei<sup>49</sup>.** Az ábrán az EHV-1 OriS régiója és azok környékéről leíró transzkriptek láthatóak. Fehérje kódoló gének: fekete nyíl; nem-kódoló géneket zöld; mRNS-ek: kék; ncRNS-ek: piros nyíl. A többi HV-vel való összehasonlíthatóság érdekében a HSV-1 terminológia szerint neveztük a géneket. A színek intenzitása az abundanciára utal: 1: 1–9 leolvasás (read), 2: 10–49 read, 3: 50–199 read, 4: 200–999 read, 5: > 1000 read. Az intronokat vízszintes vékony összekötő vonalak jelzik. A PCR során alkalmazott primerek pozícióit zöld nyilakkal jelöltük.

## Dinamikus transzkriptom: a génextpresszió időbeli változásainak elemzése

Számos vírus esetén meghatároztuk a géntermékek relatív expressziós szintjét. Az IE gének azonosításához cikloheximid (CHX) kezelést végeztünk, mely fehérjeszintézisgátló, hiányában nem termelődnek azok a TF-ok, melyek az E és L gének leíródásához szükségesek. A gének kinetikai besorolásához alkalmaztuk a PacBio, illetve az ONT szekvenálást egyaránt, előbbi esetben az amplifikáció-mentes IsoSeq módszert, míg utóbbi esetben a dcDNS szevenálást. A PacBio IsoSeq könyvtárakat szekvenáltuk az RSII-vel a PRV<sup>52</sup> és a Sequel-Iel a VACV<sup>33,41</sup> esetén, míg ONT MinION-on szekvenáltuk a HSV-1<sup>40</sup>, a VACV<sup>31,41</sup> és a BoHV-1<sup>43</sup> dcDNS könyvtárakat a dinamikus RNS profiljuk elemzéséhez.

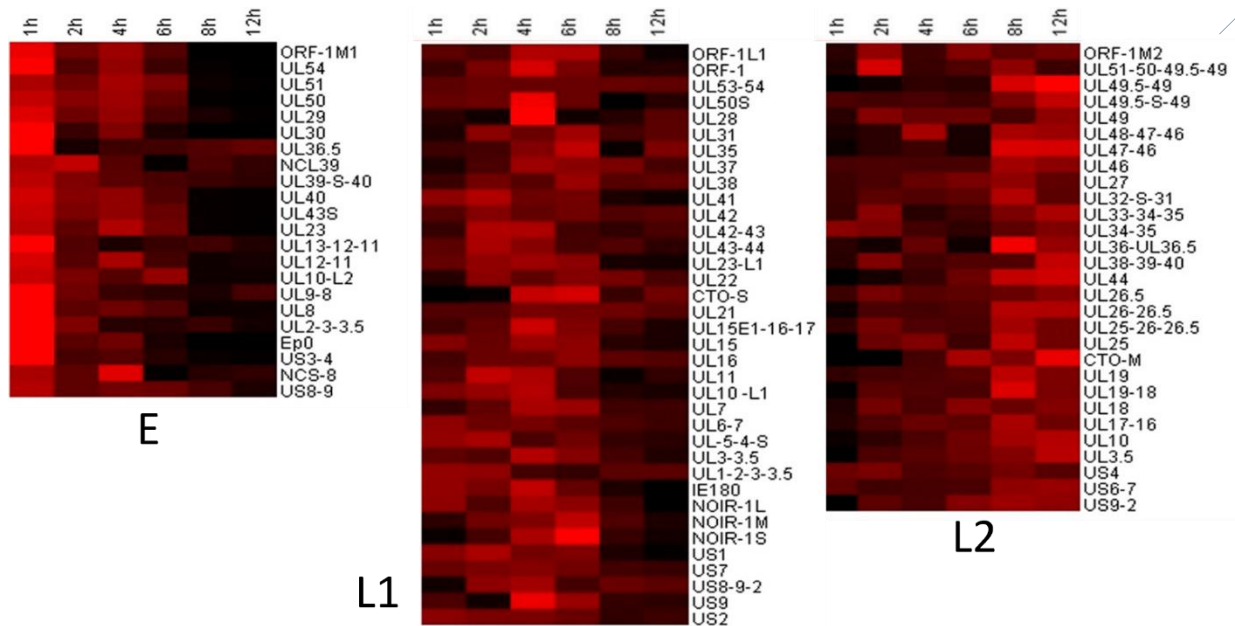
### A dinamikus PRV transzkriptom

A herpeszvírusok transzkriptomának (főként HSV-1 és PRV) dinamikus tulajdonságait korábban számos módszerrel tanulmányozták, beleértve a microarray<sup>53</sup> Illumina szekvenálást<sup>54</sup>, valamint az általunk végzett qRT-PCR elemzéseket<sup>13</sup>. Saját PhD témám is a PRV génextpresszió qRT-PCR-rel való elemzését emelte a fókuszba<sup>13</sup>, s azt követően, az LRS módszer bevezetését megelőzően is ezzel a technikával foglalkoztam.

Az LRS módszerek megjelenésével azokat elsőként vezettük be a vírusok dinamikus transzkriptom profilalkotásához is<sup>52</sup>. A PRV transzkriptom időbeni változásainak vizsgálatához a PacBio RSII szekvenálási platformját használtuk. Ennek a munkának kettős célja volt: egyrészt a PRV transzkriptek jellemzése és osztályozása kinetikai tulajdonságaik alapján egy új módszerrel, továbbá annak bemutatása, hogy az LRS módszer hogyan használható és mennyire hasznos a globális transzkripció kvantitatív elemzésében, az időben változó génextpresszió össz-genomszintű profilozásában.

A könyvtárkészítést a PacBio „*Very Low Input*” protokollja szerint végeztük, mely a relatíve rövid (1-2 kb) hosszúságú readeknek kedvez, ezért az ennél hosszabb és rövidebb leolvasások alulreprezentáltak a mintánkban. Az abszolút kvantifikálást nem tette lehetővé a módszerünk, azonban a relatív expressziós ( $F_x$ ) értékek kiszámításával a génkifejeződés mértéke összehasonlítható. Azokat a transzkripteket, amelyek átlagos  $F$  értéke kisebb volt, mint 0,05%, kizártuk az elemzésből. A PRV transzkriptek poligénes természete egy további nehézség volt a transzkriptek azonosítása során. Azokat a poligénes egységekről származó, nem teljes hosszúságú readeket, melyek legalább 8 bp-t tartalmaztak az upstream génből, a hosszabb RNS molekulához kalkuláltuk. Ezt a küszöböt a PRV statikus elemzésekor<sup>17</sup> tapasztalt, a transzkriptek TSS-ére vonatkozó (~ 4 bp-os) hossz polimorfizmus miatt állítottuk be. A vírus RNS-ek kinetikai kategorizálását azon elvek alapján végeztük, hogy az E gének a vírus fertőzési ciklus korai időszakában nagymértékben kifejeződnek, ugyanakkor a későbbi időszakokban relatíve alacsony expressziót mutatnak. Az L2 gének viszont csak kis mértékben fejeződnek ki a fertőzés első óráiban, s később válnak abundánssá. Az L1 gének köztes kinetikai profillal rendelkeznek.

A transzkriptek  $F_x$  értékeit k-közép klasztereztük, Pearson-korrelációs analízist végeztünk, s így állapítottuk meg az egyes klasztereket, s az ide tartozó transzkripteket (**9. ábra**).



9. ábra A PRV gének transzkripciós kinetikája hőterképés ábrázolása<sup>52</sup>. A PRV transzkriptek Fx értékei elkülönülő expressziós profilt mutatnak a három kinetikai klaszterben. Soronként láthatjuk az egyes transzkriptek relatív expressziós változásait. A piros a magas, míg a fekete az alacsony relatív expressziós értékre utal

Eredményeink szerint az általános E, L1 és L2 génkifejeződések jelentősen különböznek egymástól. Az E gének magas relatív kifejeződést mutatnak a DNS replikáció kezdete előtt, amely 12 h-val a fertőzés után jelentősen csökken. Az L2 gének 'inverz' kinetikát mutatnak az E-khez képest. Az L1 gének kezdetben eltérően viselkednek az E génektől, míg később expressziós profiljuk hasonlóvá lesz. Ellenben, az L1 gének az infekció első óráiban az L2 génekhez hasonló kifejeződési dinamikájúak, ami az infekció második felében eltérővé válik. Az E gén termékek maximális értékeikhez viszonyított aránya magasabb volt az L génekhez képest az infekció első 4 órájában. Minden gén expressziója nő 6 és 8 h között; látszólag részben a vírus DNS replikáció miatt. Az E gén termékek mennyisége tipikusan csökken 8 és 12 h között, ami ellentétes az ebben az időszakban növekvő L2 transzkriptumok mennyiségével.

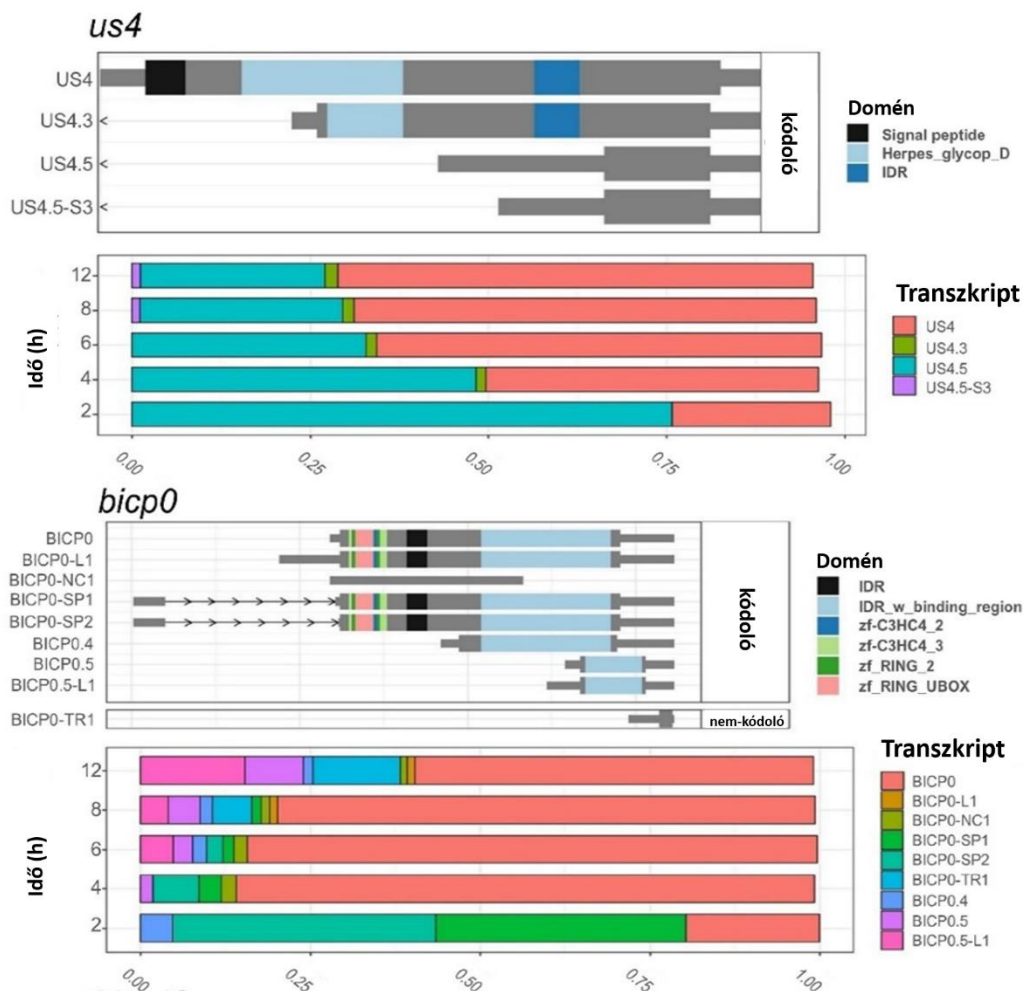
### A vírusgén domének időfüggő kifejeződése

Az izoformák (különösen a splice variánsok) amelyek különböző fehérjedoméneket hordoznak, gyakoriak például az emberi sejtekben és bizonyították, hogy fontos szerepet játszanak a rák kialakulásában<sup>55</sup>. A vírusokban, ahol a transzkriptek összetettsége és ugyanabból a génből kifejezett transzkript izoformák száma sokkal magasabb, ez a jelenség még fontosabb lehet. A transzkript izoformák arányának változása különböző funkciókat indukálhat. A BoHV-1 transzkriptom vizsgálata során a vírusfertőzés alatt e lehetséges jelenségek megfigyelése érdekében az *isoformswitchanalyser* eszközt<sup>56</sup> használtuk a LoRTIA által kapott transzkript mennyiségeken.

