

Dora_213_24

MTA DOKTORI ÉRTEKEZÉS

A VIRÁLIS TRANZKRIPTOM REJTETT KOMPLEXITÁSA - INTEGRATÍV GENOMIKAI MEGKÖZELÍTÉSEK –



TOMBÁ CZ DÓRA

SZEGEDI TUDOMÁNYEGYETEM
SZENT-GYÖRGYI ALBERT ORVOSTUDOMÁNYI KAR
ORVOSI BIOLÓGIAI INTÉZET

SZEGED, 2024

1 Rövidítések jegyzéke	9
2 Előszó	12
3 Bevezetés	13
3.1 A vírusok	13
3.1.1 A vírusok jelentősége	13
3.1.2 A vírusok rendszertana	13
3.1.3 A transzkripció és a replikáció szabályozása	15
3.1.3.1 A transzkripció szabályozása	15
3.1.3.2 A DNS replikáció szabályozása	16
3.1.3.3 Az RNS-ek szerepe a replikáció szabályozásában	17
3.1.4 Herpeszvírusok	18
3.1.4.1 Humánpatogén alfaherpeszvírusok	19
3.1.4.2 Humánpatogén betaherpeszvírusok	19
3.1.4.3 Humánpatogén gammaherpeszvírusok	20
3.1.4.4 Állati alfaherpeszvírusok	20
3.1.5 Varidnaviriák	21
3.1.5.1 A Nucleocytoviricota: nukleo-citoplazmatikus nagy DNS-vírusok (NCLDV)	21
3.1.5.2 A bakulovírusok	22
3.1.6 A cirkovírusok	23
3.1.7 Ribovírusok	23
3.1.7.1 Orthornavirae -> Negarnoviricoták	24
3.1.7.2 Orthornavirae -> Kitrinoviricoták	25
3.1.7.3 Orthornavirae -> Pisuviricoták	26
3.1.7.4 Pararnavirae -> Artverviricoták	28
3.2 A szekvenálás forradalma	28
3.2.1 Az Illumina technológia elve	30
3.2.2 A PacBio technológia elve	31
3.2.3 A nanopóruson alapuló szekvenálás elve	32
3.2.4 A Loop-Seq technológia elve	33
3.2.5 A hosszú-readek szerepe a transzkriptomikában	34
3.3 Vírus transzkriptomok analízise az LRS-alapú kutatásaink előtt	35
4 Célkitűzések	36

5 Módszerek**37**

5.1 Felhasznált vírusok, alkalmazott sejtvonalak	38
5.2 Vírusfertőzés menete, időzítéses kísérletek	39
5.2.1 PRV Kaplan törzs	39
5.2.2 További PRV törzsek	40
5.2.3 Egyéb herpeszvírusok	40
5.2.4 További nagy, DNS genomú vírusok	41
5.2.5 RNS vírusok	42
5.3 A fehérjeszintézis gátlása	43
5.4 PRV virion és vírus DNS izolálás	43
5.5 RNS izolálás	43
5.5.1 PoliA szelekció / az RNS-ek 3' végének azonosítása	43
5.5.2 Ribodepléció / a teljes RNS frakció elemzése	43
5.5.3 Terminátor kezelés / az RNS-ek 5' végének azonosítása	43
5.5.4 Az RNS minták tárolása	44
5.6 Genomikai módszerek	44
5.6.1 Szekvenálás / könyvtárkészítés	44
5.6.1.1 Genom szekvenálás	44
5.6.1.2 Rövid-read RNS szekvenálás	44
5.6.1.3 Hosszú-read RNS szekvenálás	45
a. PacBio transzkriptom szekvenálás RSII és Sequel platformokon	45
i. nem amplifikált/ izoforma szekvenálás (IsoSeq)	45
ii. amplifikált IsoSeq	46
b. ONT transzkriptom szekvenálás MinION és Mk1C készülékekkel	46
i. amplifikált cDNS szekvenálás	46
ii. direkt cDNS (dcDNS) szekvenálás	47
iii. direkt RNS (dRNS) szekvenálás	47
iv. Cap-szelekció módosított könyvtárkészítéssel	47
v. Cap-szelekció Terminátor-kezeléssel	48
5.6.1.4 Szintetikus hosszú-read szekvenálás	48
5.6.2 Biszulfid konverzió	48
5.6.3 A könyvtárak tisztítása	48

5.6.4 A nukleinsavak minőségi és mennyiségi ellenőrzése	49
5.6.5 A könyvtárak méretszelektálása	49
5.6.5.1 Manuális gélizólálás	49
5.6.5.2 BluePippin™ méretszelekció	49
<i>5.7 Bioinformatika</i>	<i>50</i>
5.7.1 Genom összeállítás és annotáció	50
5.7.2 A transzkriptom analízishez használt legfontosabb programok	51
5.7.2.1 PacBio RSII és Sequel transzkriptom szekvenálás	51
5.7.2.2 ONT transzkriptom szekvenálás	51
5.7.2.3 Illumina RNA-Seq	52
5.7.2.4 Illumina CAGE-Seq	52
5.7.2.5 Illumina LoopSeq	52
5.7.2.6 Vizualizáció	52
5.7.2.7 A transzkript izoformák funkcionális elemzése	53
5.7.3 Statikus transzkriptom analízis	53
5.7.3.1 Térképezés és adatelemzés	53
5.7.4 Dinamikus transzkriptom analízis	55
5.7.4.1 A Pseudorabies vírus (PRV) gének kinetikai besorolása PacBio RSII-vel	55
5.7.4.2 A Vakcína vírus (VACV) gének relatív expressziós szintjeinek meghatározása PacBio Sequel és ONT MinION szekvenálás alapján	56
5.7.4.3 A Bovine alphaherpesvirus-1 (BoHV-1) gének expressziós analízise	56
5.7.4.4 A herpesz szimplex vírus-1 (HSV-1) gének expressziós analízise	56
5.7.4.5 A Vesicular stomatitis Indiana virus (VSIV) sejtspecifikus, dinamikus transzkriptom elemzése	57
5.7.5 RNS módosulások analízise	57
5.7.6 Vírusfertőzés a gazdagének expressziójára gyakorolt hatásának analízise	57
5.7.6.1 BoHV-1 – B. taurus sejtvonal	57
5.7.6.2 VACV - C. sabaesus sejtvonal	58
<i>5.8 Kvantitatív real-time PCR</i>	<i>58</i>
<i>5.9 Northern blot</i>	<i>59</i>
5.9.1 Hagyományos Northern blot	59
5.9.2 Mikro RNS (miRNS) Northern blot	60
<i>5.10 Mikroszkópos technikák</i>	<i>60</i>
<i>5.11 Géntechnológiai megközelítések</i>	<i>61</i>

5.11.1 Mutáns vírusok előállítás	61
5.11.1.1 PRV ul41Δ	61
5.11.1.2 PRV ep0Δ	61
5.11.1.3 PRV ul54Δ	62
5.12 Engedélyek	62
6 Eredmények	63
6.1 Technológiai újítások: „wetlab”, bioinformatika	63
6.1.1 Új technológiai megközelítések	63
6.1.1.1 Transzkript annotálás	63
6.1.1.2 Dinamikus transzkriptom vizsgálata	64
6.1.2 Bioinformatika	65
6.1.3 Génkiütött vírusok transzkriptomának analízise RT ² -PCR-rel	65
6.2 Vírus genomok annotálása: DNS és RNS szekvenciákból	65
6.2.1 Pseudorabies vírus (PRV)	65
6.2.1.1 PRV Kaplan törzs	65
6.2.1.2 PRV MdBio törzs	66
6.2.1.3 PRV ul54Δ	66
6.2.2 Porcine circovirus 1 (PCV-1)	67
6.2.3 Porcine Endogenous Retrovirus (PERV)	67
6.2.4 Vaccinia vírus (VACV)	67
6.2.5 SARS-CoV-2	67
6.3 Vírus transzkriptomok	69
6.3.1 Read-hosszak	69
6.3.2 A virális transzkriptom rejtett komplexitása	72
6.3.2.1 PRV	72
6.3.2.2 Egyéb herpeszvírusok	73
6.3.2.3 Egyéb vírusok	77
6.3.3 Ami a strukturális komplexitás mögött rejlik	79
6.3.3.1 Új kódoló gének, emRNS-ek	80
a. PRV ul36.5	80
b. A HSV emRNS-ek	81
c. Beágyazott gének a bakulovírus genomban	82

6.3.3.2 Nem-kódoló RNS-ek	82
a. Intergenikus transzkriptek	83
b. Beágyazott, nem kódoló RNS-ek	83
c. Antiszensz transzkriptek	83
6.3.3.3 Read-through RNS-ek	84
6.3.3.4 RNS izoformák	84
a. Transzkripció start hely (TSS) variánsok	84
b. Transzkripció stop (TES) variánsok	85
c. Splice variánsok	86
6.3.3.5 Transzkripció átfedés hálózatok	90
a. Parallel (tandem) átfedések	90
b. Konvergens átfedések	90
c. Divergens átfedések	91
d. Átfedések a távoli gének között	92
6.3.4 Funkció a strukturális komplexitás mögött	92
6.3.4.1 raRNS-ek: transzkriptek a replikációs origók közelében	93
a. Alfaherpeszvírusok—OriS	93
b. Alfaherpeszvírusok—OriL	95
c. Bétaherpeszvírusok	97
d. Gammaherpeszvírusok	97
6.3.4.2 raRNS-ek: a transzkripciót szabályozó gének RNS izoformái	97
6.3.4.3 A transzkripciót szabályozó gének közelébe térképeződő ncRNS-ek	98
6.3.4.4 A DNS replikációban részt vevő gének átfedései	98
6.3.4.5 A TSS-ek meghatározása	98
6.3.5 Dinamikus transzkriptom: a génexpresszió időbeli változásainak elemzése	99
6.3.5.1 A dinamikus PRV transzkriptom	99
6.3.5.2 A HSV-1 gének kinetikai analízise	101
6.3.5.3 A BoHV-1 ősszgenom-szintű génexpressziós változásainak vizsgálata	103
6.3.5.4 A vírusgén domének időfüggő kifejeződése	105
6.3.5.5 A VACV dinamikus transzkriptoma	107
6.3.5.6 A VSIV transzkriptom időbeli elemzése	108
6.3.6 Poszttranszkripció modifikációk elemzése	109
6.3.6.1 5mC metiláció – az epitranszkriptom vizsgálata	110

6.3.6.2 A-I hiperszerkesztés /hipereditálás	111
6.3.7 A géndelációk hatása a virális transzkriptomra	113
6.3.7.1 ul54Δ	113
6.3.7.2 ep0Δ	114
6.3.7.3 ul41Δ	115
6.4 A gazdasejt transzkriptomok és a vírusfertőzés a gazdagének expressziójára kifejtett hatásának analízise	117
6.4.1 A <i>Bos taurus</i> transzkriptom és BoHV-1 fertőzés hatására bekövetkező változásai	117
6.4.2 A VACV fertőzés hatása a CV-1 sejtek génexpressziós profiljára	122
7 Az adatok elérhetősége	124
7.1 Az általunk előállított szekvenciák elérhetősége	124
7.1.1 Transzkriptom	124
7.1.2 Genom	125
7.2 Az analízisünkhöz felhasznált, mások által közölt szekvenciák elérhetősége	125
8 Aktuális projektjeink, jövőbeni terveink	126
9 Legfontosabb eredményeink összegzése	127
10 Zárszó	134
11 Irodalomjegyzék	135
12 Publikációs lista	166
12.1 Az értekezés alapjául szolgáló saját, teljes terjedelmű közlemények	166
12.2 Saját, az értekezéshez kapcsolódó rövid közlemények	168
12.3 A dolgozat témájához kapcsolódó egyéb saját, teljes terjedelmű közlemények	169
12.3.1 PhD tézishoz felhasznált, a témához kapcsolódó saját közlemény	170
12.4 Egyéb, az értekezéstől független saját publikációk	170
12.4.1 PhD tézishoz felhasznált, a jelen dolgozattól független témában megjelent saját publikációk	171
13 Tudománymetriai adatok	172
14 Köszönetnyilvánítás	173
15 Támogatók	174

The microbe

The microbe is so very small
 You cannot make him out at all,
 But many sanguine people hope
 To see him through a microscope.
 His jointed tongue that lies beneath
 A hundred curious rows of teeth;
 His seven tufted tails with lots
 Of lovely pink and purple spots,
 On each of which a pattern stands,
 Composed of forty separate bands;
 His eyebrows of a tender green;
 All these have never yet been seen--
 But Scientists, who ought to know,
 Assure us that it must be so...
 Oh! let us never, never doubt
 What nobody is sure about!

Hilaire Pierre René Belloc (1896)

A mikróba

A mikróba oly csöppnyi, hogy
 szabad szemmel nem láthatod,
 viszont vérmes reményü had
 fürkészi mikroszkóp alatt.
 Ízelt nyelvét, mit közrefog
 száz sor kíváncsi, furcsa fog;
 hét bojtos farka pettyein
 a bájos bordó s rózsaszín
 mintát, negyven különböző
 rajz szerint; s lágy, lenyűgöző
 szemöldökét, mely zöldben ég,
 mindezt nem látta senki még –
 de száz Tudós készségesen
 állítja: így kell, hogy legyen...
 Ne kételkedj hát soha ott,
 hol senki sem bizonyított!