A **10. ábra** az annotált fehérjedoméneket, jelpeptideket és IDR-eket mutatja be két példagén transzkriptjeiben. Az elemzés eredményei szerint az *us4.5* gén 5' végén csonkolt transzkriptjei nem rendelkeznek az összes azonosított fehérjedoménrel és jelpeptiddel a kanonikus US4 transzkripthez képest. Az adataink azt mutatták, hogy kezdetben az *us4.5* gén termékei nagyobb arányban keletkeznek, de a fertőzés késői szakaszában az *us4* gén US4 transzkriptjei válnak dominánssá (**10. ábra**). Ez azt jelentheti, hogy egy kezdeti "zajos" transzkripció (amely funkció



nélküli transzkripteket eredményez) után a transzkripció a funkcionális mRNS-ek kifejezésére vált. Ezt a megfigyelést az a tény is alátámasztja, hogy az *us4* egy kései gén, a fertőzés korai szakaszában nem szükséges. Nem zárható ki, hogy az US4.5 transzkript funkcionális, és lehetséges, hogy szabályozhatja az *us4* gén kifejeződését. Hasonló mintát találtunk az *ul44* gén esetében is. A *bicp0* gén két splice variánst fejez ki (azonos doménszerkezettel) a kanonikus izoformával nagyjából egyenlő mennyiségben, de két óra után jelentős csökkenést figyeltünk meg. Több 5'-csonkolt izoforma (néhány annotált domén nélkül) kisebb abundanciában fejeződött ki a vírusfertőzés későbbi szakaszaiban. Több másik gént is azonosítottunk az *isoformswitchanalyser*-rel, amelyek izoforma váltást mutatnak 'következménnyel' (*ul18*, *ul21*, *ul40*, *ul44*), de ezekben a génekben az izoforma arány nagyon hasonló volt minden mintában a 2 h-s időpontot követően.

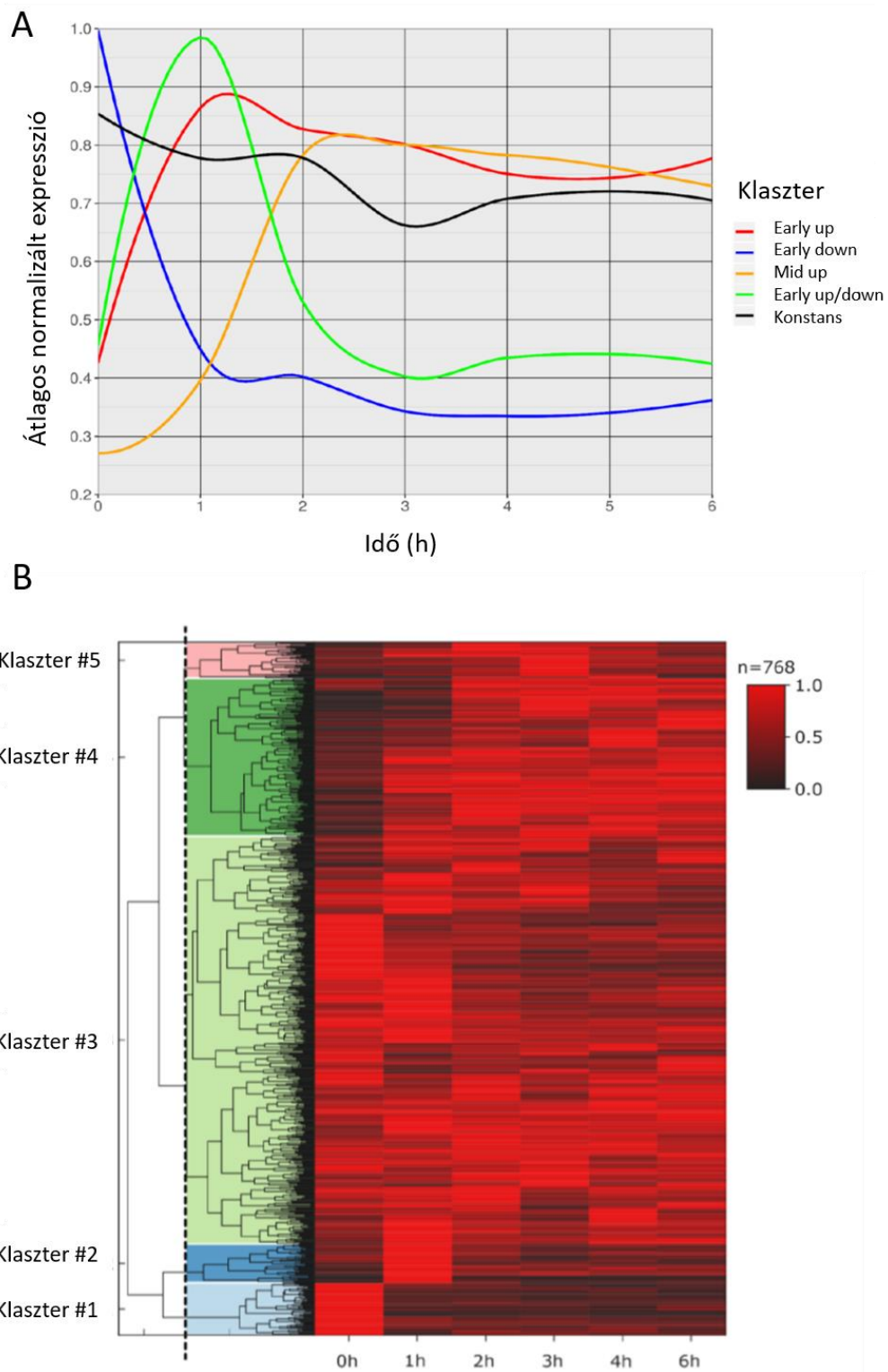


10. ábra A színes boxok a Pfam adatbázis használatával azonosított fehérje doméneket mutatják<sup>51</sup>. Az ábrán a *bicp4* és a *bicp0* gének a fertőzés során megjelenő izoformák szerkezetét és az expressziós profiljukat tüntettük fel. A megváltozott RNS expressziós profil eredményeként izoforma switchet (váltást) detektáltunk, mely váltások a fehérje domén struktúra változását is indikálják, továbbá változást a szignálpeptid tartalmában, vagy a predikált fehérje rendezetlenségében (IDR). A transzkript izoformákat kódolási valószínűségük alapján csoportosítottuk (amit a CPAT2 REF-fel értékeltünk). A 3'- és 5'-csonkolt transzkriptek csak predikált ORF-eket tartalmaznak.

## A GAZDASEJT TRANZKRIPTOMOK ÉS A VÍRUSFERTŐZÉS A GAZDAGÉNEK EXPRESSZIÓJÁRA KIFEJTETT HATÁSÁNAK ANALÍZISE

Számos vírus esetén elvégeztük a gazdasejtek expressziójára gyakorolt hatásuk elemzését is (pl. BoHV-1, EHV-1, VACV, VSIV, SARS-CoV-2). A tézisben részletesen a VACV gazdasejtjeinek a vírusfertőzés következtében megváltozó génextpressziós profilját mutatom be.

**A gazdagének válasza a VACV fertőzés hatására:** korábbi tanulmányok kevés differenciálisan expresszálandó gént azonosítottak a VACV fertőzés hatására<sup>57</sup>, míg egy nemrégiben végzett proteomikai tanulmány szintén azt mutatta, hogy a VACV fertőzés nagyon kevés gazdagént befolyásol<sup>58</sup>. Az eredményeink szerint nagymértékben expresszálandó 768 gént öt klaszterbe csoportosítottuk a vírusfertőzésre mutatott válaszuk alapján (**11. ábra**). Az úgynevezett *'early-up'* (korai-fel) klaszterbe sorolt gének a vírusfertőzést megelőzően (mock) nem, vagy nagyon alacsony mértékben fejeződtek ki, míg a fertőzés utáni későbbi időpontokban konzisztensen magas expressziójúak. Az *'early down'* (korai-le) RNS-ek a mock mintákban nagy mértékben expresszálandóak, míg a fertőzés után vagy alig fejeződtek ki, vagy egyáltalán nem íródtak le. Az *'early up/down'* (korai fel/le) kategóriába sorolt transzkriptek nem, vagy alig íródnak le a VACV fertőzés előtt, majd a fertőzés után 1 h-val magas az expressziójuk, ami szinte teljes mértékben megszűnik a következő órákra. A *'mid up'* (közép fel) transzkripteknek nincs expressziója a mock sejteken, a csúcs kifejeződésük és a plató fázisuk 2-3 h-val a fertőzés után van. A *'constant'* (konstans) transzkriptek kategóriáját olyan gének alkotják, melyek relatív expressziója mértékében nem történt szignifikáns változás a vírusfertőzés általunk vizsgált időpontjaiban (**12. ábra**). A legjobban jellemezett géntaszterek expressziós mintázatainak értékelése során a GO-t használtuk, és jelentős túlréprezentációt találtunk a "szignálreceptor aktivitás szabályozása" GO folyamatban azon géneknél, amelyek a korai szakaszokban mutattak magas expressziót, de a később nem, vagy csak alig fejeződtek ki. Továbbá, sok, a vírusfertőzés során upregulált gén játszik szerepet a sejtosztódásban, vagy a "vírus életciklusának pozitív szabályozásában". Néhány, a vírusfertőzés hatására downregulált gént pedig a "sejtnövekedés" és a "mezenchimális differenciáció" kategóriákba annotáltunk.



11. ábra Az **abundáns gazdagének expressziós változásai a VACV fertőzés során**<sup>59</sup>. (A) Gazdagén klaszterek expressziós profilja. (B) Azonosított öt elkülönülőgazdagén klaszter hő térképes ábrája.

A céljaink eléréséhez technológiai újításokat hajtottunk végre, integratív megközelítéseket alkalmazva meghatároztuk számos víruscsalád tagjainak transzkriptom profilját, melyhez lényegében az összes rendelkezésre álló szekvenáló platformot (Illumina, PacBio RSII és Sequel, ONT MinION) és számos könyvtárkészítési technikát alkalmaztunk. A sejtes szervezetekkel összehasonlítva, a vírusoknak igen kevés génjük van, genomjuk kompakt, ezért ideális modellszervezetek általános transzkriptomikai kérdések vizsgálatára is. Már az első LRS eredményeink sokkal bonyolultabb RNS profilt mutattak, mint azt a korábbi ismeretek alapján gondoltuk. Eredményeink egyben arra is rámutattak, hogy az LRS-nek számos előnye van az RNS szekvenálásban az SRS módszerekhez képest. **Eredményeink egyértelműsítették, hogy a vírus transzkripció rendkívül komplex és, hogy szinte a teljes genom transzkripcionálisan aktív.** Az intergenikus régiókat hosszú UTR szekvenciák írják át, illetve önálló transzkriptek is leíródnak erről a DNS szakasról.

**Kimutattuk, hogy sokkal több RNS és transzkript izoforma képződik egy adott ORF-ről, és, hogy a poligénus RNS-ek száma is sokkal magasabb, mint, ami eddig ismert volt.** Munkánk eredményeként **megsokszoroztuk az ismert transzkriptek számát** minden vizsgált vírus esetében. Genomméretükhöz mérten nagyszámú transzkriptet és RNS izoformát azonosítottunk még a kisméretű, DNS genommal rendelkező PCV-nél és PERV-nél, illetve az RNS vírusok közül a VSIV-nél és a SARS-CoV-2-nél. A többi vizsgált RNS vírusról ezt a nagyfokú komplexitást nem tapasztaltuk.