Csengery Kristóf fordítása



1 RÖVIDÍTÉSEK JEGYZÉKE

acDNS	amplifikált cDNS
AcMNPV	<i>Autographa californica</i> multiple nucleopolyhedrovirus
α HV	alfaherpeszvírus
AI	artificial intelligence (mesterséges intelligencia)
ASFV	African swine fever virus (Afrikai sertéspestis vírus)
AD	Aujeszky's disease (Aujeszky betegség)
asRNS	antiszensz RNS
ATCC	American Type Culture Collection
AyV	Aujeszky's disease virus or Pseudorabies virus (PRV; Aujeszky-féle vírus)
β HV	bétaherpeszvírusok
BoHV-1	Bovine alphaherpesvirus 1
bp	bázispár
CAGE	Cap analysis of gene expression
CAGE-Seq	CAGE szekvenálás
CCHFV	Crimean-Congo hemorrhagic fever virus (krími-kongói vérzéses láz vírusa)
CCS	Circular Consensus Sequence (cirkuláris konszenzus szekvencia)
CDS	coding sequence (kódoló szekvencia)
CHX	cikloheximid
CPE	cytopathic effect (citopátiás hatás)
CVB5	Coxsackievirus, group B serotype 5 (Coxsackie B5-vírus)
CV1	<i>Cercopithecus aethiops</i> monkey kidney cell line (majomvese sejtvonal)
cxRNS	complex RNS (komplex RNS)
dcDNS	direkt cDNS
DE	delayed early (késleltetett korai)
DMEM	Dulbecco's Modified Eagle's Medium (Dulbecco-féle módosított Eagle-tápközeg)
DNP	DNS polimeráz
dRNS	direkt RNS
ds	double-stranded (duplaszálú)
E	early (korai)
EBV	Epstein-Barr vírus
EHV	Equid alphaherpesvirus 1
E/L	early-late (korai-kései)
ENA	European Nucleotide Archive
ep0	early protein 0 (korai fehérje 0)
emRNS	embedded mRNA (beágyazott mRNA)
encRNS	embedded non-coding RNA (beágyazott nem-kódoló RNS)
eRNS	embedded RNA (beágyazott RNS)
FBS	Fetal Bovine Serum (magzati szarvasmarhaszérum)
FCS	Fetal Calf Serum (magzati brojú szérum)
fRNS	fúziós RNS
γ HV	gammaherpeszvírusok
GEO	Gene Expression Omnibus
GO	Gene Ontology (gén-ontológia)
gRNS	genomi RNS
HCMV	Human cytomegalovirus (humán citomegalovírus)

HSV-1	Herpes simplex virus type 1 (1-es típusú herpesz szimplex vírus)
HV	herpeszvírus
I	intermediate (köztes)
ICTV	International Committee on Taxonomy of Viruses
IE	immediate early (azonnali korai)
ifORF	in-frame ORF (kereten belüli ORF)
IGV	Integrative Genomics Viewer
IR	inverz ripít
IRS	internal repeat short (US régió belső ismétlődése)
Iso-Seq	isoform sequencing (izoforma szekvenálás)
IVA	Influenza A vírus
kbp	kilobázispár
KSHV	Kaposi's sarcoma-associated herpesvirus (Kaposi-szarkómához társuló herpeszvírus)
L	late (kései)
LAT	látencia-asszociált transzkript
LIS	late initiator sequence (kései iniciátor szekvencia)
LLT	long latency transcript (hosszú látencia transzkript)
lncRNS	long non-coding (hosszú nem-kódoló) RNS
LoRTIA	Long-read RNA-Seq Transcript Isoform Annotator
LRS	long-read sequencing (Hosszú-read/leolvasású szekvenálás)
L/ST	junction-spanning transcripts (junkció környéki transzkript)
MDBK	Madin–Darby Bovine Kidney (szarvasmarha vese sejtek)
MEM	Minimum Essential Medium, Eagle
miRNS	mikro RNS
mpox	monkeypox (majomhimlő)
MPXV	Monkeypox vírus (majomhimlő vírus)
MRC-5	human lung fibroblast cells (humán tüdő fibroblaszt sejtek)
ncRNS	non-coding RNA (nem kódoló RNS)
NCS	Newborn Calf Serum (újszülött borjú szérum)
NEB	New England Biolabs
NGS	next-generation sequencing (újgenerációs szekvenálás)
NSP	nonstructural protein (nem strukturális fehérje)
nt	nukleotid
OB	occlusion body
OBP	Ori binding protein (Ori-kötő fehérje)
ONT	Oxford Nanopore Technologies
ORC	origin recognition complex (replikációs origó felismerő komplex)
ORF	open reading frame (nyitott leolvasási keret)
OriL	origin of replication, unique long region (replikációs origó egyedi nagy régió)
OriS	origin of replication, unique short region (replikációs origó egyedi rövid régió)
PacBio	Pacific Biosciences
PA	poliadeniláció
PAM	porcine alveolar macrophage (sertés alveoláris makrofágok)
PAS	poliadenilációs szignál
PA-Seq	poliadenilált RNS-ek 3' végének szekvenálása

PBS	phosphate buffered saline (foszfáttal pufferelt sóoldat)
PCV-1	Porcine circovirus 1 (1-es típusú sertés cirkovírus)
PEL	KSHV-positive primary effusion lymphoma (PEL) cell line iBCBL1-3xFLAG-RTA102 (KSHV pozitív rekombináns sejtvonat)
PERV	Porcine endogenous retrovirus (sertés endogén retrovírus)
pfu	plaque forming unit (plakk-képző egység)
PK-15	Porcine-kidney 15 cell line (sertésvese sejtvonat)
PRV	Pseudorabies virus (Aujeszky-féle vírus; AyV)
raRNS	replication-associated RNS (replikáció asszociált RNS)
RdRp	RNA-dependent RNA polymerase (RNS-függő RNS polimeráz)
RdDp	RNA-dependent DNA polymerase (RNS-függő DNS polimeráz)
RK-13	Rabbit kidney epithelial cells (nyúlvese epitél sejtek)
RNP	RNS-polimeráz
ROI	Reads of Insert (leolvasások inszertje)
RSII	Real-Time Sequencer II
RT	Reverse transcription (reverz transzkripció)
rtRNA	read-through RNS (továbbíródó RNS)
SARS-CoV-2	Severe acute respiratory syndrome coronavirus 2 (súlyos akut légzőszervi szindrómát okozó koronavírus 2)
SBS	Sequencing by synthesis [(DNS) szintézis alapú szekvenálás]
Sf9	<i>Spodoptera frugiperda</i> cell line (Sf9 sejtvonat)
sgRNS	szubgenomi RNS
SLRS	Synthetic long-read sequencing (szintetikus hosszú-read szekvenálás)
SRS	Short-read sequencing (Rövid read/leolvasású szekvenálás)
ss	single-stranded (egyszálú)
ssDBP	single-stranded DNA binding protein (egy-szálú DNS-kötő fehérje)
TF	transcription factor (transzkripció faktor)
TGS	Third-generation sequencing (Harmadik generációs szekvenálás)
TI	transcript isoform (transzkript izoforma)
TO	transcriptional overlap (transzkripció átfedés)
TRS	terminal repeat sequence (US régió terminális ismétlődése)
TRT	transcriptional read-through (transzkripció továbbírás)
TS	template switching ('szálváltás')
UL	unique long (egyedi hosszú)
uORF	upstream ORF
US	unique short (egyedi rövid)
UTR	untranslated region (nem transzlálódó/leolvasódó régió)
VACV	Vaccinia virus (vakcína vírus)
VHS	Virion host shut off
VL	very late (nagyon kései)
VSIV	Vesicular stomatitis Indiana virus (vezikuláris sztomatitisz vírus, Indiana szerotípus)
VZV	Varicella Zoster vírus
WHO	World Health Organization (Egészségügyi Világszervezet)
WNV	West Nile virus (nyugat-nílusi vírus)
WR	Western Reserve [(strain of VACV) a vakcína vírus Western Reserve törzse]
ZIKV	Zika vírus

2 ELŐSZÓ

Molekuláris genetikával a Szegedi Tudományegyetem (SZTE) Általános (ma Szent-Györgyi Albert) Orvostudományi Kar (SZAOK) Orvosi Biológiai Intézetben (OBI) 2006 őszén, PhD hallgatóként kezdtem el foglalkozni Dr. Boldogkői Zsolt Genomika és Géntechnológia (3G) csoportjában. Az Aujeszky-féle betegséget (AD) okozó állati patogén alfaherpeszvírust (Aujeszky-féle vagy Pseudorabies vírus; AyV, PRV) használtuk idegi aktivitás mérésére, mint idegi nyomjelző és mint génbeviteli vektor¹, ehhez különféle genetikailag módosított mutáns vírusokat állítottunk elő. A gének kifejeződésének tanulmányozására elsőként a PRV-t, mint modell organizmust használtuk: a vírus vad típusú Kaplan (Ka) törzse mellett számos mutáns PRV génexpressziós kinetikáját is meghatároztuk reverz transzkripciót (RT) követő Real-Time kvantitatív PCR-ral (qRT-PCR), mely eredményeként a vírus protein kódoló génjeinek komplementer száláról is detektáltunk bizonyos mértékű antiszensz (as)RNS kifejeződést². Eredményeink szerint a vírus örökítő anyagának jelentős része - hasonlóan az emberi genomhoz - transzkripcionálisan aktív, mindkét DNS szálon. A PRV teljes transzkriptomot második generációs szekvenálással (NGS) is meghatároztuk³.

2013-tól a Covid-19 pandémia okozta lezárásokig hét éven keresztül kisebb blokkokban, összesen két évet a Stanford Egyetem Genetika Tanszékén (Snyder-lab) dolgoztam, ahol lehetőségem volt a Pacific Biosciences (PacBio) harmadik generációs (TGS), hosszú-read szekvenáló (LRS) platformjain (korábban az RSII-vel, majd a Sequellel), hazai projektjeinken dolgozni. E módszerrel megszekvenáltunk számos (elsősorban herpesz) vírusgenomot^{4,5} és transzkriptomot⁶⁻⁹. Ezzel egyrészt úttörők lettünk a vírus transzkriptomikában, továbbá hazánkban elsőként alkalmaztunk TGS módszert.

A Snyder laborban lehetőségem volt a depresszió-öngyilkosság genetikáját feltáró, szintén hazai projektünkön is dolgoznom (Illumina teljes exom szekvenálás), melynek eredményeként a TGF- β jelút, a depresszió-öngyilkosságban betöltött potenciális szerepét találtuk^{10,11}. A projekt a Cornell Egyetemmel (Tilgner-lab) való kollaborációnk keretein belül folytatódik, ahol a depresszió és az Alzheimer-kór kutatását LRS alapú, egysejt transzkriptomikai megközelítéssel célozzuk meg¹².

Az Oxford Nanopore Technologies (ONT) nanopórusos TGS módszerének megjelenésekor azonnal bevezettük¹³ a hazai laborunkban a vírusok (különös **hangsúlyt** fektettünk **a herpeszvírusokra**) RNS profiljának analízise céljából. Valójában a **TGS módszereket** olyannyira hamar vezettük be, hogy saját transzkript annotáló programcsomagot is ki kellett dolgozzunk^{14,15}. A TGS technikának rengeteg előnye van minden korábbi módszerrel szemben, **a transzkriptomok** minden várakozást felülmúló **komplexitását** tártuk fel az általunk vizsgált összes vírusban¹⁶⁻²⁶. **Erről szól e disszertáció.**

A kutatói életpálya során számos fontos akadályt kell megugranunk, egyre magasabb falakat kell megmászunk. A PhD és az MTA doktori fokozatok megszerzése két ilyen jelentős mérföldkő. Mindkét esetben összefoglaló dolgozatot írunk az addig elért eredményeinkről. Ezeket az eredményeket, bár egyénként érzük, s saját sikerként könyveljük el, nagyon fontos megemlíteni, hogy a kutatói szakma (legalábbis természettudományoké) nem egyéni sport, hanem csapatjáték.

A disszertációban bemutatott eredményeken számos kollégám dolgozott hosszabb-rövidebb időn át velem az elmúlt közel két évtizedben. Eredményeink egymásra épülnek, sok esetben egymástól nehezen, vagy nem is elkülöníthetők. A 'Köszönetnyilvánítás' fejezet javarészt nekik szól.

„The combined results of several people working together is often much more effective than could be that of an individual scientist working alone.” John Bardeen

3 BEVEZETÉS

A disszertációban a 2011-2024 közti időszakban megjelent publikációink eredményeinek széles spektrumát tárgyalom. Adott témakörön, publikáción belül a saját munkám bemutatására fókuszáltam elsősorban, mely azonban sok esetben munkatársaim munkájával együttesen diszkutálható. A publikációkban a szerzői hozzájárulások megtekinthetők. E munkához kizárólag saját ábrákat használtam, melyek közül, ha valamely publikációnk részét képezik, azt az ábraleírásukban hivatkoztam. A többi ábrát e doktori értekezés illusztrálásához készítettem, két kép pedig a mesterséges intelligencia (AI) műve.

3.1 A VÍRUSOK

A dolgozatban főként a herpeszvírusokkal kapcsolatos genomikai eredményeinkre fókuszálok, azonban számos egyéb vírussal is foglalkoztunk, s bizonyos transzkriptomikai jellegzetességeket rájuk hivatkozva tárgyalok majd. Így a bevezetés első szakaszában az e fajokra vonatkozó, aktuális rendszertani áttekintést nyújtok, tárgyalom jelentőségüket és genomi felépítésüket.