**Számos új potenciális mRNS-t, ncRNS-t azonosítottunk** mind géneken belül, mind az intergenikus régiókban egyaránt az összes nagy genomú vírusról. **Meghatároztuk a vírusgének kanonikus TSS és TES pozícióit, illetve az alternatív hossz-izoformákat.** Leírtunk számos új intront, spliceolt transzkriptet és alternatív splice variánst. Kimutattuk, hogy **a poxvírusok genomjának középső régiója a génexpresszió szempontjából 'kaotikus'**: az e genom-szegmensébe lokalizálódó L gének nem rendelkeznek fix TSS és TES pozíciókkal, hanem **a genom számos pontjáról indul és végződik a transzkripció.** Ebben a régióban **az antiszensz RNS-ek expresszió is rendkívül magas.** E kaotikus expressziós profil a vírusfertőzés kései fázisaira hatványozottan érvényes. A korai időpontokban, illetve a genom két szélső szegmensében az egyes génekről sokkal szabályosabb RNS-ek íródnak le, itt alig jellemzőek az alternatív TSS és TES variációk. **Kimutattuk, hogy a poxvírusok genomja az elsőtől az utolsó nukleotidig transzkripcionálisan aktív.** A herpeszvírusoknál ismert volt, hogy a gének jelentős hányada poligénus. Leírtuk, hogy ez a pox-és a bakulovírusokra is jellemző. Kimutattuk, hogy a konvergens, divergens és parallel transzkripciós átfedések alapvető jellemzői a virális RNS profiloknak. **A transzkripciós átfedések komplex hálózatát tártuk fel** a vizsgált vírusokban, még távoli gének között is igen gyakoriak az RNS szintű átfedések. A virális transzkriptomok feltárása nagy jelentőségű, mert betekintést nyerhetünk általa a vírusok replikációjának és patogenezisének molekuláris mechanizmusába, ami segíthet a potenciális antivirális célpontok fejlesztésében.

Amellett, hogy meghatároztuk a virális transzkriptek izoformáit, azt is vizsgáltuk, hogy ezek a vírusfertőzés során mikor íródnak le. Az időben változó transzkripciós profilok elemzése ugyanis alapvető jelentőségű a génexpresszió összgenom-szintű szabályozásának megértéséhez. **Meghatároztuk a vírusgének kinetikai osztályait,** illetve a transzkripteket is kategorizáltuk aszerint, hogy a vírusfertőzés mely időpontjában fejeződnek ki kisebb, vagy nagyobb mértékben. Az általunk leírt, saját promóterrel rendelkező ncRNS-eket (pl. a CTO, a NOIR és ezek izoformái) is be tudtuk sorolni a megállapított kinetikai osztályok valamelyikébe. Az **expressziós dinamikák alapján csoportokat alkottunk.** Meghatároztuk, hogy az adott génről leíródnak különböző

**transzkript izoformák egymáshoz viszonyított aránya hogyan változik a fertőzés egyes időpontjaiban.** Továbbá, megvizsgáltuk a transzkript izoformák kódoló kapacitását is. Megállapítottuk, hogy hosszú 5' UTR-eket eredményező TES variánsok uORF-eket kódolnak, melyek szabályozó szereppel bírnak, míg az egyéb variánsok (rövid 5' UTR, illetve splice variánsok) esetén a fehérje domén szerkezete, ezáltal a funkciójuk tér el a kanonikus variánsétól.

Összefoglalva, **bemutattuk az LRS hasznosságát a vírusok dinamikus transzkriptomának vizsgálatában is.** Megállapítottuk, hogy ezt a technikát alkalmazni lehet minden olyan folyamat tanulmányozására, amelyek egy meghatározott, jól kontrollált transzkripció időbeli lefolyást mutatnak, mint például a vírusok replikációja, az embriófejlődés, a szöveti regeneráció, stb. Továbbá, sikerült nyomon követnünk az egyes TI-k, intron variánsok és átfedő RNS-ek egyéni dinamikáját, amelyek különösen fontosak a nagy DNS genomú vírusoknál, amelyek összetett transzkriptom profillal rendelkeznek a produktív fertőzés során. Ezenkívül jellemeztük számos vírus RNS kinetikai sajátosságait is.

A vírusgének LRS által megállapított kinetikai besorolását – ahol rendelkezésre álltak adatok - összehasonlítottuk a qRT-PCR eredményeinkkel. Ezek javarészt megfeleltek egymásnak. Az LRS-alapú megközelítés előnye, hogy e módszerrel el tudtuk különíteni egymástól az izoformákat, illetve az egymással parallel átfedő RNS-eket is, melyre a qPCR nem alkalmas.

Mind a statikus, mind pedig a dinamikus elemzésekhez (beleértve a TSS-ek, TES-ek, intronok és maguknak a transzkripteknek a meghatározását) a csoportunk által fejlesztett LoRTIA transzkript annotátort, illetve számos 'in-house' scriptet alkalmaztuk. A dinamikus transzkriptom analízishez **saját matematikai megközelítéseket** dolgoztunk ki.

Megszekvenáltuk a PRV-Ka törzset, saját izolátumunkat, az MdBio törzset, valamint számos mutáns genomot. **Elsőként használtuk vírusgenom meghatározásra a PacBio SMRT technológiát, s kimutattuk, hogy a PacBio módszer optimális a magas G+C tartalmú, repeat régiókban gazdag genom szekvenciák meghatározására.** A genom szekvenálás révén meghatároztuk a saját labortörzsünk és az általunk izolált MdBio törzsek közti DNS szintű eltéréseket és meghatároztuk a hozzájuk legközelebbi PRV genomokat. Továbbá, megszekvenáltuk és annotáltuk a PCV-1 és a PERV genomokat is.

A VACV PacBio és ONT transzkriptom szekvenálása során a genom teljes, közel 200 kbp hosszúságú szakaszának minden nukleotidjáról kaptuk leolvasást, azaz **a VACV teljes genom transzkripcionálisan aktív.** Az LRS alapú RNS-Seq módszer eredményezte adataink a **genom meghatározására is alkalmazhatóak voltak.** Hasonlóképpen, a SARS-CoV-2 genomot is meghatároztuk a transzkriptom mellett.

Az ONT dRNS-Seq technikájával, illetve validálásként Illumina MiSeq készüléken biszulfid konverziós mintákon, **kimutattuk, hogy egyes AcMNPV gének metiláltak, illetve hipereditáltak,** továbbá, az **ONT natív RNS szekvenálás alkalmazhatóságát az RNS módosulások detekciójára.**

**Homológ rekombinációval géniütött PRV-ket állítottunk elő,** majd a vad típusú vírussal párhuzamosan az összes génjük expresszióját mértük real-time kvantitativ RT-PCR-rel. **Meghatároztuk az ep0, az ul41 és az ul54 gének teljes virális RNS profilra gyakorolt hatását.**

Az LRS módszert alkalmazva **meghatároztuk a BoHV-1 fertőzés gazdagének expressziójára gyakorolt hatását,** továbbá **annotáltunk gazda RNS-eket.** A **Bos taurus (MDBK) géneket expressziós dinamikájuk alapján klaszterekbe rendeztük, s meghatároztuk azokat a funkcionális**

**csoporthoz, melyek alul-, vagy felülreprezentáltak a vírusfertőzés hatására.** Az alul-reprezentált csoportokat alkotók zömében az alapvető sejtfunkciókban szerepelnek, úgy, mint morfogenezis, a sejtciklus szabályozása, a jelátvitel, a katabolikus útvonalak és az aerob légzés, míg jelentős mértékben nőtt az antivirális válaszban részt vevő gének expressziója a fertőzés következtében. Ezenkívül, a transzkripcióban, RNS lebontásában, translációban szerepet játszó gének is upreguláltak. Azonosítottunk egy kis csoportot az azonnali választ adó génekből is, amelyek jelentősen megváltozott expressziót mutattak már 1 h-val a vírusfertőzést követően. A VACV fertőzés hatását a CV-1 sejtvonalon vizsgáltuk, melynek **eredményeként azonosítottunk számos új *Chlorocebus sabaeus* (CV-1) RNS-t és transzkript izoformát, illetve a fertőzés hatására overreprezentált gének csoportját,** melyek a „jelátviteli receptor aktivitásának szabályozása” GO (Gene Ontology - Génontológia) annotációban szerepelnek. Ez azt jelenti, hogy a fertőzött sejtekben aktívabbak azok a gének, amelyek a sejt felszíni vagy intracelluláris receptorok által közvetített jelátviteli folyamatok aktivitását szabályozzák. Ez a megfigyelés arra utal, hogy a vírusfertőzés hatással van a gazdasejt jelátviteli útvonalaira, potenciálisan befolyásolva a sejt a fertőzésre adott választát, beleértve az immunválaszt és a sejtszaporodást. **A VSIV fertőzés hatását két sejtvonalon (Vero és T98G) is megfigyeltük.** Jelentős különbséget találtunk a két sejt típusban a vírusfertőzés hatékonyságában. **A gliasejtek sokkal érzékenyebbek a fertőzésre,** míg a fibroblaszt sejtek ellenállóbbak. Közel 1400 eltérően kifejeződő gént és 35 különbözően expresszálandó KEGG (Kyoto Encyclopedia of Genes and Genomes - Kyoto Gén- és Genomtudományi Enciklopédia) útvonalat észleltünk a sejt vonalak között. Körülbelül 2,5-szer több gént találtunk a gliasejtekben, amelyek kifejeződési trajektóriáját jelentősen befolyásolta a vírusfertőzés (452 a gliasejtekben, és 172 a Vero sejtekben). Ezek közül 137-et azonosítottak korábban veszélyes fertőzés során az egér agysejtekben<sup>60</sup>.

Kimutattuk az LRS módszerek alkalmazhatóságát a virális fertőzés gazdasejtekre gyakorolt hatásának elemzése, a gazdagének annotációja, s transzkript izoformáinak meghatározására egyaránt. **A vírus-gazda interakciókkal kapcsolatos adataink értékesek jövőbeli funkcionális tanulmányokhoz és annak megértéséhez, hogy a vírus hogyan küzd le a gazda védekezési mechanizmusait.** Ezenkívül ezek az **eredmények hasznosak lehetnek az új antivirális terápia fejlesztéséhez.**

A projektjeink során – illetve egyes vírusoknál mások által - kimutatott jelenségek, azaz a rendkívül összetett, genom-szerte átfedő transzkripció hálózat létezése<sup>41,50</sup>, illetve a szorosan egymás mellett elhelyezkedő gének által kódolt RNS-molekulák parallel, divergens vagy konvergens módon történő átfedése az egyes víruscsaládoknál, a **transzkripció apparátusok közötti kölcsönhatásra utal az egész vírusgenom.** Kimutattuk, hogy a legtöbb, általunk vizsgált vírusban **a replikációs origókat átfedik bizonyos lncRNS-ek, valamint az mRNS-ek hosszú 5' vagy 3' UTR izoformái.**

Munkánk egy fontos célja az volt, hogy az Ori-k környékén leíró RNS-eket, melyek kulcsszerepet játszanak a herpeszvírusok replikációjának és/vagy genom-szerte történő transzkripciójának szabályozásában, annotáljuk. Nagyobb hangsúlyt fektettünk az  $\alpha$ HV-ok vizsgálatára, mivel esetükben ezekről az rRNS-ekről kevesebb információ állt rendelkezésre a többi herpeszvírushoz képest. Munkánk **során leírtunk új és korrigáltunk már leírt rRNS-ek korábbi annotációit.** Továbbá **azonosítottunk egy bonyolult TO-hálózatot a replikációs és transzkripció szabályozó gének által kódolt transzkriptek és körülöttük lévő specifikus lncRNS-ek között is.** Promóter konszenzus szekvenciákat is detektáltunk az összes vizsgált herpeszvírus OriS-ében. Bár a

herpeszvírus OriS-ai AT-gazdag szekvenciákat tartalmaznak, amelyeket potenciálisan tévesen TATA boxoknak azonosíthatók, minden esetben azonosítottuk a megfelelő TSS-eket ezekhez a szekvenciákhoz közeli helyeken. Lehetségesnek tartjuk, hogy néhány, a különféle vírusokban általunk azonos névvel illetett lncRNS valójában polifiletikus eredetű. Mindazonáltal ***a CTO és NOIR-1 transzkriptcsaládok ortológok a PRV és EHV-1 esetében.***

A DNS-replikáció és a transzkripció egyidejű aktivitása ugyanazon genomi régióban interferenciát generál a két folyamat között az egész genom mentén<sup>61</sup>. Korábbi eredmények azt sugallják, hogy ezek az ütközések drámaibb következményekkel járnak, amikor konvergens orientációban történnek, nem pedig parallel<sup>62</sup>. Számos molekuláris mechanizmus alakult ki a RNP és a replikációs villa közötti konfliktus minimalizálására. Azonban ***eredményeink arra utalnak, hogy létezik egy beépített, belső szabályozó rendszer, amely a két apparátus közötti kölcsönhatáson alapul, hogy együttműködve szabályozza mind a replikációt, mind a transzkripció kezdetét.*** Úgy gondoljuk, hogy a leíró RNP és a repliszóma közötti ütközés és versengés, valamint a folyamatban lévő DNS és RNS szintézisek által létrehozott pre-replikációs és transzkripció-iniciációs komplexek összeállítása közvetíti ezt az interakciót (60. ábra).

Az  $\alpha$ HV-okban az OriS pozíciója rendkívül konzervatív: az US repeat régiók területén helyezkedik el, és a fő transzkripciót szabályozó gének, az *icp4* és *us1* veszik körül. ***Kimutattuk, hogy ezek a gének transzkript izoformákat állítanak elő, amelyek nem csak az OriS-sel, hanem a replikációs origó másik oldalán lévő TF RNS-ekkel és/vagy specifikus Ori-közeli lncRNS-ekkel is átfedésben vannak.*** A TO-k valószínűsíthető funkciója, hogy további interakció formákat segítsen elő ezek között a gének között (a TF/promóter interakción túl), mely interakciók jelenthetnek RNS:DNS és valószínűleg RNS:RNS hibridizációkat egyaránt, valamint a transzkripció apparátusok közötti interferenciát. Ennek eredményeként feltételezhető, hogy ***az  $\alpha$ HV-ok US-IR régiója egyfajta „szuper szabályozó központként” működik,*** ahol a DNS replikáció és az összgenom-szintű transzkripció iniciáció egy kölcsönhatásokon alapuló többszintű rendszeren keresztül, együttesen szabályozott. Továbbá, ez a genomi szegmens szabályozza a litikus ciklus és a latencia közötti átmenetet, valamint ezeknek a folyamatoknak a fenntartását. Így ez a régió funkcionálisan az  $\alpha$ HV-ok legösszetettebb genomi lokusza, amely lncRNS-eket, azaz intergenikus transzkripteket, asRNS-eket, s mRNS-ek hosszú izoformáit kódolja, melyek együttese egy bonyolult TO mintázatot hoz létre. A litikus transzkriptek mellett, több látens lncRNS (LAT, LLT, L/ST) és miRNS is kifejeződik ebből a genomi régióból.

***Azonosítottuk a NOIR-1 transzkript családot, amelyet az US-IR régió kódol, és amely a Varicellovírus nemzetség evolúciós újdonsága.*** Ezeknek az lncRNS-eknek a pontos helye jelentős változatosságot mutat, azonban közös bennük, hogy 3'-koterminálisban végződnek a latenciában működő LLT transzkripttel, ugyanakkor a lítikus ciklusban expozálódnak. Emellett, a különféle NOIR RNS izoformák az OriS régió környéki transzkripció szabályozó faktort kódoló génekkel és/vagy magával a replikációs origóval fednek át. ***A NOIR-1 RNS-eknek szerepe lehet a transzkripció szabályozásában azáltal,*** hogy átfedik az *icp4* TF gént, illetve ***kontrollálhatják a DNS replikációt*** az OriS-sel való átfedésük révén. Azon esetekben, amikor mind az OriS-t, mind az *icp4*-et átfedik, hatásuk lehet mindkét mechanizmusra. Hasonlóképpen működhet a HSV-1-ben az általunk leírt OriS-RNS1-is.

A Simplexvírusok OriL-jét a fő replikációs gének (*ul29/ul30*) veszik körül. Feltételezzük, hogy ezek a gének nem csak a hagyományos TF/promóter kötődésen keresztül léphetnek interakcióba a

replikáció iniciációs folyamatával, hanem más módon is, beleértve az RNS:DNS és/vagy RNS:RNS hibridizációt, valamint az átfedő régióban a transzkripció és replikáció apparátusai közti interferenciát. A ***CTO ncRNS-eket először PRV-ben azonosítottuk<sup>15,17</sup>, majd detektáltuk az ortológjaikat a közeli rokon EHV-1-ben<sup>45</sup>***. A kanonikus ***CTO-S*** nem fedi át az OriL-t, de ***befolyásolhatja a replikációt*** azáltal, hogy segít szétválasztani a két DNS szálat, s így például így meghatározza a replikáció irányát. Az egyik CTO-S TI-t az OriL-en belüli promóter szabályozza, míg a másik az *ul21* gén TES variánsa.

Észrevettük, hogy az  $\alpha$ HV-okban, összesen három kemény konvergens átfedés van, s azt is, hogy e génpárok egyik tagja mindig egy, a replikációban szerepet játszó gén. A jelenség valószínűleg funkcióval bír: feltételezésünk szerint a Simplexvírusok *ul30/ul31* génpárja közti kemény TO hatására az *ul31* hosszú, az OriL-t átfedő TES izoformákat produkál transzkripció read-through révén. Bár a PRV-ben és az EHV-1-ben az OriL az *ul21* és *ul22* gének között helyezkedik el, az OriL-t az *ul21* hosszú TES izoformája is átfedi. Mivel a Simplexvírusokban az *ul21*-nek, a Varicellovírusokban pedig az *ul31*-nek nincsenek hosszú TES izoformái, azt feltételezzük, hogy ezeknek a TI-knek a szerepe a DNS replikáció kezdetének megzavarása.

Eredményeink azt mutatják, hogy ***az raRNS-ek jelenléte általános jelenség a herpeszvírusok körében***, azonban még ***a közeli rokon fajok is egyedi stratégiákat fejlesztettek ki***, önálló lncRNS-ek, mRNS-ek izoformái, vagy ezek kombináció alkotják őket. Ugyanakkor ***az általános előfordulásuk a jelenség jelentőségét mutatja, a DNS replikáció és az általános transzkripció szabályozásában***. Míg az  $\alpha$ HV-ok OriS régiójában a fő TR gének úgy tűnik, hogy egymást és a DNS replikáció kezdetét szabályozzák, a Simplexvírusok OriL régiójában a fő replikációs gének inkább egymást és a DNS replikációt szabályozhatják a TO-kon keresztül. Ezek a feltételezett mechanizmusok több szabályozási szintet biztosítanak a hagyományos transzkripció faktor/promóter interakción túl.



## IRODALOMJEGYZÉK

1. Prorok J, Kovács PP, Kristóf AA, Nagy N, Tombácz D, Tóth JS, Ordög B, Jost N, Virág L, Papp JG, Varró A, Tóth A, Boldogkoi Z. Herpesvirus-mediated delivery of a genetically encoded fluorescent Ca(2+) sensor to canine cardiomyocytes. *J Biomed Biotechnol*. 2009;2009:361795.
2. Marshall E. Gene therapy. Viral vectors still pack surprises. *Science*. 2001 Nov 23;294(5547):1640.
3. Boldogkoi Z, Balint K, Awatramani GB, Balya D, Busskamp V, Viney TJ, Lagali PS, Duebel J, Pásti E, Tombácz D, Tóth JS, Takács IF, Scherf BG, Roska B. Genetically timed, activity-sensor and rainbow transsynaptic viral tools. *Nature Methods* 2009 Feb;6(2):127-30.
4. Wang S, Liang B, Wang W, Li L, Feng N, Zhao Y, Wang T, Yan F, Yang S, Xia X. Viral vectored vaccines: design, development, preventive and therapeutic applications in human diseases. *Signal Transduct Target Ther*. 2023 Apr 7;8(1):149.
5. Reale A, Vitiello A, Conciatori V, Parolin C, Calistri A, Palù G. Perspectives on immunotherapy via oncolytic viruses. *Infect Agent Cancer*. 2019 Feb 11;14:5.
6. Thiel K. Old dogma, new tricks--21st Century phage therapy. *Nat Biotechnol*. 2004 Jan;22(1):31-6.
7. Gorbalenya A, Krupovic M, Siddell S, Varsani A, Kuhn J. Establishing a single taxon that comprises RNA viruses at the basal rank of virus taxonomy. Report number: ICTV Proposal No. 2017.006G.
8. Howley PM, Knipe DM, Enquist LW: Fields Virology: Fundamentals. 7th ed. 2023. Wolters Kluwer.
9. Elias P, Lehman IR. Interaction of origin binding protein with an origin of replication of herpes simplex virus 1. *Proc Natl Acad Sci U S A*. 1988 May;85(9):2959-63.
10. Zhao W, Zeng W, Pang B, Luo M, Peng Y, Xu J, Kan B, Li Z, Lu X. Oxford nanopore long-read sequencing enables the generation of complete bacterial and plasmid genomes without short-read sequencing. *Front Microbiol*. 2023 May 15;14:1179966.
11. Tombácz D, Balázs Z, Csabai Z, Snyder M, Boldogkői Z. Long-Read Sequencing Revealed an Extensive Transcript Complexity in Herpesviruses. *Front Genet*. 2018 Jul 17;9:259.
12. Workman RE, Tang AD, Tang PS, Jain M, Tyson JR, Razaghi R, Zuzarte PC, Gilpatrick T, Payne A, Quick J, Sadowski N, Holmes N, de Jesus JG, Jones KL, Soulette CM, Snutch TP, Loman N, Paten B, Loose M, Simpson JT, Olsen HE, Brooks AN, Akeson M, Timp W. Nanopore native RNA sequencing of a human poly(A) transcriptome. *Nat Methods*. 2019 Dec;16(12):1297-1305.
13. Tombácz D, Tóth JS, Petrovszki P, Boldogkoi Z. Whole-genome analysis of pseudorabies virus gene expression by real-time quantitative RT-PCR assay. *BMC Genomics*. 2009 Oct 23;10:491.
14. Oláh P, Tombácz D, Póka N, Csabai Z, Prazsák I, Boldogkői Z. Characterization of pseudorabies virus transcriptome by Illumina sequencing. *BMC Microbiol*. 2015 Jul 1;15:130.
15. Tombácz D, Csabai Z, Oláh P, Havelda Z, Sharon D, Snyder M, Boldogkői Z. Characterization of novel transcripts in pseudorabies virus. *Viruses*. 2015 May 22;7(5):2727-44.
16. Tombácz D, Sharon D, Oláh P, Csabai Z, Snyder M, Boldogkői Z. Strain Kaplan of Pseudorabies Virus Genome Sequenced by PacBio Single-Molecule Real-Time Sequencing Technology. *Genome Announc*. 2014 Jul 17;2(4):e00628-14.
17. Tombácz D, Csabai Z, Oláh P, Balázs Z, Likó I, Zsigmond L, Sharon D, Snyder M, Boldogkői Z. Full-Length Isoform Sequencing Reveals Novel Transcripts and Substantial Transcriptional Overlaps in a Herpesvirus. *PLoS One*. 2016 Sep 29;11(9):e0162868.