3.1.1 A vírusok jelentősége

A vírusok obligát paraziták, azaz a gazdasejten kívül nem képesek szaporodni. Minden sejtes életforma rendelkezik vírus kórokozókkal, ideértve a baktériumokat, gombákat, növényeket, állatokat és az embert. E szervezetek nagy károkat képesek okozni a mezőgazdaságban és az emberi egészségben, de a vadon élő fajok és a házi kedvencek körében is. A vírusok hatását a világra az elmúlt években jól megfigyelhettük a COVID-19 világjárvány során, amely világszerte jelentős társadalmi és gazdasági hatásokat váltott ki. A vírusok, különösen a pandémiás potenciállal rendelkező vírusok, komoly hatással lehetnek a világra és komplex válaszokat igényelnek mind az egészségügyi, mind a társadalmi és gazdasági rendszerek részéről. Létezik számos vírus, mely egyértelműen mutualisztikus: egyesek létfontosságúak a gazdaszervezet túléléséhez, míg mások például versenyelőnyt adnak a gazdaszervezetnek a természet versengő világában²⁷.

A vírusok a tudományos kutatásban pozitív szerepet töltenek be, hiszen relatíve kisméretű és egyszerű genomjuk és gyors reprodukciós rátájuk kiváló modellszervezetté teszi őket. Például, bakteriofágok segítségével igazolták, hogy a DNS az örökítőanyag²⁸. Továbbá, vírusokat használnak génbeviteli²⁹⁻³¹ és génterápiás vektorokként³²⁻³⁵ és idegpálya jelölésre^{1,36,37}, valamint vakcinafejlesztésre^{38,39} és immunterápiás eszközként is⁴⁰⁻⁴³. A bakteriofágok pedig nagy potenciállal rendelkeznek az antibiotikum rezisztens baktériumok elleni védekezésben⁴⁴⁻⁴⁷, illetve a fágterápia révén a bélrendszeri mikrobiom helyreállításának klinikai alternatívájaként is⁴⁸.

3.1.2 A vírusok rendszertana

A vírusok polifiletikus eredetűek⁴⁹. A nagy DNS genommal rendelkező vírusok feltehetően baktériumsejtek leegyszerűsödésével jöttek létre, míg a kisebb RNS vírusok a gazdasejt nukleinsavaiból 'barkácsoltak' komplex struktúrákat. A vírusok rendszerezésének fő problémája, hogy a rendkívül gyors evolúciójuk elmosza a távoli rokonok közötti genetikai hasonlóságokat. Emiatt a vírusok taxonómiája eltér az összes többi szervezet rendszertanától: míg a sejttel rendelkező élőlényeket a genetikai rokonság alapján osztályozzák, a vírusok osztályozása más

logikát követ. A vírusok rendszertana dinamikus jellegű, folyamatosan változik, ahogy új vírusokat fedeznek fel, és ahogy a genetikai és biológiai ismeretek bővülnek, szekvenálási technológiák fejlődnek⁵⁰. Általában két fő séma kombinációja alapján osztályozzák őket. Egyrészt a genomjuk típusa és a replikációjuk módja szerint, ez a Baltimore-rendszer (I-VII csoportok)⁵¹⁻⁵³, mely alapján megkülönböztetünk például DNS (egyszálú és duplaszálú), RNS [duplaszálú, illetve (+), vagy (-) egyszálú], és reverz transzkripciót alkalmazó vírusokat. Ehhez kapcsolódnak a nevezéktan szabályok és irányelvek, amiket az International Committee on Taxonomy of Viruses (ICTV) dolgozott ki, s változtat nagy gyakorisággal; legutóbb 2022-ben⁵⁴. Az ICTV a vírusokat hat birodalomba sorolja. A birodalmakban található vírusoknak nem feltétlenül vannak közös őseik és a birodalmaknak sincsenek közös őseik. A vírusokat konkrét, konzervatív tulajdonságok alapján csoportosítják, melyek lehet, hogy egy, vagy több alkalommal alakultak ki az evolúció során. A vírusok ilyen irányú rendszerezését a 21. század módszerei, mint a metagenomika és a krio-elektronmikroszkópia tették lehetővé, ezek vezettek a *Riboviria* 2018-as, majd 2019-ben további három és végül 2020-ban a maradék két birodalom megalapításához.

2018-ban gyakorlatilag az összes RNS vírust a *Riboviria* birodalomba sorolták⁵⁵, először csak az RNS-függő RNS polimerázt [RNA-dependent RNA polymerase (RdRp)] kódoló vírusokat, majd egy évvel később kibővítették a retrovírusokkal, melyek RNS-függő DNS polimerázt [RNA-dependent DNA polymerase (RdDp)] kódolnak. Előbbiek az *Orthornavirae*, utóbbiak a *Pararnavirae* királyságot alkotják. A *Riboviriák* közé kevés prokarióta vírus tartozik, azonban a legtöbb eukarióta vírust magában foglalja, ideértve a legtöbb emberi, állati és növényi vírust egyaránt (a metagenomikai tanulmányok megváltoztathatják még ezt a perspektívát).

2019-ben alapították a legtöbb kettős szálú (ds)DNS-vírust magába foglaló birodalmakat. A dsDNS vírusok többségét a *Duplodnaviria* birodalomba sorolják. Ide tartoznak két külön törzset alkotva az eukarióta herpeszvírusok (*Herpesvirales* nemzetség) és a bakteriofágok (*Caudoviricetes* osztály) képviselői egyaránt⁵⁶. A dsDNS vírusok egy kisebb csoportja, többek között az óriás eukarióta DNS-vírusok⁵⁶ és a poxvírusok is a *Varidnaviria* birodalomba kerültek besorolásra (ICTV). A *Peploviricota* (herpeszvírusok) és *Uroviricota* (bakteriofágok) törzsei mellett 2023-ban egy feltételezett új, a *Duplodnaviria* birodalomba sorolt törzset is leírtak: ez a *Mirusviricota*⁵⁶, melynek tagjait az óceánok mélyén fedezték fel. E vírusok a virion morfogenezis alapján kerültek a *Duplodnaviria* birodalomba, azonban számos tulajdonságuk a *Varidnaviria* birodalom *Nucleocytoviricota* (pl. óriás- és poxvírusok) törzséhez hasonlít, s ezáltal összekapcsol két távoli rokon vírus birodalmat.

Szintén 2019-ben alapították az egyszálú (ss)DNS genomú vírusok többségét magába foglaló birodalmat, a *Monodnaviriata*. Prokarióta vírusok mellett számos, humán betegséget is okozó víruscsoport (pl. papilloma – és polioma vírusok), illetve a cirkuláris genomú cirkovírusok családja is ide került besorolásra⁵⁷.

A hipertermofil archaeákat fertőző vírusokat 2020-ban az *Adnaviria*⁵⁸, míg azokat a szatellit nukleinsavakat (virusoid), melyek replikációjukhoz egy másik vírust igényelnek (pl. Hepatitis delta), a *Ribozyviria* birodalomba sorolták⁵⁹. A nemrégiben felfedezett viroid-szerű elemek, az obeliszkék⁶⁰ besorolása még várat magára.

A fent említett, a vírusok klasszifikációjára és nevezéktanára vonatkozó változatosság miatt, mely a gyakori fajszerű névváltoztatást is magában foglalja, a disszertációban az általunk vizsgált vírusok esetén első említéskor megjelölöm a legújabb, 2022-es nevezéktan szerinti elnevezéseket, azonban a hagyományos neveket használom a továbbiakban.

3.1.3 A transzkripció és a replikáció szabályozása

A vírusok igen változatos fertőzési mechanizmusokat fejlesztettek az évmilliók során. Például, a bakteriofágok belövellik a DNS-üket a baktérium sejtekbe, míg az eukarióta vírusok endocitózissal vagy membránfúzióval jutnak be a sejtekbe. A buroknélküli vírusok csak az első mechanizmust vehetik igénybe, a membrán burokkal rendelkezők pedig fajtól függően mindkét módot alkalmazhatják. A sejtbe bejutott vírusok szállíthatóak a sejtmagba (a citoskeleton rendszert használva erre), de maradhatnak a citoplazmában is. Az előbbi a DNS-, az utóbbi pedig az RNS vírusokra jellemző, melynek oka, hogy a DNS vírusok a gazdasejt DNS szintetizáló és splice mechanizmusait használják, míg az RNS vírusok saját replikációs apparátust kódolnak. E szabály alóli kivétel például a nagy DNS genommal rendelkező poxvírusok, melyek maguk állítják elő a replikációjukhoz szükséges molekuláris apparátust, illetve az RNS-alapú influenza vírusok, melyek transzkriptjei éréséhez a gazdasejt sejtmagjában lévő splice apparátust használják⁶¹.

A génextpresszió szabályozása rendkívüli változatosságot mutat a különféle víruscsaládok között. Az egészen kicsi pozitív-szálú RNS vírusok esetében a teljes genom gyakran egyetlen mRNS molekulaként funkcionál, ami néhány gént kódol (pl. filovírusok). A negatív szálú RNS vírusoknál a genomról egy komplementer RNS képződik és ez szolgál mRNS-ként. A nagyobb RNS vírusok (pl. koronavírusok) esetében a genom egy része mRNS-ként működik, a másik részéről (szubgenomikus régió) viszont egy komplementer RNS szál képződik, ami templátként szolgál az mRNS képződéshez. A DNS vírusoknál az egyes gének kinetikai osztályokba sorolhatók, attól függően, hogy a vírus életciklusának melyik stádiumában aktívak. A herpeszvírusoknál (HV-ok) a legkorábban leíródozó (*immediate-early*; IE) gének rendszerint transzkripció faktorokat (TF) kódolnak, a korai (*early*; E) gének általában a DNS replikációhoz szükséges fehérjéket határozzák meg, míg a kései (*late*; L) gének a vírus szerkezeti elemeinek genetikai információját hordozzák⁶². A többi vírus esetén is hasonló a logika, a kinetikai osztályok nevezéktana tér csak el, illetve az E és az L osztályt is szokás ketté bontani mind a HV-oknál, mind más vírusok esetén.

3.1.3.1 A transzkripció szabályozása

A vírusos életciklus központi eleme a transzkripció szabályozása. Ebben a szekcióban kizárólag a herpeszvírusok mechanizmusait ismertetem. A ***Simplexvirus humanalpha1*** [Herpes simplex vírus 1-es típusa (HSV-1)] az alfaherpeszvírusok (α HV) modellvírusa. A HSV-1 transzkripciójának szabályozásában négy IE fehérje vesz részt. Az esszenciális ICP4 vírusfehérjét az *rs1* gén kódolja. E gén a virális transzkripció fő szabályozója, fehérje terméke a gazdasejt transzkripció faktorait (pl. TFIID) toborozza a HSV-1 promótereire⁶³. Az *us1* gén által kódolt ICP22 fő funkciója a transzkripció elongációjának szabályozása⁶⁴. Az HSV-1 *us1* génje az egyedi rövid (US) genomi régióban található egyetlen kópiában, míg annak promótere duplán van jelen az inverz ripit (IR) régióban (a másik IR kópia az *ul12* gén kifejeződését szabályozza). A *Varicellovirus* nemzetségben az *us1* gén az IR régióba transzlokálódott, így ezekben a szervezetekben e gén két példányban van jelen. Az ICP0, amelyet az *rl2* gén kódol, szigorú értelemben véve nem TF, de a pre-kromatin kölcsönhatások szabályozásával elősegíti a transzkripciót⁶⁵. Az ICP27, amelyet az *ul54* kódol, az RNS-polimeráz (RNP) enzim virális promóterekhez való kötődését segíti elő⁶⁶. A litikus ciklus szabályozásának mechanizmusa más α HV-ok esetében is hasonló a fent ismertetettekhez. Egy jelentős különbség azonban az *icp0*, *us1* és *ul54* ortológ gének expressziós kinetikájának megváltozása: e gének az E kinetikai fázisban fejeződnek ki a ***Varicellovirus suidalpha1*** [Pseudorabies vírus (PRV)] és az ***Varicellovirus equidalpha1*** [Equid alphaherpeszvírus 1 (EHV-1)]

esetében. Ezenkívül, az *icp0* gén strukturálisan és funkcionálisan is egyszerűsödött a HSV-1-hez képest ezekben a vírusokban. A bétaherpeszvírusok (β HV) és gammaherpeszvírusok (γ HV) alapvetően az α HV-okhoz hasonló mechanizmusokat alkalmaznak a transzkripció szabályozására. A **Cytomegalovirus humanbeta5** [humán cytomegalovírus (HCMV; β HV)] esetében két fő IE gén (*ie1* és *ie2*) szabályozza a transzkripciót⁶⁷. A **Lymphocryptovirus humangamma4** [Epstein-Barr vírus (EBV; γ HV)] IE fehérjéi (BZLF1 és BRLF1) transzaktivátorok, amelyek az E vírusgének transzkripcióját szabályozzák⁶⁸.

A herpeszvírusok alternatív fertőzési módját a látencia jelenti. A látens állapotba vonulás során a vírus genom túlnyomó része transzkripcionálisan represszált marad, csak néhány specifikus vírus RNS fejeződik ki. Az HSV-1 'látencia-asszociált transzkript'-je (LAT) az egyetlen vírus gén termék, amely magas szinten fejeződik ki látens állapotban⁶⁹. Ez a nem kódoló RNS (ncRNS) elnyomja a litikus gének kifejeződését az *icp4* aktivitásának blokkolásával és a heterokromatin kialakulásának elősegítésével a HSV-1 genomján⁷⁰. Bár a LAT számos nyílt olvasási keretet (open reading frame; ORF) tartalmaz, valószínűleg nem kódol polipeptideket^{71,72}. Más hosszú ncRNS-ek (lncRNS-ek) is kifejeződnek a látens állapotban, például a 'hosszú-látencia transzkript' (LLT; mely az *icp0* és az *icp4* géneket fedi át antiparallel módon), valamint az egyedi hosszú (UL) és US genomi régiók közti, úgynevezett L/S junkció környéki transzkriptek (L/STs), melyek az *icp34.5* és az *icp4* géneket fedik át fordított orientációban⁷³.

3.1.3.2 A DNS replikáció szabályozása

Míg a prokarióta genomokban egyetlen pontból (replikációs origó; Ori) indul a DNS szintézis, addig az eukarióta genomokban tipikusan több tízezer Ori található. A bakteriális Ori-t konszenzus szekvenciák határozzák meg, az eukarióták Ori-jei viszont nem szekvenciálisan, hanem a kromatin-szerkezetük alapján meghatározottak⁷⁴. A vírusoknak egy vagy néhány Ori-juk van, amelyeket részben az adott DNS-szegmens szekvenciája, részben pedig adott genomi régió strukturális tulajdonságai határoznak meg⁷⁵. Az eukarióta genomok replikációja az *origin recognition complex* (ORC) Ori-hoz történő kötésével kezdődik⁷⁶. Az ORC funkciója, hogy platformként szolgáljon a replizóma összeállításához, amely számos fehérjét tartalmaz, mint például a DNS helikáz, a DNS polimeráz (DNP), a topoizomeráz, primáz, a DNS giráz, az egy-szálú DNS-kötő fehérje (ssDBP), az RNáz H, DNS ligáz és a telomeráz enzimek.

A herpeszvírusok több replikációs fehérjét is kódolnak a DNS-ük szintéziséhez. Például a HSV-1 kódol egy Ori-kötő fehérjét (OBP; *ul9*), egy ssDBP-t (*ul29*), két DNP-t (*ul30* és *ul42*), és három helikáz/primáz molekulát (*ul5*, *ul8* és *ul52*)⁷⁷. Több vírus által kódolt ko-faktor is szerepet játszik a nukleotid anyagcserében, mint például a ribonukleotid-reduktáz (*ul39/40*), a timidin kináz (*ul23*), az uracil-DNS glikoziláz (*ul2*), a dezoxiuridin trifoszfátáz (*ul50*) és az alkalikus nukleáz (*ul12*), lehetővé téve a herpeszvírus replikációját nem-osztódó sejtekben is⁷⁸.

A HSV-1-nek három Ori-ja van, kettő az IR-ekben, melyek az US régiót veszik körül (OriS), és egy az UL régióban (OriL). A Simplexvírusokban az OriL a DNS replikációban szerepet játszó *ul29* és *ul30* között helyezkedik el, míg az OriS két példányát az *icp4* és az *us1* IE gének veszik körül az US régió belső ismétlődésében (IRS), és az *icp4* és *us12* gének az US régió terminális ismétlésében (TRS). Néhány Varicellovírus, mint például a **Varicellovírus humanalpha3** [Varicella Zoster vírus (VZV)] és a **Varicellovírus bovinealpha1** [Bovine alphaherpeszvírus 1 (BoHV-1)] nem rendelkeznek OriL-el. Ezzel szemben más Varicellovírusokban, mint a PRV és az EHV-1, az OriL az *ul21* és *ul22* gének

közötti intergénikus régióba került át az evolúció során. A HCMV-nek egyetlen replikációs origója (OriLyt) van, míg az emberi gammaherpeszvírusok, az EBV és a **Rhadinovirus humangamma8** [Kaposi-szarkómával társított herpeszvírus (KSHV)] két litikus (OriLyt-L és OriLyt-R) és egy látens replikációs origóval rendelkeznek (az EBV-ben OriP-ként és a KSHV-ben terminális ripítnek nevezve)⁷⁹⁻⁸¹.

3.1.3.3 Az RNS-ek szerepe a replikáció szabályozásában

Az elmúlt évtizedben az élővilág minden doménjében találtak az Ori-k környékéről leíró IncRNS-eket⁸². Egy friss kutatás szerint az emlős replikációs origó felismerő komplex 1-es alegysége (ORC1) körülbelül 72%-a aktív promóterekkel asszociált, melyek több mint a fele ncRNS és aktív promóterek által szabályozott⁸³.

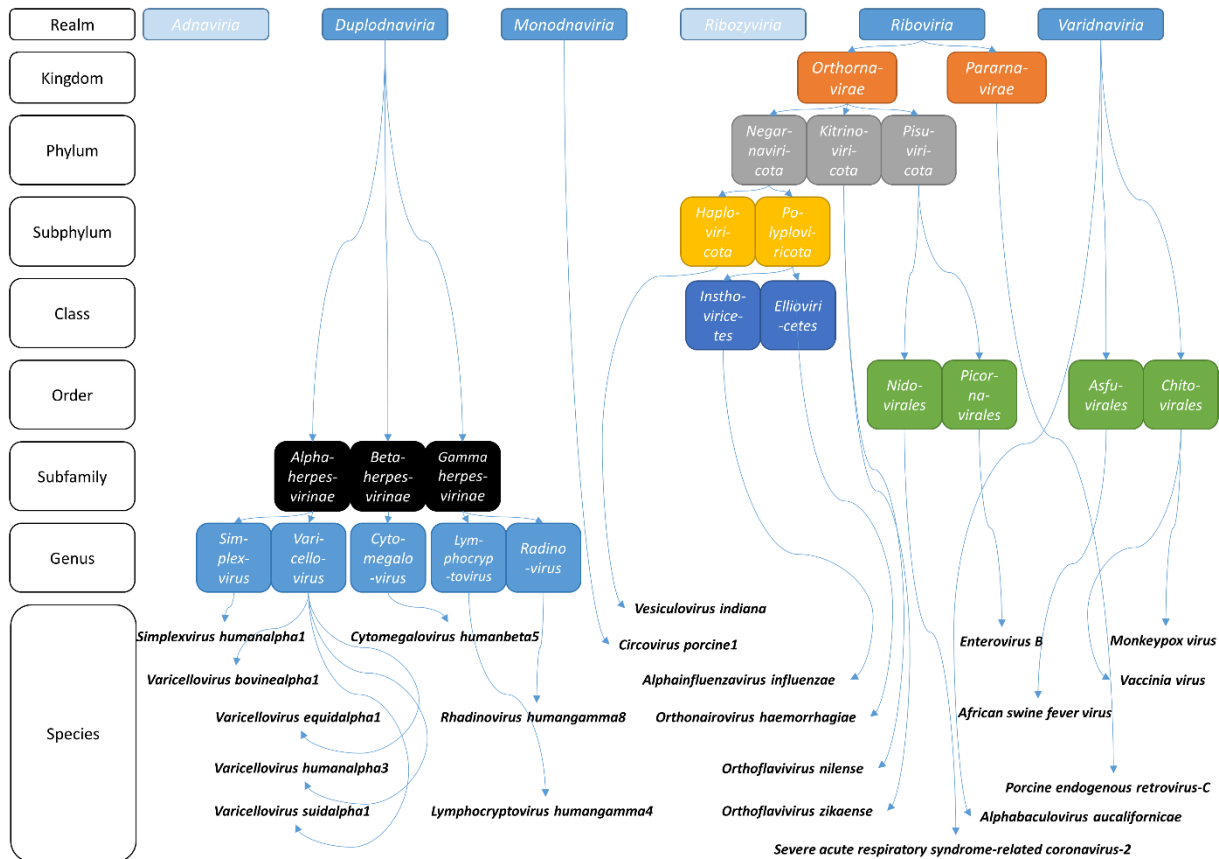
A replikációs RNS-ek számos módon szabályozzák a DNS replikációt. Többek között az RNS primer szintézisének szabályozásával azáltal, hogy a DNS szekvenciával hibridizálnak⁶⁹, vagy az mRNS-sel alakítanak ki hibrideket. Ez utóbbi folyamat az RNáz H általi degradációjukat inicializálja, s ezáltal gátolja a replikációs fehérjék transzlációját. Továbbá, ezek a transzkriptek az ORC Ori-hoz való verbuválását is segíthetik⁸⁴.

Míg ezek a transzkriptek a β - és γ HV-ok esetében részben már ismertek voltak, az α HV-oknál gyakorlatilag észrevétlenek maradtak egészen mostanáig. Például HCMV esetén az OriLyt-ből leíró⁸⁵ RNA4.9 ncRNS-t funkcionálisan is jellemezték⁸⁶. Ez az RNS a vírus DNS replikációját szabályozza DNS:RNS hibrid kialakításával, befolyásolja az *uI57* gén által kódolt ssDBP expressziójának mértékét. Feltételezik, hogy az RNA4.9-nek további szerepei is vannak mind *cisz*, mind *transz* módon, például elősegíti a fő IE promóter transzkripciójának represszióját látencia alatt⁸⁷. Az RNA4.9 és más, a HCMV Ori régiója környékéről leíró RNS-ek felfedezését^{88,89} követően az volt a feltételezés, hogy ennek a vírusnak egyedi replikációs szabályozási rendszere van, ami később nem bizonyult igaznak.

Szintén régebb óta ismert az EBV OriLyt régiójában a BHLF1 replikációs RNS által kialakított RNS:DNS hibrid⁹⁰. Ezenkívül egy bidirekcionális promóter⁹¹ és egy 'nagyon strukturált' RNS is azonosításra került ebben a genomi régióban⁸⁷. Ez utóbbi transzkriptnek a funkciója, hogy segítse a vírus EBNA1 és HMGA1a fehérjéit az ORC⁹² toborzásában.

A HSV-1 OriS-e közelében két, koterminálisban végződő IncRNS töredék szekvenciáit korábban már szintén leírták⁹³, funkciót azonban nem rendeltek hozzájuk.

A következő szekciókban bemutatom rendszertani kategóriákra bontva a vírusokat, melyekkel transzkriptomika szintjén foglalkoztunk, s amelyekből e dolgozat született. Elsősorban tüneteket, epidemiológiai, illetve állategészségügyi jelentőségüket foglalom össze, illetve a genomjuk, transzkripciójuk és replikációjuk jellegzetességeit tárgyalom. Az **1. ábrán** a legújabb rendszertan szerint ábrázoltam őket.



1. ábra Az ICTV 2022-es taxonómiai besorolása és nevezéktana alapján az általunk (mind laboratóriumi, mind bioinformatikai szinten) vizsgált 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.

3.1.4 Herpeszvírusok

A *Duplodnaviria* birodalom *Orthoherpesviridae* (herpeszvírusok) családjába közel 120 faj tartozik⁵⁴, melyek különféle gerinces élőlényeket képesek megfertőzni⁹⁴, így számos humán patogént is ismerünk köztük. A herpeszvírusok jellegzetes tulajdonsága a látens fertőzés képessége, vagyis képesek hosszú ideig, esetleg egy életen át a gazdaszervezetben rejtőzködni, majd bizonyos körülmények között újra aktívvá válni. A család három alcsaládra oszlik. Az *Alphaherpesvirinae* képviselői pl. a HSV-1, vagy például az Aujeszky-féle betegséget okozó sertés herpeszvírus, a PRV. A *Betaherpesvirinae* alcsaládba tartozik pl. a HCMV és a *Roseolovirus humanbeta6* [humán herpeszvírus 6 (HHV-6)], míg *Gammaherpesvirinae* ismert humánpatogén tagjai az EBV és a KSHV. A három alcsalád tagjai különböznek a biológiai tulajdonságaikban, mint például a látens fertőzés helye és a replikációs ciklus ideje⁹⁵. A herpeszvírusok burkoltak, lineáris dsDNS genomjuk 125–240 kilobázispár (kbp) között változik^{96,97}.

A herpeszvírusok egyes tagjai általában specifikusak az adott gazdaszervezetre, és a fertőzés többnyire élethosszig tart. A kezelés leginkább a tünetek enyhítésére és a vírus replikációjának gátlására összpontosít⁹⁸. Az immunrendszer és a herpeszvírusok közötti interakciók, valamint a látens fertőzések reaktiválódásának mechanizmusai aktív kutatási területek, mivel ezek az információk alapvető fontosságúak az új terápiás stratégiák és vakcinák fejlesztéséhez.

Az emberek általában több herpeszvírus fertőzésen is átesnek életük során. Az átlagos szám meglehetősen változó lehet, és függ a földrajzi helyzettől, az életkori csoporttól, az immunrendszer állapotától, a szociális és gazdasági körülményektől, valamint az egyéni életmódtól és viselkedési tényezőktől. Sok esetben ezek a fertőzések enyhe vagy észrevehetetlen tünetekkel járnak, de bizonyos esetekben súlyos betegségeket is okozhatnak. A herpeszvírusokkal való fertőzöttség általánossága miatt ezek a vírusok közegészségügyi és orvostudományi szempontból is a figyelem középpontjában vannak.

3.1.4.1 *Humánpatogén alfaherpeszvírusok*

HSV-1: a fertőzés élethosszig tart, tünetei enyhék, vagy egyáltalán nem is okoz problémát. A vírus leggyakrabban szájüregi herpeszt eredményez, immunhiányos betegeknél azonban akut encefalitiszt (agyvelőgyulladás) okozhat. Globális gyakorisága rendkívül magas, becslések szerint a világ népességének akár 60-90%-a is fertőzött lehet már gyermek-, vagy serdülőkorban. Az Egészségügyi Világszervezet (WHO) globális becslései szerint világszerte több mint 3,7 milliárd, ötven év alatti ember érintett⁹⁹.

VZV: a bárányhimlő kórokozója, amely az elsődleges fertőzés során alakul ki. A bárányhimlő szövődményei relatíve gyakoriak, sokszor súlyosak is (pl. tüdőgyulladás, agyvelőgyulladás), főként felnőtt korban. A későbbi életkorban övsömört (zoster) okozhat, amely a látens állapotból történő reaktiválódás során jön létre¹⁰⁰. Az övsömör gyakorisága az életkor előrehaladtával nő¹⁰¹. A bárányhimlőn gyermekkorban a legtöbb ember átesik, különösen azokban a régiókban, ahol nincs rutinszerű oltási program. Hazánkban 1998-óta elérhető az oltás, 2019-től kötelező.