18. Moldován N, Tombácz D, Szűcs A, Csabai Z, Snyder M, Boldogkői Z. Multi-Platform Sequencing Approach Reveals a Novel Transcriptome Profile in Pseudorabies Virus. **Front Microbiol.** 2018 Jan 22;8:2708.
19. Boldogkői Z, Szűcs A, Balázs Z, Sharon D, Snyder M, Tombácz D. Transcriptomic study of Herpes simplex virus type-1 using full-length sequencing techniques. **Sci Data.** 2018 Nov 27;5:180266.
20. Moldován N, Tombácz D, Szűcs A, Csabai Z, Balázs Z, Kis E, Molnár J, Boldogkői Z. Third-generation Sequencing Reveals Extensive Polycistronism and Transcriptional Overlapping in a Baculovirus. **Sci Rep.** 2018 Jun 5;8(1):8604.
21. Tombácz D, Prazsák I, Szucs A, Dénes B, Snyder M, Boldogkői Z. Dynamic transcriptome profiling dataset of vaccinia virus obtained from long-read sequencing techniques. **Gigascience.** 2018 Dec 1;7(12):giy139.
22. Balázs Z, Tombácz D, Csabai Z, Moldován N, Snyder M, Boldogkői Z. Template-switching artifacts resemble alternative polyadenylation. **BMC Genomics.** 2019 Nov 8;20(1):824.
23. Tombácz D, Tóth JS, Boldogkői Z: Deletion of the virion host shut: off gene of pseudorabies virus results in selective upregulation of the expression of early viral genes in the late stage of infection. **Genomics** 2011. 98(1):15-25.
24. Csabai Z, Takács IF, Snyder M, Boldogkői Z, Tombácz D. Evaluation of the impact of ul54 gene-deletion on the global transcription and DNA replication of pseudorabies virus. **Arch Virol.** 2017 Sep;162(9):2679-2694.
25. Csabai Z, Tombácz D, Deim Z, Snyder M, Boldogkői Z. Analysis of the Complete Genome Sequence of a Novel, Pseudorabies Virus Strain Isolated in Southeast Europe. **Can J Infect Dis Med Microbiol.** 2019 Apr 4;2019:1806842.
26. Tombácz D, Moldován N, Balázs Z, Csabai Z, Snyder M, Boldogkői Z. Genetic Adaptation of Porcine Circovirus Type 1 to Cultured Porcine Kidney Cells Revealed by Single-Molecule Long-Read Sequencing Technology. **Genome Announc.** 2017 Feb 2;5(5):e01539-16.
27. Szűcs A, Moldován N, Tombácz D, Csabai Z, Snyder M, Boldogkői Z. Long-Read Sequencing Reveals a GC Pressure during the Evolution of Porcine Endogenous Retrovirus. **Genome Announc.** 2017 Oct 5;5(40):e01040-17.
28. Prazsák I, Tombácz D, Szűcs A, Dénes B, Snyder M, Boldogkői Z. Full Genome Sequence of the Western Reserve Strain of Vaccinia Virus Determined by Third-Generation Sequencing. **Genome Announc.** 2018 Mar 15;6(11):e01570-17.
29. Tombácz D, Dörmő Á, Gulyás G, Csabai Z, Prazsák I, Kakuk B, Harangozó Á, Jankovics I, Dénes B, Boldogkői Z. High temporal resolution Nanopore sequencing dataset of SARS-CoV-2 and host cell RNAs. **Gigascience.** 2022 Oct 17;11:giac094.
30. Tombácz D, Csabai Z, Szűcs A, Balázs Z, Moldován N, Sharon D, Snyder M, Boldogkői Z. Long-Read Isoform Sequencing Reveals a Hidden Complexity of the Transcriptional Landscape of Herpes Simplex Virus Type 1. **Front Microbiol.** 2017 Jun 20;8:1079.
31. Balázs Z, Tombácz D, Szűcs A, Csabai Z, Megyeri K, Petrov AN, Snyder M, Boldogkői Z. Long-Read Sequencing of Human Cytomegalovirus Transcriptome Reveals RNA Isoforms Carrying Distinct Coding Potentials. **Sci Rep.** 2017 Nov 22;7(1):15989.
32. Balázs Z, Tombácz D, Szűcs A, Snyder M, Boldogkői Z. Long-read sequencing of the human cytomegalovirus transcriptome with the Pacific Biosciences RSII platform. **Sci Data.** 2017 Dec 19;4:170194.

33. Tombácz D, Prazsák I, Csabai Z, Moldován N, Dénes B, Snyder M, Boldogkői Z. Long-read assays shed new light on the transcriptome complexity of a viral pathogen. *Sci Rep*. 2020 Aug 14;10(1):13822.
34. Moldován N, Balázs Z, Tombácz D, Csabai Z, Szűcs A, Snyder M, Boldogkői Z. Multi-platform analysis reveals a complex transcriptome architecture of a circovirus. *Virus Res*. 2017 Jun 2;237:37-46.
35. Moldován N, Szucs A, Tombácz D, Balázs Z, Csabai Z, Snyder M, Boldogkői Z. Multiplatform next-generation sequencing identifies novel RNA molecules and transcript isoforms of the endogenous retrovirus isolated from cultured cells. *FEMS Microbiol Lett*. 2018 Mar 1;365(5):fny013.
36. Balázs Z, Tombácz D, Szűcs A, Snyder M, Boldogkői Z. Dual Platform Long-Read RNA-Sequencing Dataset of the Human Cytomegalovirus Lytic Transcriptome. *Front Genet*. 2018 Sep 27;9:432.
37. Boldogkői Z, Moldován N, Szűcs A, Tombácz D. Transcriptome-wide analysis of a baculovirus using nanopore sequencing. *Sci Data*. 2018 Dec 4;5:180276.
38. Prazsák I, Moldován N, Balázs Z, Tombácz D, Megyeri K, Szűcs A, Csabai Z, Boldogkői Z. Long-read sequencing uncovers a complex transcriptome topology in varicella zoster virus. *BMC Genomics*. 2018 Dec 4;19(1):873.
39. Fülöp Á, Torma G, Moldován N, Szenthe K, Bánáti F, Almsarrhad IAA, Csabai Z, Tombácz D, Minárovits J, Boldogkői Z. Integrative profiling of Epstein-Barr virus transcriptome using a multiplatform approach. *Viol J*. 2022 Jan 6;19(1):7.
40. Tombácz D, Moldován N, Balázs Z, Gulyás G, Csabai Z, Boldogkői M, Snyder M, Boldogkői Z. Multiple Long-Read Sequencing Survey of Herpes Simplex Virus Dynamic Transcriptome. *Front Genet*. 2019 Sep 24;10:834.
41. Tombácz D, Prazsák I, Torma G, Csabai Z, Balázs Z, Moldován N, Dénes B, Snyder M, Boldogkői Z. Time-Course Transcriptome Profiling of a Poxvirus Using Long-Read Full-Length Assay. *Pathogens*. 2021 Jul 21;10(8):919.
42. Torma G, Tombácz D, Csabai Z, Moldován N, Mészáros I, Zádori Z, Boldogkői Z. Combined Short and Long-Read Sequencing Reveals a Complex Transcriptomic Architecture of African Swine Fever Virus. *Viruses*. 2021 Mar 30;13(4):579. doi: 10.3390/v13040579.
43. Tombácz D, Moldován N, Torma G, Nagy T, Hornyák Á, Csabai Z, Gulyás G, Boldogkői M, Jefferson VA, Zádori Z, Meyer F, Boldogkői Z. Dynamic Transcriptome Sequencing of Bovine Alphaherpesvirus Type 1 and Host Cells Carried Out by a Multi-Technique Approach. *Front Genet*. 2021 Apr 7;12:619056.
44. Prazsák I, Csabai Z, Torma G, Papp H, Földes F, Kemenesi G, Jakab F, Gulyás G, Fülöp Á, Megyeri K, Dénes B, Boldogkői Z, Tombácz D. Transcriptome dataset of six human pathogen RNA viruses generated by nanopore sequencing. *Data Brief*. 2022 Jun 18;43:108386.
45. **Tombácz D**, Torma G, Gulyás G, Fülöp Á, Dörmő Á, Prazsák I, Csabai Z, Mizik M, Hornyák Á, Zádori Z, Kakuk B, Boldogkői Z: Hybrid sequencing discloses unique aspects of the transcriptomic architecture in equid alphaherpesvirus 1. *Heliyon* 2023. 9(7):e17716.
46. Kakuk B, Dörmő Á, Csabai Z, Kemenesi G, Holoubek J, Růžek D, Prazsák I, Dani VÉ, Dénes B, Torma G, Jakab F, Tóth GE, Földes FV, Zana B, Lanszki Z, Harangozó Á, Fülöp Á, Gulyás G, Mizik M, Kiss AA, Tombácz D, Boldogkői Z. In-depth Temporal Transcriptome Profiling of Monkeypox and Host Cells using Nanopore Sequencing. *Sci Data*. 2023 May 9;10(1):262.

47. Prazsák I, Tombácz D, Fülöp Á, Torma G, Gulyás G, Dörmő Á, Kakuk B, McKenzie Spires L, Toth Z, Boldogkői Z. KSHV 3.0: a state-of-the-art annotation of the Kaposi's sarcoma-associated herpesvirus transcriptome using cross-platform sequencing. *mSystems*. 2024 Feb 20;9(2):e0100723.
48. Tombácz D, Torma G, Gulyás G, Moldován N, Snyder M, Boldogkői Z. Meta-analytic approach for transcriptome profiling of herpes simplex virus type 1. *Sci Data*. 2020 Jul 9;7(1):223.
49. Torma G, Tombácz D, Csabai Z, Almsarrhad IAA, Nagy GÁ, Kakuk B, Gulyás G, Spires LM, Gupta I, Fülöp Á, Dörmő Á, Prazsák I, Mizik M, Dani VÉ, Csányi V, Harangozó Á, Zádori Z, Toth Z, Boldogkői Z. Identification of herpesvirus transcripts from genomic regions around the replication origins. *Sci Rep*. 2023 Sep 29;13(1):16395.
50. Boldogkői Z, Moldován N, Balázs Z, Snyder M, Tombácz D. Long-Read Sequencing - A Powerful Tool in Viral Transcriptome Research. *Trends Microbiol*. 2019 Jul;27(7):578-592.
51. Tombácz D, Kakuk B, Torma G, Csabai Z, Gulyás G, Tamás V, Zádori Z, Jefferson VA, Meyer F, Boldogkői Z. In-Depth Temporal Transcriptome Profiling of an Alphaherpesvirus Using Nanopore Sequencing. *Viruses*. 2022 Jun 13;14(6):1289. doi: 10.3390/v14061289.
52. Tombácz D, Balázs Z, Csabai Z, Moldován N, Szűcs A, Sharon D, Snyder M, Boldogkői Z. Characterization of the Dynamic Transcriptome of a Herpesvirus with Long-read Single Molecule Real-Time Sequencing. *Sci Rep*. 2017 Mar 3;7:43751.
53. Aguilar JS, Devi-Rao GV, Rice MK, Sunabe J, Ghazal P, Wagner EK. Quantitative comparison of the HSV-1 and HSV-2 transcriptomes using DNA microarray analysis. *Virology*. 2006 Apr 25;348(1):233-41.
54. O'Grady T, Cao S, Strong MJ, Concha M, Wang X, Splinter Bondurant S, Adams M, Baddoo M, Srivastav SK, Lin Z, Fewell C, Yin Q, Flemington EK. Global bidirectional transcription of the Epstein-Barr virus genome during reactivation. *J Virol*. 2014 Feb;88(3):1604-16.
55. Vitting-Seerup K, Sandelin A. The Landscape of Isoform Switches in Human Cancers. *Mol Cancer Res*. 2017 Sep;15(9):1206-1220.
56. Vitting-Seerup K, Sandelin A. IsoformSwitchAnalyzeR: analysis of changes in genome-wide patterns of alternative splicing and its functional consequences. *Bioinformatics*. 2019 Nov 1;35(21):4469-4471.
57. Rubins KH, Hensley LE, Relman DA, Brown PO. Stunned silence: gene expression programs in human cells infected with monkeypox or vaccinia virus. *PLoS One*. 2011 Jan 18;6(1):e15615.
58. Soday L, Lu Y, Albarnaz JD, Davies CTR, Antrobus R, Smith GL, Weekes MP. Quantitative Temporal Proteomic Analysis of Vaccinia Virus Infection Reveals Regulation of Histone Deacetylases by an Interferon Antagonist. *Cell Rep*. 2019 May 7;27(6):1920-1933.e7.
59. Maróti Z, Tombácz D, Prazsák I, Moldován N, Csabai Z, Torma G, Balázs Z, Kalmár T, Dénes B, Snyder M, Boldogkői Z. Time-course transcriptome analysis of host cell response to poxvirus infection using a dual long-read sequencing approach. *BMC Res Notes*. 2021 Jun 24;14(1):239.
60. García-Muse T, Aguilera A. Transcription-replication conflicts: how they occur and how they are resolved. *Nat Rev Mol Cell Biol*. 2016 Sep;17(9):553-63.
61. Srivatsan A, Tehranchi A, MacAlpine DM, Wang JD. Co-orientation of replication and transcription preserves genome integrity. *PLoS Genet*. 2010 Jan 15;6(1):e1000810. doi: 10.1371/journal.pgen.1000810.
62. Brambati A, Colosio A, Zardoni L, Galanti L, Liberi G. Replication and transcription on a collision course: eukaryotic regulation mechanisms and implications for DNA stability. *Front Genet*. 2015 Apr 28;6:166. doi: 10.3389/fgene.2015.00166.

## TUDOMÁNYMETRIAI ADATOK

---

- Tudományos folyóiratcikkek száma: 69
  - teljes cikk, nemzetközi folyóiratban: 65
  - rövid közlemény, nemzetközi folyóiratban: 4
  - az utolsó 10 év tudományos teljes, lektorált folyóiratcikkeinek száma: 60
- Első és utolsó szerzős közlemények száma: 44
- Összesített impakt faktor: 324,144
- Első és utolsó szerzős közlemények impakt faktora: 182,366
- Az értekezés alapját képező közlemények impakt faktora: 186,983
- Hirsch index: 21
- Összes hivatkozás: 1225

## AZ ÉRTEKEZÉS ALAPJÁUL SZOLGÁLÓ KÖZLEMÉNYEK LISTÁJA

### AZ ÉRTEKEZÉS ALAPJÁUL SZOLGÁLÓ SAJÁT, TELJES TERJEDELMŰ KÖZLEMÉNYEK

34. Prazsák I, **Tombácz D\***, Fülöp Á, Torma G, Gulyás G, Dörmő Á, Kakuk B, McKenzie Spires L, Toth Z, Boldogkői Z: KSHV 3.0: a state-of-the-art annotation of the Kaposi's sarcoma-associated herpesvirus transcriptome using cross-platform sequencing. *mSystems* 2024. 11:e0100723. doi: 10.1128/msystems.01007-23.
33. Lee S-C, Naik NG, **Tombácz D**, Gulyás G, Kakuk B, Boldogkői Z, Hall K, Papp B, Boulant S, Toth Z: Hypoxia and HIF-1 $\alpha$  promote lytic *de novo* KSHV infection. *Journal of Virology* 2023. 97(11):e0097223. doi: 10.1128/jvi.00972-23.
32. **Tombácz D**, Torma G, Gulyás G, Fülöp Á, Dörmő Á, Prazsák I, Csabai Z, Mizik M, Hornyák Á, Zádori Z, Kakuk B, Boldogkői Z: Hybrid sequencing discloses unique aspects of the transcriptomic architecture in equid alphaherpesvirus 1. *Heliyon* 2023. 9(7):e17716. doi: 10.1016/j.heliyon.2023.e17716.
31. **Tombácz D**, Dörmő Á, Gulyás G, Csabai Z, Prazsák I, Kakuk B, Harangozó Á, Jankovics I, Dénes B, Boldogkői Z: High temporal resolution Nanopore sequencing dataset of SARS-CoV-2 and host cell RNAs. *Gigascience* 2022. 11:giac094. doi: 10.1093/gigascience/giac094.
30. Prazsák I, Csabai Z, Torma G, Papp H, Földes F, Kemenesi G, Jakab F, Gulyás G, Fülöp Á, Megyeri K, Dénes B, Boldogkői Z, **Tombácz D**<sup>✉</sup>: Transcriptome dataset of six human pathogen RNA viruses generated by nanopore sequencing. *Data in Brief* 2022. 43:108386. doi: 10.1016/j.dib.2022.108386.
29. **Tombácz D**, Kakuk B, Torma G, Csabai Z, Gulyás G, Tamás V, Zádori Z, Jefferson VA, Meyer F, Boldogkői Z: In-Depth Temporal Transcriptome Profiling of an Alphaherpesvirus Using Nanopore Sequencing. *Viruses* 2022. 14(6):1289. doi: 10.3390/v14061289.
28. **Tombácz D**, Prazsák I, Torma G, Csabai Z, Balázs Z, Moldován N, Dénes B, Snyder M, Boldogkői Z: Time-Course Transcriptome Profiling of a Poxvirus Using Long-Read Full-Length Assay. *Pathogens* 2021. 10(8):919. doi: 10.3390/pathogens10080919.
27. Kakuk B, **Tombácz D\***, Balázs Z, Moldován N, Csabai Z, Torma G, Megyeri K, Snyder M, Boldogkői Z: Combined nanopore and single-molecule real-time sequencing survey of human betaherpesvirus 5 transcriptome. *Scientific Reports* 2021. 11(1):14487. doi: 10.1038/s41598-021-93593-y.
26. Maróti Z, **Tombácz D\***, Moldován N, Torma G, Jefferson VA, Csabai Z, Gulyás G, Dörmő Á, Boldogkői M, Kalmár T, Meyer F, Boldogkői Z: Time course profiling of host cell response to herpesvirus infection using nanopore and synthetic long-read transcriptome sequencing. *Scientific Reports* 2021. 11(1):14219. doi: 10.1038/s41598-021-93142-7.
25. Maróti Z, **Tombácz D\***, Prazsák I, Moldován N, Csabai Z, Torma G, Balázs Z, Kalmár T, Dénes B, Snyder M, Boldogkői Z: Time-course transcriptome analysis of host cell response to poxvirus infection using a dual long-read sequencing approach. *BMC Research Notes* 2021. 14(1):239. doi: 10.1186/s13104-021-05657-x.
24. **Tombácz D**, Moldován N, Torma G, Nagy T, Hornyák Á, Csabai Z, Gulyás G, Boldogkői M, Jefferson VA, Zádori Z, Meyer F, Boldogkői Z: Dynamic Transcriptome Sequencing of Bovine

- Alphaherpesvirus Type 1 and Host Cells Carried Out by a Multi-Technique Approach. *Frontiers in Genetics* 2021. 12:619056. doi: 10.3389/fgene.2021.619056.
23. Moldován N, Torma G, Gulyás G, Hornyák Á, Zádori Z, Jefferson VA, Csabai Z, Boldogkői M, **Tombácz D**, Meyer F, Boldogkői Z: Time-course profiling of bovine alphaherpesvirus 1.1 transcriptome using multiplatform sequencing. *Scientific Reports* 2020. 10(1):20496. doi: 10.1038/s41598-020-77520-1.
22. **Tombácz D**, Prazsák I, Csabai Z, Moldován N, Dénes B, Snyder M, Boldogkői Z: Long-read assays shed new light on the transcriptome complexity of a viral pathogen. *Scientific Reports* 2020. 10(1):13822. doi: 10.1038/s41598-020-70794-5.
21. **Tombácz D**, Torma G, Gulyás G, Moldován N, Snyder M, Boldogkői Z: Meta-analytic approach for transcriptome profiling of herpes simplex virus type 1. *Scientific Data*. 2020. 7(1):223. doi: 10.1038/s41597-020-0558-8.
20. **Tombácz D**, Moldován N, Balázs Z, Gulyás G, Csabai Z, Boldogkői M, Snyder M, Boldogkői Z: Multiple Long-Read Sequencing Survey of Herpes Simplex Virus Dynamic Transcriptome. *Frontiers in Genetics*. 2019. 10:834. doi: 10.3389/fgene.2019.00834.
19. Boldogkői Z, Moldován N, Balázs Z, Snyder M, **Tombácz D**: Long-Read Sequencing - A Powerful Tool in Viral Transcriptome Research. *Trends in Microbiology* 2019. 27(7):578-592. doi: 10.1016/j.tim.2019.01.010.
18. Boldogkői Z, **Tombácz D**, Balázs Z: Interactions between the transcription and replication machineries regulate the RNA and DNA synthesis in the herpesviruses. *Virus Genes* 2019. 55(3):274-279. doi: 10.1007/s11262-019-01643-5.
17. Boldogkői Z, Balázs Z, Moldován N, Prazsák I, **Tombácz D**: Novel classes of replication-associated transcripts discovered in viruses. *RNA Biology* 2019. 16(2):166-175. doi: 10.1080/15476286.2018.1564468.
16. Boldogkői Z, Moldován N, Szűcs A, **Tombácz D**<sup>✉</sup>: Transcriptome-wide analysis of a baculovirus using nanopore sequencing. *Scientific Data* 2018. 5:180276. doi: 10.1038/sdata.2018.276.
15. Boldogkői Z, Szűcs A, Balázs Z, Sharon D, Snyder M, **Tombácz D**<sup>✉</sup>: Transcriptomic study of Herpes simplex virus type-1 using full-length sequencing techniques. *Scientific Data* 2018. 5:180266. doi: 10.1038/sdata.2018.266.
14. **Tombácz D**, Prazsák I, Szucs A, Dénes B, Snyder M, Boldogkői Z: Dynamic transcriptome profiling dataset of vaccinia virus obtained from long-read sequencing techniques. *Gigascience* 2018. 7(12):giy139. doi: 10.1093/gigascience/giy139.
13. **Tombácz D**, Balázs Z, Csabai Z, Snyder M, Boldogkői Z: Long-Read Sequencing Revealed an Extensive Transcript Complexity in Herpesviruses. *Frontiers in Genetics* 2018. 9:259. doi: 10.3389/fgene.2018.00259.
12. **Tombácz D**, Sharon D, Szűcs A, Moldován N, Snyder M, Boldogkői Z: Transcriptome-wide survey of pseudorabies virus using next- and third-generation sequencing platforms. *Scientific Data* 2018. 5:180119. doi: 10.1038/sdata.2018.119.
11. Moldován N, **Tombácz D**<sup>\*</sup>, Szűcs A, Csabai Z, Snyder M, Boldogkői Z: Multi-Platform Sequencing Approach Reveals a Novel Transcriptome Profile in Pseudorabies Virus. *Frontiers in Microbiology* 2018. 8:2708. doi: 10.3389/fmicb.2017.02708.