Az alcsalád harmadik ismert humánpatogén képviselője a **Simplexvirus humanalpha2** (HSV-2).

3.1.4.2 *Humánpatogén betaherpeszvírusok*

HCMV: előfordulási gyakorisága földrajzi régióként változó, de különösen magas a fejlődő országokban, ahol a fertőzöttség gyakran már csecsemőkorban megtörténik. Világszerte a lakosság 60–90%-át fertőzi meg¹⁰². Felnőtteknél mononukleózis-szerű tüneteket okozhat¹⁰³, míg újszülötteknél életveszélyes állapotot¹⁰⁴. Különbféle humán sejtet fertőz, beleértve a fibroblasztokat, az epitél és az endotél sejteket, a simaizom-sejteket és a monocitákat egyaránt¹⁰⁵.

HHV-6: két fő variánst foglal magába, az HHV-6A-t és az HHV-6B-t. Ezek a vírusok (a *Roseolovirus humanbeta7*-hez hasonlóan) nagyon gyakoriak, általában a korai gyermekkorban megfertőzik a populációk jelentős részét¹⁰⁶. Kétéves korra a gyerekek 90%-a átesik a HHV-6 fertőzésen. A HHV-6 a kora gyermekkorban gyakori exanthema subitum más néven roseola infantum (azaz a hirtelen kiütés, háromnapos láz) fő kórokozója¹⁰⁷. Az HHV-6 szorosan kapcsolódik egyes krónikus állapotokhoz, mint a krónikus fáradtság szindróma¹⁰⁸, valamint szerepet játszhat bizonyos neurodegeneratív betegségek kialakulásában is^{109,110}. (Mi major depresszióval diagnosztizált, később öngyilkosságot elkövetett páciensek agymintái szekvenálását követően, az analízis során detektáltunk HHV-6-ot, szemben a kontroll mintákkal¹⁰.)

3.1.4.3 *Humánpatogén gammaherpeszvírusok*

EBV: az infekciós mononukleózis és egyes ráktípusok okozója. Elsődleges fertőzése korai gyermekkorban általában enyhe vagy tünetmentes. Elsősorban nyál útján terjed, s a világ népességének jelentős része, akár 90%-a is fertőzött lehet már a korai felnőttkorra^{111,112}. Az EBV szerepet játszik pl. a Burkitt-limfóma patogenezisében, valamint részt vesz a nasopharyngealis carcinoma és a gyomorrák egyes típusainak kialakulásában is^{113,114}. Látenciából többek közt stressz hatására (legyengült immunrendszer) reaktiválódhat. Az EBV újraaktiválódását a hosszú COVID szindróma kialakulása egyik okaként határozták meg¹¹⁵.

KSHV: ez a vírus kevésbé gyakori, és kapcsolatban áll bizonyos típusú rákokkal, különösen azokban a populációkban, ahol magas a prevalenciája¹¹⁶. A fertőzése általában tünetmentes az egészséges egyéneknél, de súlyos betegségeket okozhat az immunrendszerrel kompromittált betegekben, mint például az HIV-fertőzöttekben¹¹⁷. A KSHV a Kaposi-szarkóma etiológiai kórokozója¹¹⁸, melyet a bőrben, a nyálkahártyákban és a belső szervekben kialakuló rendellenes véredények képződése jellemez¹¹⁹. Ezen túlmenően a KSHV-t más ráktípusokkal is összefüggésbe hozták, mint a limfómák egyik típusa (primary effusion lymphoma) és a Castleman-betegség¹²⁰.

3.1.4.4 *Állati alfaherpeszvírusok*

A herpeszvírusok családja rendkívül változatos gazdaspektrummal rendelkezik, és a legkülönbözőbb állatcsoportokban is megtalálhatók, az osztrigáktól a teknősökön át, a madaraktól az emlősökig számos gazdaállatban¹²¹. Projektjeink során az alábbi állati patogénekkal foglalkoztunk.

PRV: elsősorban sertéseket fertőz, agy- és idegrendszeri tüneteket okoz, mint például viselkedészavarok, légzési nehézség, de akár végzetes kimenetelű is lehet. Más állatfajokban (pl. szarvasmarhák, macskák, kutyák) is okozhat fertőzést, de ezekben az esetekben gyakran és gyorsan halálos kimenetelű. Az emberre általában nem veszélyes¹²². A PRV világszerte előfordul a vaddisznó populációkban, amelyek fontos rezervoárjai a vírusnak. A nemzeti eradikációs programok eredményeképpen az Aujeszky-féle betegség (AD) gyakorlatilag eltűnt a házi sertésekből több európai országban, valamint Kanadában, Új-Zélandon és az Egyesült Államokban is¹²³, azonban ez nem feltétlenül jelenti a vírus teljes megsemmisítését. Szerbiában például előfordul, az elmúlt évtizedekben számos sertésfarmon jelentettek AD járványt¹²⁴⁻¹²⁶. A házi sertésekből történő sikeres víruselimináció ellenére különböző európai országokban évről évre jelentenek PRV fertőzéseket¹²⁷⁻¹²⁹ vaddisznókban, farkasokban, vörös rókákban, ibériai hiúzokban és kutyákban. Legutóbb 2022 decemberében, Csehországban okozta a vírus vadászkutyák sajnálatos halálát¹³⁰. Látható, hogy az eradikációs és oltási erőfeszítések ellenére potenciális vírusátvitel fordulhat elő a vadon élő állatokkal való érintkezés révén¹³¹.

BoHV-1: elsősorban szarvasmarhákat fertőz, és többek között okozhatja az úgynevezett fertőző bovine rhinotracheitist, amely egy légúti betegség. A vírus stressz hatására reaktiválódhat¹³².

EHV-1: főként lovakat fertőz, okozhat légúti betegségeket, idegrendszert érintő tüneteket, illetve vetélést a vemhes kancáknál¹³³. Az Európai Unióban endémiásnak tekintendő, 2021-ben elsősorban Spanyolországban okozott komolyabb járványt¹³⁴.

A fent említett állati herpeszvírusok komoly károkat okozhatnak az állattartásban, de vakcinázással lehet védekezni ellenük.

A ***Varicellovirus cercopithecinealpha9*** [Simian varicella virus (SVV)] számos óvilági majomfaj nagy mortalitású fertőzését okozza. (E vírussal csak bioinformatikai elemzést végeztünk.)

3.1.5 **Varidnaviriák**

A birodalom jelentős képviselői a nukleo-citoplazmatikus nagy DNS-vírusok (NCLDV), melyek a *Nucleocyotviricota* törzsbe tartoznak. A törzs egyik rendjébe (*Citovirales*) a poxvírusok tartoznak közel száz fajjal, míg a másik, az *Asfuvirales* törzset egyetlen faj, az afrikai sertéspestis betegséget okozó ***African swine fever virus*** (ASFV) képviseli.

3.1.5.1 **A Nucleocyotviricota: nukleo-citoplazmatikus nagy DNS-vírusok (NCLDV)**

Az NCLDV-k egy diverz csoportot képeznek, a jelenleg ismert legnagyobb virionok, illetve a legnagyobb genomú vírusok (óriásvírusok) mind ide tartoznak¹³⁵.

Poxvírusok (*Poxviridae*): a családba számos olyan vírus tartozik, amelyek különféle állatcsoportokat fertőznek, gerinctelenektől az emberig. A család képviselői általában téglatest alakú¹³⁶, burkolt vírusok, bonyolult belső szerkezettel rendelkeznek, beleértve a viszonylag nagy (~ 200 kbp)¹³⁷, lineáris dsDNS genomot és a hozzá kapcsolódó enzimeket¹³⁸. A legtöbb DNS vírussal ellentétben a poxvírusok a fertőzött sejt citoplazmájában replikálódnak és az RNS-ek képződése, a transzkripció is ott történik, úgynevezett „vírus gyárakban”, függetlenül a gazdasejtől, s nem a sejtmagban^{139,140}.

A család embereket is fertőző, *Orthopoxvirus* nemzetségébe tartozik a ***Vaccinia vírus*** (VACV), a majomhimlő vírus (***Monkeypox vírus***; MPXV), a tehénhimlő és a rendkívül veszélyes fekete himlőt okozó *Variola virus*^{141,142}. A (fekete) himlő fertőzések a történelem során milliók halálát okozták¹⁴³. A VACV amellet, hogy humánpatogén, orvostörténetileg is jelentős vírus; sikeresen használták (Edward Jenner) oltóanyagként a fekete himlő elleni immunizálásra¹⁴⁴⁻¹⁴⁶. A betegséget a globális oltási programnak köszönhetően sikeresen eradikálták az emberi populációból, melyet a WHO 1980-ban hirdetett ki^{143,147}. A vírus lineáris DNS-e körülbelül 190 kb hosszúságú, inverz terminális ismétlődő szekvenciákkal határolt és körülbelül 220 fehérjekódoló génje van¹⁴⁸. A VACV legvirulensebb törzse a Western Reserve (WR)¹⁴⁹.

A majomhimlő egy zoonotikus kórokozó, amely Nyugat- és Közép-Afrikában endemikus, és az elmúlt évtizedekben, néhány ritka esettől eltekintve, az emberi MPXV fertőzések csak az említett térségben lokalizálódtak. A 2022-es évben azonban egyre több esetről számoltak be olyan országokból, ahol a betegség nem endemikus^{150,151}. A 2022-es majomhimlő járvány során a vírusgenom monitorozása felfedte, hogy a keringő MPXV törzs kapcsolódik a kevésbé patogén Nyugat-Afrikai MPXV kládhoz, de egy magas mutációs aránnyal rendelkező, erősen divergens új kládot alkot¹⁵²⁻¹⁵⁴. Ennek következtében a WHO vészhelyzeti bizottsága nemzetközi aggodalomra okot adó közegészségügyi vészhelyzetnek (Public Health Emergency of International Concern) minősítette a vírus járványpotenciálját az endemikus régiókon kívül is.

A poxvírus gének az expressziós kinetikájuk alapján E, köztes (I) és L génekre csoportosíthatók¹⁵⁵⁻¹⁵⁸. A vírusgenom jellegzetessége, hogy az E gének a genom két végéhez közel csoportosulnak¹⁵⁹, míg az I és L gének a virális DNS központi részén gyűlnek össze. A szomszédos gének főként azonos irányban helyezkednek el; ezért a konvergens és divergens gén párok száma viszonylag alacsony például a herpeszvírusokhoz képest¹⁶⁰.

ASFV: az afrikai sertéspestis betegséget (ASF) okozza, mely gyakran halálos kimenetelű a sertések és vaddisznók körében. Jelenleg nincs engedélyezett vakcina az ASF ellen, ami jelentősen megnehezíti a betegség elleni küzdelmet¹⁶¹. Számos európai országban jelen van a vírus, Németországtól Görögorszáig, Észtországtól Olaszorszáig, így hazánkban is¹⁶². Viszonylag gyakran előfordulnak helyi járványkitörések Európában, legutóbb 2023 nyarán Horvátországban¹⁶³.

Az ASFV terjedhet közvetlenül fertőzött állatok és azok váladékai, például a vér, a nyál érintkezése révén, valamint közvetetten fertőzött tárgyak, például etető- és itatóeszközök, járművek, ruházat, valamint a vírus ellenálló képessége miatt a fertőzött hús- és hústermékek révén is¹⁶⁴. Az ASFV okozta gazdasági kár jelentős. A betegség elleni védekezés fő eszközei a megelőző intézkedések, a szigorú karantén rendelkezések, valamint a fertőzés és terjedés elleni gyors és hatékony fellépés.

Az ASFV egy nagyméretű, dsDNS-vírus, genetikai állománya összetett, több mint 150 gént kódol. Az IE és E gének a DNS replikációt megelőzően, míg az I és L gének azt követően expresszálódnak¹⁶⁵. Az ASFV replikációs ciklusa részben a gazdaszervezet citoplazmájában zajlik¹⁶⁶.

3.1.5.2 A *bakulovírusok*

A *Varidnaviriák* ízeltlábúakat fertőző, nagy, burkolt, pálcika alakú virionnal, s dsDNS genommal rendelkező képviselői önálló osztályt (*Naldaviricetes*) alkotnak. A vírus replikáció a sejtmagban zajlik. A főként rovarokat fertőző vírusok a Bakulovírusok (*Baculoviridae*) családjába tartoznak. A család tagjai két fő nemzetséget alkotnak: az *Alphabaculovirusok* (lepkék és hernyók fertőzése) és a *Betabaculovirusok* (molyok fertőzése). Egyik jellegzetes tulajdonságuk, hogy polihéder formájú struktúrákat [úgynevezett occlusion bodies (OB-k)] képeznek, amelyek védik a vírusokat a környezeti tényezőktől és segítik a vírus terjedését az egyik gazdaszervezettől a másikig. Az OB-ken belül található virionok számos sejtbe képesek bejutni, de a vírus általában csak specifikus rovarsejtekben képes szaporodni¹⁶⁷.

A család legismertebb tagja az *Alphabaculovirusok* közé tartozó ***Alphabaculovirus aupalifornicae*** [korábbi, ismertebb nevén *Autographa californica* multiple nucleopolyhedrovirus (AcMNPV)]. Az AcMNPV-t különösen a biotechnológiai kutatásokban és alkalmazásokban vizsgálják, például a génexpressziós rendszerekben^{168,169} rekombináns fehérjék előállítására használják, valamint biológiai alapú növényvédőszerként (bioinsekticid) a mezőgazdaságban, a káros rovarpopulációk szabályozására. A bakulovírusokkal kapcsolatos kutatások a rovarpatológia, a molekuláris biológia, a biotechnológia és a mezőgazdasági tudományok területén egyaránt fontosak¹⁶⁷.