10. **Tombácz D**, Prazsák I, Moldován N, Szűcs A, Boldogkői Z: Lytic Transcriptome Dataset of Varicella Zoster Virus Generated by Long-Read Sequencing. *Frontiers in Genetics* 2018. 9:460. doi: 10.3389/fgene.2018.00460.
9. Moldován N, Szucs A, **Tombácz D**, Balázs Z, Csabai Z, Snyder M, Boldogkői Z: Multiplatform next-generation sequencing identifies novel RNA molecules and transcript isoforms of the endogenous retrovirus isolated from cultured cells. *FEMS Microbiology Letters* 2018. 365(5):fny013. doi: 10.1093/femsle/fny013.
8. **Tombácz D**, Csabai Z, Szűcs A, Balázs Z, Moldován N, Sharon D, Snyder M, Boldogkői Z: Long-Read Isoform Sequencing Reveals a Hidden Complexity of the Transcriptional Landscape of Herpes Simplex Virus Type 1. *Frontiers in Microbiology* 2017. 8:1079. doi: 10.3389/fmicb.2017.01079.
7. Csabai Z, Takács IF, Snyder M, Boldogkői Z, **Tombácz D**✉. Evaluation of the impact of ul54 gene-deletion on the global transcription and DNA replication of pseudorabies virus. *Archives of Virology* 2017. 162(9):2679-2694. doi: 10.1007/s00705-017-3420-3.
6. Moldován N, Balázs Z, **Tombácz D**, Csabai Z, Szűcs A, Snyder M, Boldogkői Z: Multi-platform analysis reveals a complex transcriptome architecture of a circovirus. *Virus Research* 2017. 237:37-46. doi: 10.1016/j.virusres.2017.05.010.
5. **Tombácz D**, Balázs Z, Csabai Z, Moldován N, Szűcs A, Sharon D, Snyder M, Boldogkői Z: Characterization of the Dynamic Transcriptome of a Herpesvirus with Long-read Single Molecule Real-Time Sequencing. *Scientific Reports* 2017. 7:43751. doi: 10.1038/srep43751.
4. **Tombácz D**, Csabai Z, Oláh P, Balázs Z, Likó I, Zsigmond L, Sharon D, Snyder M, Boldogkői Z: Full-Length Isoform Sequencing Reveals Novel Transcripts and Substantial Transcriptional Overlaps in a Herpesvirus. *PLoS One* 2016. 11(9):e0162868. doi: 10.1371/journal.pone.0162868.
3. **Tombácz D**, Csabai Z, Oláh P, Havelda Z, Sharon D, Snyder M, Boldogkői Z: Characterization of novel transcripts in pseudorabies virus. *Viruses* 2015. 7(5):2727-44. doi: 10.3390/v7052727.
2. **Tombácz D**, Tóth JS, Boldogkői Z: Effects of deletion of the early protein 0 gene of pseudorabies virus on the overall viral gene expression. *Gene* 2012. 493(2):235-42. doi: 10.1016/j.gene.2011.11.049.
1. **Tombácz D**, Tóth JS, Boldogkői Z: Deletion of the virion host shut: off gene of pseudorabies virus results in selective upregulation of the expression of early viral genes in the late stage of infection. *Genomics* 2011. 98(1):15-25. doi: 10.1016/j.ygeno.2011.03.010.

#### SAJÁT, AZ ÉRTEKEZÉSHEZ KAPCSOLÓDÓ RÖVID KÖZLEMÉNYEK

4. Prazsák I, **Tombácz D**, Szűcs A, Dénes B, Snyder M, Boldogkői Z: Full Genome Sequence of the Western Reserve Strain of Vaccinia Virus Determined by Third-Generation Sequencing. *Genome Announcements* 2018. 6(11):e01570-17. doi: 10.1128/genomeA.01570-17.
3. Szűcs A, Moldován N, **Tombácz D**, Csabai Z, Snyder M, Boldogkői Z: Long-Read Sequencing Reveals a GC Pressure during the Evolution of Porcine Endogenous Retrovirus. *Genome Announcements* 2017. 5(40):e01040-17. doi: 10.1128/genomeA.01040-17.
2. **Tombácz D**, Moldován N, Balázs Z, Csabai Z, Snyder M, Boldogkői Z: Genetic Adaptation of Porcine Circovirus Type 1 to Cultured Porcine Kidney Cells Revealed by Single-Molecule Long-Read Sequencing Technology. *Genome Announcements* 2017. 5(5):e01539-16. doi: 10.1128/genomeA.01539-16.



1. **Tombácz D**, Sharon D, Oláh P, Csabai Z, Snyder M, Boldogkői Z: Strain Kaplan of Pseudorabies Virus Genome Sequenced by PacBio Single-Molecule Real-Time Sequencing Technology. *Genome Announcements* 2014. 2(4):e00628-14. doi: 10.1128/genomeA.00628-14.

#### A DOLGOZAT TÉMÁJÁHOZ KAPCSOLÓDÓ EGYÉB SAJÁT, TELJES TERJEDELMŰ KÖZLEMÉNYEK:

21. Nagy GÁ, **Tombácz D\***, Praszák I, Csabai Z, Dörmő Á, Gulyás G, Kemenesi G, Tóth GE, Holubek J, Růžek D, Kakuk B, Boldogkői Z: Exploring the Transcriptomic Profile of Human Monkeypox Virus via CAGE and Native RNA Sequencing Approaches. *mSphere* 2024 – bíráló alatt, preprint: *bioRxiv* 2024.04.29.591702; doi: <https://doi.org/10.1101/2024.04.29.591702>

20. Torma G, **Tombácz D\***, Csabai Z, Almsarrhad IAA, Nagy GÁ, Kakuk B, Gulyás G, Spires LM, Gupta I, Fülöp Á, Dörmő Á, Praszák I, Mizik M, Dani VÉ, Csányi V, Harangozó Á, Zádori Z, Toth Z, Boldogkői Z: Identification of herpesvirus transcripts from genomic regions around the replication origins. *Scientific Reports* 2023. 13(1):16395. doi: 10.1038/s41598-023-43344-y.

19. Kakuk B, Dörmő Á, Csabai Z, Kemenesi G, Holoubek J, Růžek D, Praszák I, Dani VÉ, Dénes B, Torma G, Jakab F, Tóth GE, Földes FV, Zana B, Lanszki Z, Harangozó Á, Fülöp Á, Gulyás G, Mizik M, Kiss AA, **Tombácz D**, Boldogkői Z: In-depth Temporal Transcriptome Profiling of Monkeypox and Host Cells using Nanopore Sequencing. *Scientific Data* 2023. 10(1):262. doi: 10.1038/s41597-023-02149-4.

18. Torma G, **Tombácz D\***, Moldován N, Fülöp Á, Praszák I, Csabai Z, Snyder M, Boldogkői Z: Dual isoform sequencing reveals complex transcriptomic and epitranscriptomic landscapes of a prototype baculovirus. *Scientific Reports* 2022. 12(1):1291. doi: 10.1038/s41598-022-05457-8.

17. Fülöp Á, Torma G, Moldován N, Szenthe K, Bánáti F, Almsarrhad IAA, Csabai Z, **Tombácz D**, Minárovits J, Boldogkői Z: Integrative profiling of Epstein-Barr virus transcriptome using a multiplatform approach. *Virology Journal* 2022. 19(1):7. doi: 10.1186/s12985-021-01734-6.

16. Kakuk B, Kiss AA, Torma G, Csabai Z, Praszák I, Mizik M, Megyeri K, **Tombácz D**, Boldogkői Z: Nanopore Assay Reveals Cell-Type-Dependent Gene Expression of Vesicular Stomatitis Indiana Virus and Differential Host Cell Response. *Pathogens* 2021. 10(9):1196. doi: 10.3390/pathogens10091196.

15. Torma G, **Tombácz D\***, Csabai Z, Moldován N, Mészáros I, Zádori Z, Boldogkői Z: Combined Short and Long-Read Sequencing Reveals a Complex Transcriptomic Architecture of African Swine Fever Virus. *Viruses* 2021. 13(4):579. doi: 10.3390/v13040579.

14. Torma G, **Tombácz D\***, Csabai Z, Göbhardt D, Deim Z, Snyder M, Boldogkői Z: An Integrated Sequencing Approach for Updating the Pseudorabies Virus Transcriptome. *Pathogens* 2021. 10(2):242. doi: 10.3390/pathogens10020242.

13. Olasz F, **Tombácz D**<sup>✉</sup>, Torma G, Csabai Z, Moldován N, Dörmő Á, Praszák I, Mészáros I, Magyar T, Tamás V, Zádori Z, Boldogkői Z: Short and Long-Read Sequencing Survey of the Dynamic Transcriptomes of African Swine Fever Virus and the Host Cells. *Frontiers in Genetics* 2020. 11:758. doi: 10.3389/fgene.2020.00758.

12. Balázs Z, **Tombácz D**, Csabai Z, Moldován N, Snyder M, Boldogkői Z: Template-switching artifacts resemble alternative polyadenylation. *BMC Genomics*. 2019. 20(1):824. doi: 10.1186/s12864-019-6199-7.

11. Csabai Z, **Tombácz D**, Deim Z, Snyder M, Boldogkői Z: Analysis of the Complete Genome Sequence of a Novel, Pseudorabies Virus Strain Isolated in Southeast Europe. *Canadian Journal of Infectious Diseases and Medical Microbiology* 2019. 2019:1806842. doi: 10.1155/2019/1806842.
10. Prazsák I, Moldován N, Balázs Z, **Tombácz D**, Megyeri K, Szűcs A, Csabai Z, Boldogkői Z: Long-read sequencing uncovers a complex transcriptome topology in varicella zoster virus. *BMC Genomics* 2018. 19(1):873. doi: 10.1186/s12864-018-5267-8.
9. Balázs Z, **Tombácz D**, Szűcs A, Snyder M, Boldogkői Z: Dual Platform Long-Read RNA-Sequencing Dataset of the Human Cytomegalovirus Lytic Transcriptome. *Frontiers in Genetics* 2018. 9:432. doi: 10.3389/fgene.2018.00432.
8. Moldován N, **Tombácz D**, Szűcs A, Csabai Z, Balázs Z, Kis E, Molnár J, Boldogkői Z: Third-generation Sequencing Reveals Extensive Polycistronism and Transcriptional Overlapping in a Baculovirus. *Scientific Reports* 2018. 8(1):8604. doi: 10.1038/s41598-018-26955-8.
7. Balázs Z, **Tombácz D\***, Szűcs A, Snyder M, Boldogkői Z: Long-read sequencing of the human cytomegalovirus transcriptome with the Pacific Biosciences RSII platform. *Scientific Data* 2017. 4:170194. doi: 10.1038/sdata.2017.194.
6. Balázs Z, **Tombácz D\***, Szűcs A, Csabai Z, Megyeri K, Petrov AN, Snyder M, Boldogkői Z: Long-Read Sequencing of Human Cytomegalovirus Transcriptome Reveals RNA Isoforms Carrying Distinct Coding Potentials. *Scientific Reports* 2017. 7(1):15989. doi: 10.1038/s41598-017-16262-z.
5. Póka N, Csabai Z, Pásti E, **Tombácz D**, Boldogkői Z: Deletion of the us7 and us8 genes of pseudorabies virus exerts a differential effect on the expression of early and late viral genes. *Virus Genes* 2017. 53(4):603-612. doi: 10.1007/s11262-017-1465-8.
4. Oláh P, **Tombácz D\***, Póka N, Csabai Z, Prazsák I, Boldogkői Z: Characterization of pseudorabies virus transcriptome by Illumina sequencing. *BMC Microbiology* 2015. 15:130. doi: 10.1186/s12866-015-0470-0.
3. Takács IF, **Tombácz D**, Berta B, Prazsák I, Póka N, Boldogkői Z: The ICP22 protein selectively modifies the transcription of different kinetic classes of pseudorabies virus genes. *BMC Molecular Biology* 2013 Jan 29;14:2. doi: 10.1186/1471-2199-14-2.
2. Tóth JS, **Tombácz D\***, Takács IF, Boldogkői Z: The effects of viral load on pseudorabies virus gene expression. *BMC Microbiology* 2010. 10:311. doi: 10.1186/1471-2180-10-311.
1. Márton G, **Tombácz D**, Tóth JS, Szabó A, Boldogkői Z, Dénes A, Hornyák A, Nógrádi A: Ex vivo infection of human embryonic spinal cord neurons prior to transplantation into adult mouse cord. *BMC Neuroscience* 2010. 11:65. doi: 10.1186/1471-2202-11-65.

#### PhD tézishez felhasznált, a témához kapcsolódó saját közlemény

1. **Tombácz D**, Tóth JS, Petrovszki P, Boldogkői Z: Whole-genome analysis of pseudorabies virus gene expression by real-time quantitative RT-PCR assay. *BMC Genomics* 2009. 10:491. doi: 10.1186/1471-2164-10-491.