A bakulovírus gének expresszió alapján három csoportra oszthatók: E, L és nagyon kései [very late (VL)]; az E-ket gyakran kettéosztják IE és késleltetett korai [delayed early (DE)] génekre¹⁶⁷. Az IE és E gének promotereit gyakran a kanonikus TATA motívum veszi körül, melyet a gazda RNS polimeráz II enzime ismer fel, majd a transzkripció a korai iniciátor CAGT szekvenciától indul¹⁷⁰. Az L és VL transzkriptek a kései iniciátor szekvenciához [late initiator sequence (LIS)] kötődnek, mely egy TAAG motívumot vesz körül, s ezt a virális RNS polimeráz ismeri fel és indítja a transzkripciót a motívum második nukleotidjától^{171,172}. Az E transzkriptek konszenzus AAUAAA vagy AUUAAA poliadenilációs (PA) szignált (PAS) tartalmazznak, PA-farkukat a gazdasejt enzimej alakítják ki. Az L és VL transzkripteknek azonban nincs szükségük konszenzus PAS-ra. Kimutatták, hogy a virális RNS polimeráz templát függetlenül képes adozinokat az RNS végéhez adni, amint elérte a T-gazdag szekvenciát¹⁷³.

3.1.6 A cirkovírusok

A *Monodnavirae* birodalom egyik családját (*Circoviridae*) a cirkuláris ssDNS genommal rendelkező, madarakat és emlősöket fertőző vírusok alkotják. A cirkovírusok a legkisebb állati vírusok, genomjuk mindössze ~ 1000 - 2000 bp¹⁷⁴. Lipidburokkal nem rendelkeznek, kapszidjuk ikozaéder alakú. A vírusgenom a sejtmagban dsDNS-sé alakul, majd erről íródnak le az mRNS-ek.

A ma ismert cirkovírusok egy kivételével mind patogének. Ez a kivétel a ***Circovirus porcine1*** [sertés 1-es típusú cirkovírus, korábbi tudományos nevén *Porcine circovirus type 1* (PCV-1)]. A viriont Tischer és kollégái fedezték fel¹⁷⁵ a PK-15 (sertésvese epitél) sejtvonalban [porcine kidney epithelial cell line]. A PCV-1 genom az egyik legrövidebb a DNS vírusok közt, mindösszesen 1759 nukleotid alkotja¹⁷⁶⁻¹⁷⁸. Az Orit két gén, a *rep* (ORF1) és a *cap* (ORF2) veszi körül, míg az ORF3 antiszensz orientációban áll az ORF1-gyel¹⁷⁹⁻¹⁸¹. A génekről alternatív splicinggal különféle mRNS-ek képződnek. A *rep* két, a replikáció iniciációjáért felelős fehérjét (Rep és Rep') kódol, valamint 9 ncRNS-t, a *cap* a kapszid (Cap) fehérjét, míg az ORF3-ról nem azonosítottak korábban transzkriptet.

3.1.7 Ribovírusok

A Riboviria (ribovírusok, vagy riboviriák) birodalma az RNS-alapú vírusok összességét foglalja magába, amelyek genetikai információként ribonukleinsavat használnak. Az RNS, mint genetikai anyag, lehetővé teszi ezeknek a vírusoknak, hogy gyorsan mutálódjanak és alkalmazkodjanak a változó környezeti feltételekhez, így rendkívül változatos és dinamikus csoportot alkotnak^{51,52}. A birodalom vírusai közé tartoznak a közegészségügy szempontjából jelentős patogének, mint például az influenzavírusok, a hepatitis C vírus, a HIV, az ebola és a veszettség vírus, valamint a különböző koronavírusok, beleértve a SARS-CoV-2-t is⁵⁴. Továbbá, ide tartozik a dohánymozaik vírus is, az első vírus, amelyet felfedeztek. A genomba integrálódó retrovírusok (pl. HIV) szintén a birodalom képviselői. Az emberi genom mintegy 7–8%-a eredeztethető különféle retrovírus fertőzésekből^{182,183}, mint pl. a legaktívabb humán endogén retrovírus (HERV), a HERV-K¹⁸⁴.

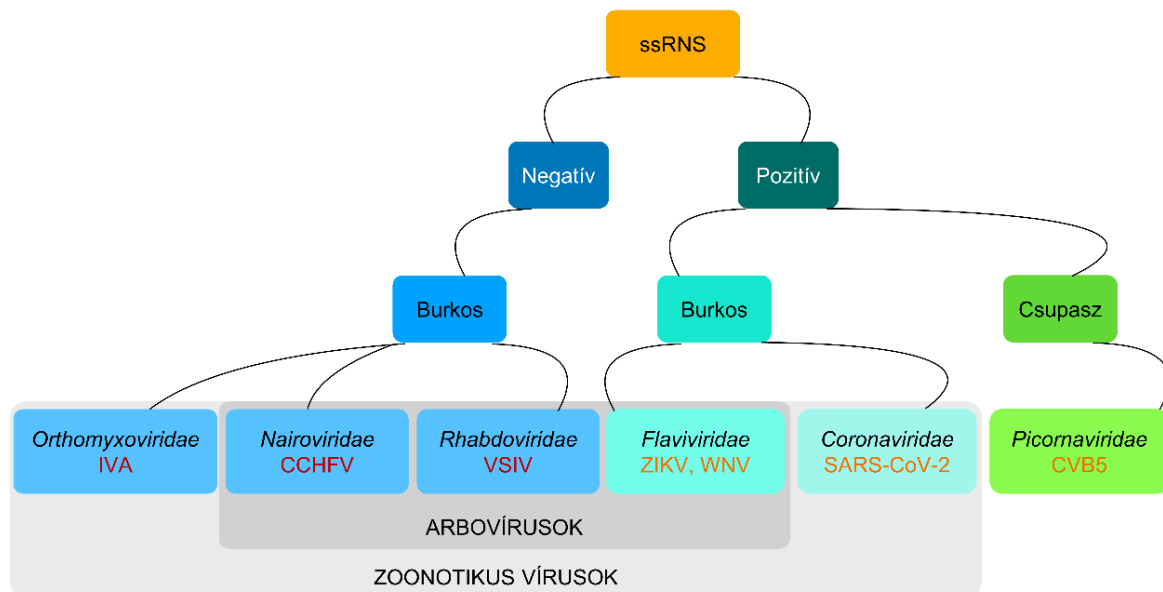
A ribovírusokat többféle szempont alapján csoportosíthatjuk, többek között a genetikai anyaguk szerkezete (pl. ss-, vagy dsRNS), a vírus részecskék felépítése (pl. burkolt vagy burkolatlan), és a vírusok által alkalmazott replikációs stratégia alapján^{51,52}. Ezek a vírusok különféle gazdaszervezetekben képesek megbetegedéseket előidézni. A legtöbbjük eukariótákat fertőz, beleértve az embereket, állatokat és a növényeket is. Van egy kisszámú RNS vírus, amely baktériumok kórokozója, azonban még nem írtak le RNS vírust archaeákban¹⁸⁵. Metagenomikai megközelítéssel, a Yellowstone archea-dominálta melegvízű forrásaiban detektáltak putatív archea RNS vírusokat, azonban a gazdaszervezetet még nem azonosították^{186,187}.

A ribovírusok kutatása kiemelten fontos a virológiában és az orvostudományban, mivel gyors adaptációs képességük és változatos biológiai tulajdonságaik révén komoly kihívást jelentenek a betegségmegelőzés és -kezelés terén. A megfelelő vakcinák és terápiák kifejlesztéséhez alapos ismeretekre van szükség a ribovírusok szerkezetéről, életciklusáról, patogenitásáról és interakciójáról a gazdaszervezetekkel^{98,188}.

3.1.7.1 *Orthornavirae* -> *Negarnoviricoták*

Az *Orthornavirae* királyságon belül számos család képviselői úgynevezett zoonotikus megbetegedést okozó vírusok, köztük találjuk a rendkívül súlyos betegségeket okozó arbovírusokat (ízeltlábúak által terjesztett vírusok) is (2. ábra).

A *Negarnoviricoták* (-) ssRNS genomjáról az RdRp szintetizálja az mRNS-t, mely tulajdonképpen a komplementer, pozitív szál. E pozitív szálról képződnek az újonnan szintetizálódó vírusok (-)ssRNS genomjai, míg a pozitív szálról (mRNS-ek) a szerkezeti és egyéb fehérjéik¹⁸⁹. A törzs számos altörzsre és osztályokra tagozódik. Mi három, külön osztályba tartozó *Negarnoviricota* fajjal, az ***Alphainfluenzavirus influenzae***-vel [Influenza A vírus, (IVA)], a ***Vesiculovirus indiana***-val [Vascular Stomatitis Indianavirus, (VSIV)], és az ***Orthonairovirus haemorrhagiae***-vel [Crimean-Congo hemorrhagic fever virus, (CCHFV)] foglalkoztunk.



2. ábra Az ábrán azon ssRNS genomú víruscsaládokat tüntettem fel, melyek transzkriptom profiljával foglalkoztunk. Az ábra a vírusok genom irányultságára (negatív, pozitív), a vírusburok jelenlétére, vagy hiányára is utal. A rövidítések az alábbi vírusokra utalnak: IVA: Influenza A vírus; CCHFV: Krími-kongói vérzéses láz vírus; VSIV: Vesicular stomatitis indianavirus; ZIKV: Zika vírus; WNV: Nyugat-nílusi vírus; SARS-CoV-2: súlyos akut légzőszervi szindróma-koronavírus 2; CVB5: Coxsackie vírus B5.

IVA: az *Orthomyxoviridae* család képviselője, a szezonális influenzajárványokért felelős (az Influenza B vírussal karöltve, de pandémiát csak az IVA okozhat). Az IVA-kat a felszínükön található két fehérje, a hemagglutinin (H, vagy HA; 18 altípus) és a neuraminidáz (N, vagy NA; 11 altípus) alapján osztályozzák. A H és N kombinációi révén potenciálisan 198 altípus létezhet, de természetben (elsősorban madarakban) mostanáig csak 131 altípust azonosítottak¹⁹⁰. Az influenzavírusok genomja szegmentált [az IVA genom nyolc (-)ssRNS szegmensből áll], s az egyes vírusok szegmenseket cserélhetnek egymással, amikor egyszerre, ugyanazt a gazdát fertőzik meg. A szegmensek egy része csak egy-egy gént kódol, míg néhány szegmens kettőt¹⁹¹. Az IVA esetén az 1-es szegmens a polimeráz PB2 alegységét, a 3-as a polimeráz proteáz aktivitású (PA) alegységét, a 4-es a HA, az 5-ös a nukleoprotein (NP) fehérjét, míg a 6-os az NA-t kódolja. Az influenza vírusok mindegyike a polimeráz PB1 alegységét a 2-es szegmensben kódolja. Sok IVA törzs a PB1-F1 kiegészítő fehérjét (accessory protein), melynek apoptotikus aktivitása van, szintén a 2-es szegmensben, azonban a +1 leolvasási keretben kódolja¹⁹². A 7-es szegmens az M1 (mátrix) fehérjét, valamint az IVA-k esetén alternatív splicinggal az M2 ioncsatornát kódolja¹⁹³. A 8-as

szegmensről az interferon-antagonista NS1 fehérje képződik^{194,195}, valamint splicinggal a NEP/NS2, mely a virális ribonukleoprotein (RNP) gazdasejt nukleusából való transzportjárt felel^{196,197}.

VSIV: a *Rhabdoviridae* család tagja, zoonózisos betegséget okoz, a fertőzés rovarcsípések (pl. *Culicoides* fajok), vagy közvetlen érintkezés révén terjed az emlős gazdák között¹⁹⁸. Bár az emberben csak enyhe tünetek jellemzőek¹⁹⁹, többek között láz, izomfájdalom, fejfájás, hányás²⁰⁰, a természetes gazdáiban (lovak, szarvasmarhák és főként sertések²⁰¹) vezikuláris betegséget okoz. A vezikuláris betegség nagyon hasonló a száj- és körömfájáshoz, s jelentős veszteséget eredményezhet a háziállat-állományban. A VSIV-fertőzés korábban gyakori volt laboratóriumi dolgozók és állatgondozók körében is.

A vírusgenom kicsi (~ 11 kbp)¹⁹⁸, öt polipeptidet kódol: N, P, M, G és L. Utóbbi a nagy (large) RdRp, mely a vírus mRNS-einek cappeléséért és poliadenilációjáért is felel^{202,203}. Az N a nukleoprotein, mely körülveszi az RNS genomot²⁰⁴. A P (foszfoprotein) az L fehérje katalitikus kofaktora²⁰⁵, az M a mátrix fehérje számos funkcióval (pl. gazda RNS-ek gátlása, apoptózis)²⁰⁶, míg a G (glükoprotein) a vírus sejtbe való bejutásáért felel²⁰⁷.

CCHFV: a *Bunyavirales* rend, *Nairoviridae* családjának tagja, mely család rovarok által terjesztett vírusokat foglal magába²⁰⁸. A CCHFV a krími-kongói vérzéses láz okozója, mely egy súlyos, zoonotikus, bizonyos kullancsok (az elsődleges vektorok a *Hyalomma* genusz tagjai) által közvetített betegség. Halálozási arány 30% feletti²⁰⁸. A betegséget először Kongóban írták le az 1960-as években²⁰⁹. A vírus Afrikában, Közel-Keleten, Ázsiában, Dél- és Kelet-Európában gyakran okoz fertőzést²¹⁰, az utóbbi években (a klímaváltozás következtében) hazánkban és a régiókban (pl.: Ausztria, Csehország, Németország) is megjelent a betegséget terjesztő kullancs²¹¹⁻²¹³. A vírus hosszútávú, kontinenseket átívelő transzportját²¹⁴ CCHFV-fertőzött kullancsokat hordozó madarak biztosítják. Jelenleg nincs a betegség elleni vakcina és specifikus antivirális szerek sem (preklinikai stádiumban van számos ígéretesnek tűnő szer). A CCHFV patomechanizmusa nem ismert. Számos állatfajt képes megfertőzni, súlyos betegséget azonban csak az emberben okoz. A rend más tagjaihoz hasonlóan a genomja három szegmensből áll, melyek a kis (small, S), a közepes (medium, M) és a nagy (large, L). A genomot az NP borítja, melyet az S szegmens kódol az egyik reading frame-ben, míg egy szembeni irányban álló ORF-ben a kis, nem strukturális fehérjét (NSP). Az M szegmens komplex, egyrészt egy glükoprotein prekuzort kódol, melyet a gazda proteázai először a GP160/85 doménre bontják, majd ez tovább alakul mucin-szerű doméné és GP38-cá. Továbbá, az M kódolja a Gn, Gc glükoproteineket és a közepes NSP-t. Az L a CCHFV esetén a többi Bunyavirushoz képest szokatlanul nagy, RdRP-t és egy proteázt kódol²⁰⁸.

3.1.7.2 *Orthornavirae* -> *Kitrinoviricoták*

A *Kitrinoviricota* törzs neve a sárga és vírus szavakra utal, jelezvénny, hogy a sárgalázat okozó vírus is e csoportot képviseli. A törzs tagjai eukariótákat fertőző, (+)ssRNS genomú vírusok. Mi a *Flaviviridae* család két tagja, az ***Orthoflavivirus nilense*** [West Nile virus (WNV)] és az ***Orthoflavivirus zikaense*** [Zika virus (ZIKV)] RNS profilját határoztuk meg.

WNV: először 1937-ben azonosították Ugandában, a Nyugat-Nílus körzetben, amelyről a nevét is kapta²¹⁵. A WNV a legelterjedtebb flavivirus az egész világon, jelentős közegészségügyi problémát jelentve számos országban. Főként *Culex* nemzetségű szúnyogok terjesztik, a madarak a természetes rezervoárjai²¹⁶. Az emberek és egyéb emlősök véletlenszerű gazdaként szolgálnak, amelyekben a vírus nem terjed tovább. A WNV terjedése szorosan összefügg a szúnyogpopulációk

dinamikájával és az éghajlati változásokkal. A WNV fertőzés embereken leggyakrabban tünetmentes, de körülbelül 20%-ban lázas megbetegedést okozhat, influenzaszerű tünetekkel. Súlyosabb esetekben a vírus a központi idegrendszert is megtámadhatja, meningitist, encephalitist vagy akár halált is okozhat, főként időseknel és immunkompromittált személyeknél. A neuroinvaszív tünetek 200 emberből egyet érintenek átlagosan²¹⁷.

A WNV genom körülbelül 11 kbp, s ez egyben, mint a vírus egyetlen mRNS-e is funkcionál. Továbbá, mint templát, erről képződik a komplementer, negatív RNS szál is²¹⁸. A genom 3'-végén nincs poli(A) farok. Az egyetlen ORF egy poliproteint kódol, melyet a vírus saját, illetve gazda eredetű proteázai tíz proteinre hasítanak²¹⁹. A három virális struktúrfehérje a kapszid (C), az előmembrán/membrán (prM/M) és a burok (En), mind az ORF 5' végén kódolódnak, míg a hét NSP (NS1, NS2A, NS2B, NS3, NS4A, NS4B, and NS5) a 3' végén.

ZIKV: 2015–2016-ban a világ a ZIKV által okozott globális járvány terjedésével szembesült. Ezt az arbovírust először 1947-ben, Ugandában azonosították majmokban, de embereken csak 60 évvel később detektálták²²⁰. 2007-ben először Mikronéziában jelentettek járványt, mely a lakosok 73%-át érintette, majd egy újabb kitörés következett Francia Polinéziában 2013-ban, amely a lakosság 70%-át érintette^{221,222}. A ZIKV fertőzések súlyos következményeiről azonban nem számoltak be egészen addig, amíg Dél-Amerikába nem érkezett a vírus, amikor is a ZIKV fertőzéseket összefüggésbe hozták különféle születési rendellenességekkel, mint például a vírus anyáról gyermekre történő átadása miatti mikrocefáliával és egyéb neurológiai betegségekkel („congenital Zika szindróma”)²²³, mint például a Guillain–Barré szindróma²²⁴. A ZIKV terjedése elsősorban fertőzött szúnyogok csípése útján történik (fő vektorként az *Aedes aegypti*-t tartják számon), de más, nem vektor által terjesztett fertőzési útvonalakat is jelentettek, mint például véradás, transzplacentális fertőzés^{225,226}. Ezenkívül tanulmányok igazolták a fertőző ZIKV jelenlétét és perzisztálását pl. nyálban, vizeletben²²⁷.

A ZIKV genom körülbelül 11 kb, mely egyetlen poliproteint kódol²²², ami pedig három szerkezeti fehérjét (C, prM és E); valamint hét NS fehérjét kódol²²⁸. A szerkezeti fehérjék felelősek a vírus partikulumok kialakulásáért, míg az NS fehérjék fontos funkciókat töltenek be a genom replikációjában, a poliprotein feldolgozásában és a gazda válaszreakcióban²²⁹.

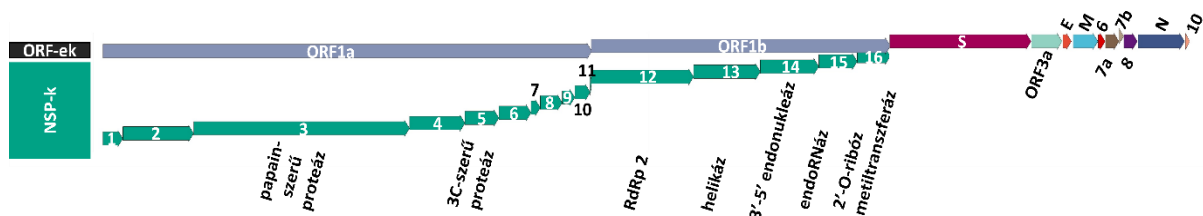
3.1.7.3 *Orthornavirae* -> *Pisuviricoták*

A *Pisuviricoták* közé (+)ssRNS és dsRNS genomú, eukariótákat fertőző vírusokat soroltak. A törzs *Nidovirales* rendjébe tartoznak a zoonózist okozó koronavírusok (*Coronaviridae*, CoV), a ma ismert legnagyobb genomú RNS vírusok^{230,231}. A genom méretük ~ 30 kbp²³². A koronavírusok legismertebbike a **Severe acute respiratory syndrome-related coronavirus-2** (SARS-CoV-2), mely a Covid-19 pandémiát okozta a 2020-as év elejétől^{233,234}. Ugyanakkor a legkisebb RNS vírusok, a *Picornaviridae* [**pico**(=kicsi) + **RNA**] család tagjai is a *Pisuviricota* törzsbe tartoznak.

SARS-CoV-2: a vírust először 2019 végén azonosították a kínai Vuhanban, és rövid idő alatt globális pandémiát (Covid-19) okozott, jelentős hatással az egészségügyi rendszerekre, a globális gazdaságra és a mindennapi életre. A SARS-CoV-2 főként cseppfertőzéssel terjed, tünetei nagyon változatosak, az enyhe légúti tünetektől kezdve a súlyos tüdőgyulladásig és akár halálos kimenetelig. A leggyakrabban láz, köhögés, fáradtság, íz- és szaglászvesztés jelentkezik. Súlyos esetekben légszomj, mellkasi fájdalom és zavartság is előfordulhat. Néhány emberben súlyosabb szövődmények, mint például akut légzési distressz szindróma, szív- és érrendszeri problémák vagy

trombózis is kialakulhat. A vírus jellegzetessége a tüskefehérjéi a vírusburkon, amelyek a gazdasejt ACE2 receptorához kötődnek, lehetővé téve a vírus bejutását a sejtekbe. A tüskefehérje egyik fő jellemzője, hogy a mutációkra való hajlam lehetővé teszi a vírusnak, hogy új variánsokat hozzon létre. Az újabb variánsok újabb tüneteket is produkálnak. 2024. április közepére közel 705 millió SARS-CoV-2 beteget jelentettek, melyek közül több mint 7 millió halállal végződött²³⁵.

Genomja körülbelül 30 kb hosszúságú, 38% G+C tartalmú, szekvenciája ~ 82%-ban azonos a 2000-es évek elején szintén Kínában megjelent SARS-CoV és a 2012-ben Szaud-Arábiában megjelent Middle East Respiratory Syndrome CoV (MERS-CoV) vírusok genomjával. Az esszenciális enzimek és struktúrfehérjék szekvenciái tekintetében ez a hasonlóság > 90%-os, mely közös patogenezis mechanizmusra utal, s terápiás célpontot jelenthet. A SARS-CoV-2 genom négy struktúrfehérjét kódol, a tüske (spike, S), az E, M és a nukleokapszid (N) fehérjéket²³¹. A genom 12 funkcionális ORF-et tartalmaz (3. ábra).



3. ábra A SARS-CoV-2 genom struktúrája²⁴. A két nagy ORF (1a és 1b) poliproteint kódol, melyekről számos NSP képződik proteolízissel. E két ORF a genom közel kétharmadát jelenti, a maradék régió pedig a struktúrfehérjéket, illetve további „kiegészítő” fehérjéket kódol.

A vírusgenom replikációja és a transzkripciója összefüggnek, hiszen ugyanúgy az RdRp enzim vesz részt a két folyamatban²³⁶. Először a (-)ssRNS szálak íródnak le, melyek templátként szolgálnak mind a genomi (gRNS), mind pedig a szubgenomi RNS-ek (sgRNS) szintéziséhez²³⁷. A gRNS-nek és az sgRNS-eknek közös 5' és 3' végződése van, hiszen az RdRp a genom ezen végéről szintetizálja a (+)ssRNS-eket. Az sgRNS-ek negatív szálának szintézise során történik meg a template switching (szálváltás, TS), amit a 'genom testben' (genome body; TRS-B) és a 'vezető', azaz leader szekvenciákban (TRS-L) lévő transzkripció-szabályozó szekvenciák (TRSs) közvetítenek, mely a leader és body szekvenciák fúzióját eredményezi^{238,239}.

Enterovirus B [*Human coxsackievirus B5* (CVB5)]: a Picornavírusok Enterovirus nemzetiségébe tartozó CVB5 által okozott tünetek közé többek között láz, aszeptikus agyhártyagyulladás, szívizomgyulladás, hasmenés, makulopapuláris, nem viszkető kiütés tartoznak. Elsősorban szájon át terjed, gyakran széklettel szennyezett víz, vagy élelmiszer fogyasztásával. A vírus főként gyermekkorban és a fiatal felnőtteknél fordul elő. A fertőzés összefüggésbe hozható az 1-es típusú cukorbetegség kialakulásával is, különösen gyermekkorban.

Genomja ~ 7,4 kbp, egyetlen, hosszú ORF-en kódol egy poliproteint, amit aztán proteázok bontanak le különböző funkcionális fehérjékre²⁴⁰.

3.1.7.4 *Pararnavirae* -> *Artverviricoták*

A *Pararnavirae* birodalom egyetlen törzse az *Artverviricota*, ennek egyetlen osztálya a *Retraviricetes*, mely név a **reverz transzkriptáz** és a **viricetes** (vírusosztályok rendszertani végződése) szavakból származik. A *Retraviricetes* magában foglal minden reverz transzkriptáz (RT) enzimet kódoló vírust, az összes ssRNS-RT vírust (beleértve a retrovírusokat) és a dsDNS-RT vírusokat is²⁴¹.

A retrovírusok szaporodásuk során integrálódnak a gazdagenomba, s ha ez csírvonalban történik meg, akkor a gazdagenom részévé válhatnak. Ilyenek a HERV-ek, a koalákban exogén és endogén formában is előforduló koala retrovírus (KoRV; az endogenizáció folyamata zajlik jelenleg is) és a sertések endogén retrovírusa (PERV) is. Az eukarióta fajokban előforduló endogén retrovírusok általában egy retrovírus, fajok közti transzmissziójának és integrációjának eredményei, az új gazda ivarsejt-vonalába²⁴². A PERV-ek eredete valószínűleg egy egérből származó retrovírus volt²⁴³.