#### EGYÉB, AZ ÉRTEKEZÉSTŐL FÜGGETLEN SAJÁT PUBLIKÁCIÓK

9. Kakuk B, Gulyás G, Dörmő Á, Járay T, Asaduzzaman M, Taifi A, Csabai Z, Boldogkői Z, **Tombácz D**<sup>✉</sup>: Ultra-deep, hybrid metagenomic sequencing of canine gut microbiome. *Scientific Data* 2024. – bírálat alatt

8. Gulyás G, Kakuk B, Dörmő Á, Járay T, Praszák I, Csabai Z, Henkrich MM, Boldogkői Z, **Tombácz D**: Cross-Comparison of Gut Metagenomic Profiling Strategies. *Communications Biology* 2024. – bírálattal alatti, preprint: *bioRxiv* 2023.11.25.568646; doi: 10.1101/2023.11.25.568646
7. Hardwick SA, Hu W, Joglekar A, Fan L, Collier PG, Foord C, Balacco J, Lanjewar S, Sampson MM, Koopmans F, Prjibelski AD, Mikheenko A, Belchikov N, Jarroux J, Lucas AB, Palkovits M, Luo W, Milner TA, Ndhlovu LC, Smit AB, Trojanowski JQ, Lee VMY, Fedrigo O, Sloan SA, **Tombácz D**, Ross ME, Jarvis E, Boldogkői Z, Gan L, Tilgner HU: Single-nuclei isoform RNA sequencing unlocks barcoded exon connectivity in frozen brain tissue. *Nature Biotechnology* 2022. 40(7):1082-1092. doi: 10.1038/s41587-022-01231-3.
6. Boldogkői Z, Csabai Z, **Tombácz D**, Janovák L, Balassa L, Deák Á, Tóth PS, Janáky C, Duda E, Dékány I: Visible Light-Generated Antiviral Effect on Plasmonic Ag-TiO<sub>2</sub>-Based Reactive Nanocomposite Thin Film. *Frontiers in Bioengineering and Biotechnology* 2021. 9:709462. doi: 10.3389/fbioe.2021.709462.
5. Csáki Á, Köves K, Kiss AL, Röhlich P, Boldogkői Z, Vereczki V, Puskár Z, **Tombácz D**, Csabai Z: Pinealocytes can not transport neurotropic viruses. Pinealo-to-retinal connection in prepubertal rats originates from pineal neurons: Light and electron microscopic immunohistochemical studies. *Neuroscience Letters* 2021. Jan 23;744:135517. doi: 10.1016/j.neulet.2020.135517.
4. Csáki Á, Köves K, Boldogkői Z, **Tombácz D**, Tóth ZE: The Same Magnocellular Neurons Send Axon Collaterals to the Posterior Pituitary and Retina or to the Posterior Pituitary and Autonomic Preganglionic Centers of the Eye in Rats. *NeuroSci* 2021. 2(1):27-44. doi: 10.3390/neurosci2010002.
3. **Tombácz D**, Maróti Z, Kalmár T, Palkovits M, Snyder M, Boldogkői Z: Whole-exome sequencing data of suicide victims who had suffered from major depressive disorder. *Scientific Data* 2019. 6:190010. doi: 10.1038/sdata.2019.10.
2. Maróti Z, Boldogkői Z, **Tombácz D**, Snyder M, Kalmár T: Evaluation of whole exome sequencing as an alternative to BeadChip and whole genome sequencing in human population genetic analysis. *BMC Genomics* 2018. 19(1):778. doi: 10.1186/s12864-018-5168-x.
1. **Tombácz D**, Maróti Z, Kalmár T, Csabai Z, Balázs Z, Takahashi S, Palkovits M, Snyder M, Boldogkői Z: High-Coverage Whole-Exome Sequencing Identifies Candidate Genes for Suicide in Victims with Major Depressive Disorder. *Scientific Reports* 2017. 7(1):7106. doi: 10.1038/s41598-017-06522-3.

**PhD tézishez felhasznált, a jelen dolgozattól független témában megjelent saját publikációk**

3. Prorok J, Kovács PP, Kristóf AA, Nagy N, **Tombácz D**, Tóth JS, Ordög B, Jost N, Virág L, Papp JG, Varró A, Tóth A, Boldogkői Z: Herpesvirus-mediated delivery of a genetically encoded fluorescent Ca(2+) sensor to canine cardiomyocytes. *Journal of Biomedicine and Biotechnology* 2009. 2009:361795. doi: 10.1155/2009/361795.
2. Boldogkői Z, Balint K, Awatramani GB, Balya D, Buskamp V, Viney TJ, Lagali PS, Duebel J, Pásti E, **Tombácz D**, Tóth JS, Takács IF, Scherf BG, Roska B: Genetically timed, activity-sensor and rainbow transsynaptic viral tools. *Nature Methods* 2009. 6(2):127-30. doi: 10.1038/nmeth.1292.
1. Rezek O, Boldogkői Z, **Tombácz D**, Kovágó C, Gerendai I, Palkovits M, Tóth IE: Location of parotid preganglionic neurons in the inferior salivatory nucleus and their relation to the superior salivatory nucleus of rat. *Neuroscience Letters* 2008. 440(3):265-9. doi: 10.1016/j.neulet.2008.05.099.

## KÖSZÖNETNYILVÁNÍTÁS

---

Hálával tartozom **Prof. Dr. Boldogkői Zsoltnak**, az SZTE SZAOK OBI tanszékvezető egyetemi tanárának, aki PhD hallgató koromtól, témavezetőként megismertette velem a vírusok molekuláris világát. Köszönöm a bizalmat, hogy doktori fokozatszerzésemet követően közvetlen munkatársa lehettem, az elmúlt 18 évben nyújtott inspiráló tudományos közeget, a sok kiváló javaslatot, s az ötleteim megvalósításában, helyes irányba terelésében nyújtott segítségét. A támogatást mind szakmailag, mind emberileg. Naponta példát mutat hihetetlen munkatempójával, s mindezt baráti, s általában „laza” hangulatban, melyért külön hálás vagyok.

Köszönet illeti **Prof. Dr. Michael Snyder**t a Stanford Egyetem (Stanford, CA, USA) Genetika Tanszék vezető professzorát, akinek révén elsajátíthattam a hosszú-read szekvenálás módszerét, akinél a PacBio kísérleteket és az Illumina szekvenálások egy részét elvégezhettem. Köszönet illeti azért is, mert az extrém elfoglaltsága ellenére gyakran biztosított számomra is szakmai, ugyanakkor közvetlen hangulatú beszélgetést, s lehetőséget adott a legtöbb ötletünk, így a vírusokkal kapcsolatos kutatásaink mellett a humán genomikai projektjeink kivitelezéséhez is. Köszönet illeti **Dr. Alexander Urbant** (Stanford Egyetem) és **Dr. Hagen Tilgnert** (Cornell Egyetem, New York, NY), akiknél kollaborációs partnerként dolgozhattam. Köszönet a Snyder-, Urban- és Tilgner-laborok (ahol két évet tölthettem vendégkutatóként) munkatársainak, (elsősorban **Dr. Donald Sharon**, **Dr. Brian Piening**, **Dr. Shinichi Takahashi**, **Dr. Julien Jarroux**, **Yi 'Francis' He**) akik a legújabb genomikai technikák elsajátításában és a felmerülő logisztikai problémák megoldásában segítettek.

Projektjeinkhez további kollaborációs partnerek is hozzájárultak, ezúton köszönöm nekik: **Prof. Dr. Palkovits Miklós** (Semmelweis Egyetem), **Prof. Dr. Jakab Ferenc** és **Dr. Kemenesi Gábor** (PTE), **Dr. Zádori Zoltán** (ÁTKI), **Dr. Dénes Béla** (NÉBIH), **Dr. Maróti Zoltán** és **Dr. Kalmár Tibor** (SZTE SZAOK Gyermekgyógyászati Klinika), **Dr. Megyeri Klára** (SZTE SZAOK Mikrobiológiai Intézet), **Prof. Dr. Dékány Imre** (SZTE TTIK Fizikai Kémiai és Anyagtudományi Intézet), **Prof. Dr. Varró András** (SZTE SZAOK Farmakológiai és Farmakoterápiás Intézet), **Prof. Dr. Roska Botond** (FMI/IOB Bázél, Svájc), **Dr. Florencia Meyer** (Mississippi State University), **Dr. Zsolt Toth** (University of Florida).

Hálás köszönettel tartozom a közvetlen jelenlegi (**Dr. Csabai Zsolt**, **Dr. Kakuk Balázs**, **Dr. Prazsák István**, **Dr. Torma Gábor**, **Dörmő Ákos**, **Gulyás Gábor**) és egykori (**Dr. Balázs Zsolt**, **Dr. Moldován Norbert**, **Dr. Oláh Péter**, **Dr. Tóth Judit**) munkatársaimnak, PhD hallgatóinknak a sokéves, kitűnő munkájukért, mely alapvető jelentőségű volt a dolgozatban bemutatott eredmények eléréséhez.

Köszönöm az SZTE SZAOK OBI 3G és az MTA-SZTE Lendület GeMiNI Csoportok mindenkori munkatársainak (**Dr. Fülöp Ádám**, **Dr. Kiss András**, **Dr. Ördög Balázs**, **Dr. Póka Nándor**, **Dr. Szűcs Attila**, **Dr. Takács Irma**, **Csányi Viktor**, **Dani Virág**, **Harangozó Ákos**, **Henkrich Miksa**, **Járay Tamás**, **Mizik Máté**, **Nagy G. Ármin**, **Petrovszki Pál**, **Dr. Islam Almsarrhad**, **Md Asaduzzaman**, **Ahmed Taifi**), hogy hozzátettek egy-egy szeletet az elért eredményeinkhez. Köszönet minden kollégámnak a barátságos atmoszféréért. Külön köszönöm **Jász Anikónak** a rengeteg adminisztratív teendő elvégzésében nyújtott segítségét, **Magyarné Papdi Csillának** a labortechnikai segítséget és a sok nevetést, **Bíbor Bettynek** a 'mindennapos', míg **Prof. Dr. Duda Ernőnek** a szakmai diskurzusokat.

Köszönet illeti **szüleimet** és **öcsémet**, akik első pillanattól kezdve támogattak a hatéves koromban megfogalmazott céloom 'biológus leszek' megvalósításában, ill. Totit és TBGt. Rengeteget segítettek az elmúlt évtizedekben. Köszönöm **A barátomnak** a hosszú, évtizedes támogatását az élet minden kalandjában és kihívásában. Az ő bennem való hite is kellett ahhoz, hogy most itt álljak.

## TÁMOGATÓK

---

Az értekezés alapját képező kutatásainkat, publikációink megjelenését és a munkámat támogatták:

- a Szegedi Tudományegyetem
- a University of Szeged Open Access Fund
- a Human Frontier Science Program
  - Young Investigator Grant (RGY0073/2006; *Boldogkői Zsolt*)
- a Magyar Tudományos Akadémia
  - Bolyai János Kutatási Ösztöndíj x2 (*Tombácz Dóra*)
  - Lendület Pályázat (LP2020-8/2020; *Tombácz Dóra*)
- a Nemzeti Kutatási Fejlesztési és Innovációs Hivatal
  - Fiatal kutatói kiválósági program (FK142676; *Tombácz Dóra*)
  - Fiatal kutató által kezdeményezett témapályázat (FK128252; *Tombácz Dóra*)
  - Kutatási témapályázat (K128247, K142674; *Boldogkői Zsolt*)
- a Svájci–Magyar Együttműködési Program (SH/7/2/8; *Boldogkői Zsolt*)
- a Nemzeti Kiválóság Program
  - Magyar Zoltán Posztdoktori Ösztöndíj (*Tombácz Dóra*)
- a Nemzeti Tudósképző Akadémia
- a Campus Hungary Program (*Tombácz Dóra*)
- a Tempus Közalapítvány (*Tombácz Dóra*)
- a Társadalmi Megújulás Operatív Program (TÁMOP)
- a Társadalmi Infrastruktúra Operatív Program (TIOP)
- a Magyar Genetikusok Egyesülete
- a Magyar Mikrobiológiai Társaság
- a National Institutes of Health (NIH CEGS Center for Personal Dynamic Regulomes [5P50HG00773502]; *Michael Snyder*)
- a Stanford Egyetem (Stanford University)
- a Weill Cornell Medicine