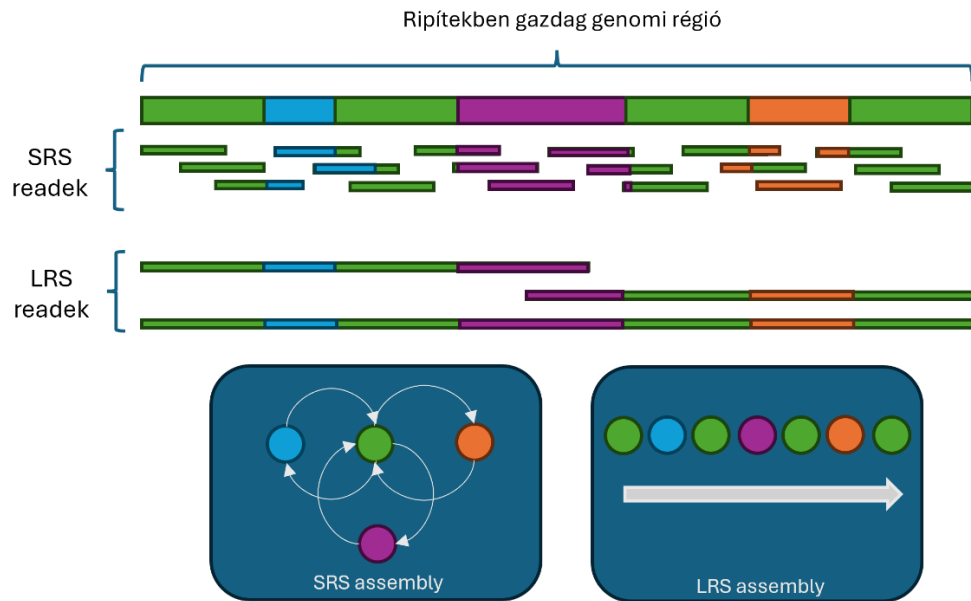
Porcine endogenous retrovirus-C (PERV-C): C-típusú gammaretrovírus²⁴⁴, mely kizárólag malacokat fertőz, míg az A és a B típusok akár embert is²⁴⁵. A PK-15 sejt vonal a PERV genom legalább két különböző klónjának 10-20 kópiáját is tartalmazza²⁴⁶. A genomja ~ 8,6 kbp. A *gag* gén a strukturális fehérjéket (M, C és N) kódolja, a *pol* a proteáz (PR), a reverz transzkriptáz és az integráz enzimeket, míg az *env* a burok fehérjéit. A Gag és Pol poliproteineket a fent említett fehérjékké a PR hasítja²⁴⁷.

Kutatócsoportunk, s az én elmúlt évtizedes munkám alapvetően a legmodernebb genomikai módszerek a molekuláris virológiai kutatásainkba való integrálásán, alkalmazásán alapult. A következő nagy fejezet a szekvenálási technikák alapjait hivatott bemutatni.

3.2 A SZEKVENÁLÁS FORRADALMA

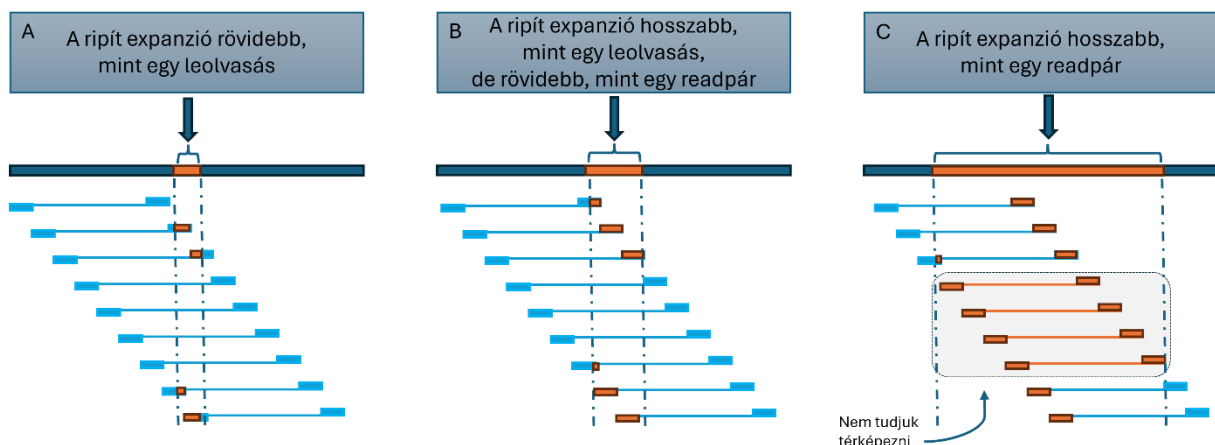
DNS és RNS szekvenálás mára a modern genomikai kutatások nélkülözhetetlen eszközeivé váltak. Az egyik legkorábbi módszer az első generációs Sanger-féle szekvenálás, mely rendkívül drága, időigényes és körülményes a végrehajtása²⁴⁸. Az utóbbi másfél-két évtizedben azonban a szekvenálás igen gyors technológiai fejlődésen ment keresztül, ami drasztikus idő- és költségcsökkenést is eredményezett, a genomikai kutatások óriási ütemű progresszióját lehetővé téve ezáltal. A szekvenálás revolúcióját két fő technológiai ugrás jellemezte: először az ún. második generációs, rövid-leolvasású szekvenálás (*short-read sequencing*; SRS), majd a harmadik generációs hosszú-leolvasású szekvenálás (*long-read sequencing*; LRS) megjelenése. Az SRS (elsősorban az Illumina technológia) forradalmasította a genomszekvenálást a 2000-es évek elejétől²⁴⁹. E módszer rövid [cc. 25-300 bázispár (bp) hosszúságú (két irányból összesen akár 600 bp)] szekvenciákat olvas le, amelyeket összerakva hozzuk létre a teljes genomot bioinformatikai módszerekkel. Az Illumina technológián alapuló, az elmúlt évtizedben leginkább alkalmazott platformok akár több százmillió read egyidejű leolvasásával igen nagy áteresztőképességet produkálnak²⁵⁰, ma a NovaSeq 6000 és az X több 10 milliárd bázispárnyi adatot képes generálni egyidejűleg²⁵¹. Ugyanakkor az SRS komoly korlátokkal is rendelkezik, mivel az e módszer által produkált szekvencia adatokból igen nehéz összerakni a genom komplex és ismétlődő régióit, illetve detektálni a hosszú strukturális variánsokat (deléciók, duplikációk, splice). E szekvenciák ugyanis gyakran hosszabbak, mint amit az SRS technológia leolvasni képes (**4. ábra**). Az LRS [a Pacific Biosciences (PacBio) és az Oxford Nanopore Technologies (ONT) által kidolgozott]

metodikák a genom- (és transzkriptom) szekvenálás újabb forradalmát hozták el, lehetővé téve hosszú szekvenciák leolvasását. Az LRS képes natív RNS szekvenálásra²⁵², illetve epigenetikai mintázatok és RNS szerkesztés közvetlen detektálására is²⁵³.



4. ábra Az ábrán szematikusan ábrázoltam egy ismétlődésekben gazdag genomi régiót, fent SRS, alatta LRS módszerrel megszekvenálva. 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.

Az LRS-nek számos előnye van az SRS-hez képest. Az LRS lehetővé teszi a nagy szerkezeti variációk és az ismétlődő DNS elemek elemzését, míg az SRS e tekintetben korlátozottan alkalmazható²⁵⁴. Az LRS lehetővé teszi a teljes genomok összeállítását, beleértve a korábban megközelíthetetlen régiókat is²⁵⁵. Emellett lehetővé teszi a ritka patogén strukturális variánsok és repeat-expanziók azonosítását és karakterizálását is, szemben az SRS-sel, mely esetén problémák merülhetnek fel (5. ábra)²⁵⁶.



5. ábra Az SRS módszerek nem alkalmasak az ismétlődésekben gazdag genomi régiók, pl. ripít expanziók pontos meghatározására. Ha a ripít expanzió rövidebb, mint egy leolvasás (A), vagy mint a paired-end, azaz a páros végről történő szekvenáláskor a két leolvasás együttese (B), akkor nincs komolyabb gond, a bioinformatikai programok képesek kezelni a problémát és meghatározni a ripít expanzió egészét. Ha azonban hosszabb a ripít régió, mind egy read pár, akkor nem megbízható az SRS módszer (C).

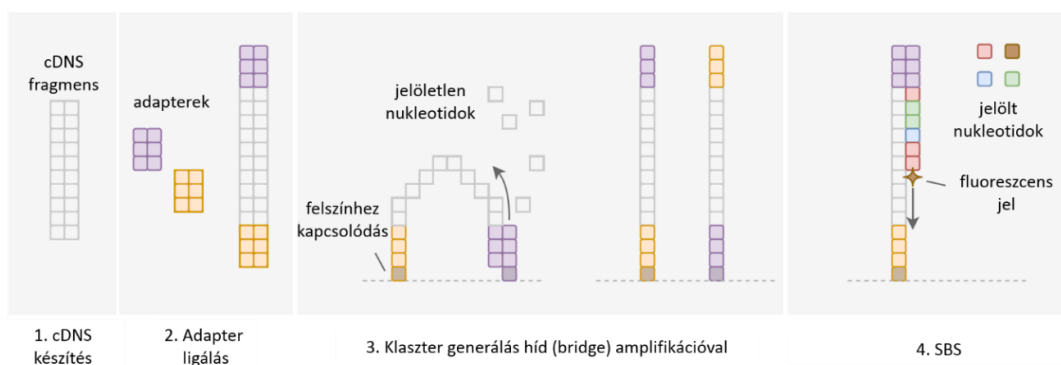
Azonban az LRS viszonylagosan pontatlan, ami javítható pl. SRS-sel történő kombinálással²⁵⁷. Az SRS nagyobb áteresztőképessége miatt előnyösebb lehet például a metagenomikában, mivel több genomot, s így több fajt azonosíthat az LRS-hez képest. Ugyanakkor az LRS révén teljes és megbízható genomokat kaphatunk diverz metagenomikai mintákból is, a módszer precízebb annotációt eredményezhet, alkalmazása révén valószínűbb képet kapunk a mikrobák diverzitásáról²⁵⁸⁻²⁶⁰.

Összességében az LRS lehetőséget kínál a genetikai diverzitás és a patogén mutációk mélyebb megértésére az emberi populációkban és a mikrobiális közösségekben egyaránt, de az optimális pontosság és lefedettség érdekében még további szekvenálási módszereket igényelhet, mely szükség a technikák fejlődését figyelembe véve hamarosan megszűnhet.

A LoopGenomics (mára az Element Biosciences-szel egyesült) kidolgozott egy SRS technikára épülő szintetikus LRS módszer (SLRS), melyet LoopSeq-nek nevezünk. Tudomásom szerint mi próbáltuk ki először a világon e technikát RNS szekvenálásra¹⁸. Úgy tűnik, hogy egyelőre a piac nem fogadta kedvezően e megközelítést. Az LRS technológia hibatája egyelőre nagyobb, mint az SRS-é, és általában drágább is, habár mindkét paraméter javulását tekintve igen gyors progresszió figyelhető meg napjainkban.

3.2.1 Az Illumina technológia elve

Az Illumina szekvenálási technika az úgynevezett szintézisen alapuló szekvenálás (sequencing by synthesis, SBS) elvén működik (6. ábra), melynek lényege, hogy az egyszálú, merev felületre rögzített DNS-mintához, szekvenciálisan hozzáadják az – általában - négy különböző, fluoreszcensen jelölt dNTP-t (dATP, dCTP, dGTP, dTTP)²⁶¹. (A NovaSeq készülék kétcsatornás, csak kétféle színt alkalmazunk). A DNS-polimeráz enzim beépíti ezeket a jelölt nukleotidokat az újonnan szintetizálódó DNS szálba, és minden beépülés után, lézerfényrel való gerjesztés következtében a fluoreszcens csoportokat egyedi emissziós színük alapján azonosítja a készülék. A terminátor csoport megakadályozza a következő nukleotid beépülését, így lehetővé téve az egyes beépülések pontos sorrendjének meghatározását²⁶¹. A fluoreszcens jel eltávolítása után a folyamat ismétlődik, így lépésről lépésre olvasható le a teljes szekvencia²⁶². Az Illumina platformok lehetővé teszik akár több millió/milliárd szekvencia egyidejű olvasását. Az eljárást magas áteresztőképessége és pontos szekvencia meghatározási képessége tette az egyik legelterjedtebb módszerre.

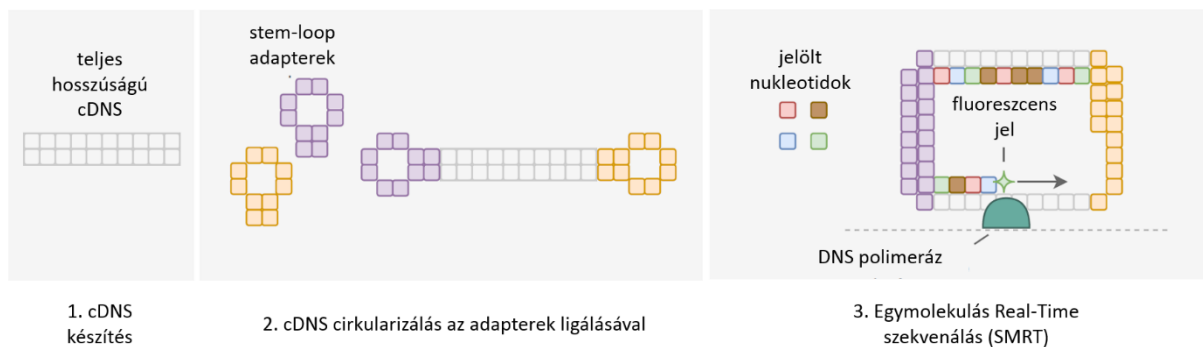


6. ábra Az Illumina SRS működésének alapelvei, az SBS technológia²⁶². Az Illumina szekvenálás DNS (vagy RNS szekvenálás esetén cDNS) fragmenseket használ templátként, melyeket adapterek segítségével kapcsol a Flow Cellhez, majd bridge amplifikációval megsokszoroz, klasztereket képezve. A szekvenálás a fluoreszcens nukleotidok, az újonnan szintetizálódó DNS szálba való beépülésein alapul. Azokból a klaszterekből, amelyek megfelelően erős, specifikus szignált generálnak, a detektor érzékeli a fluoreszcens jelet.

3.2.2 A PacBio technológia elve

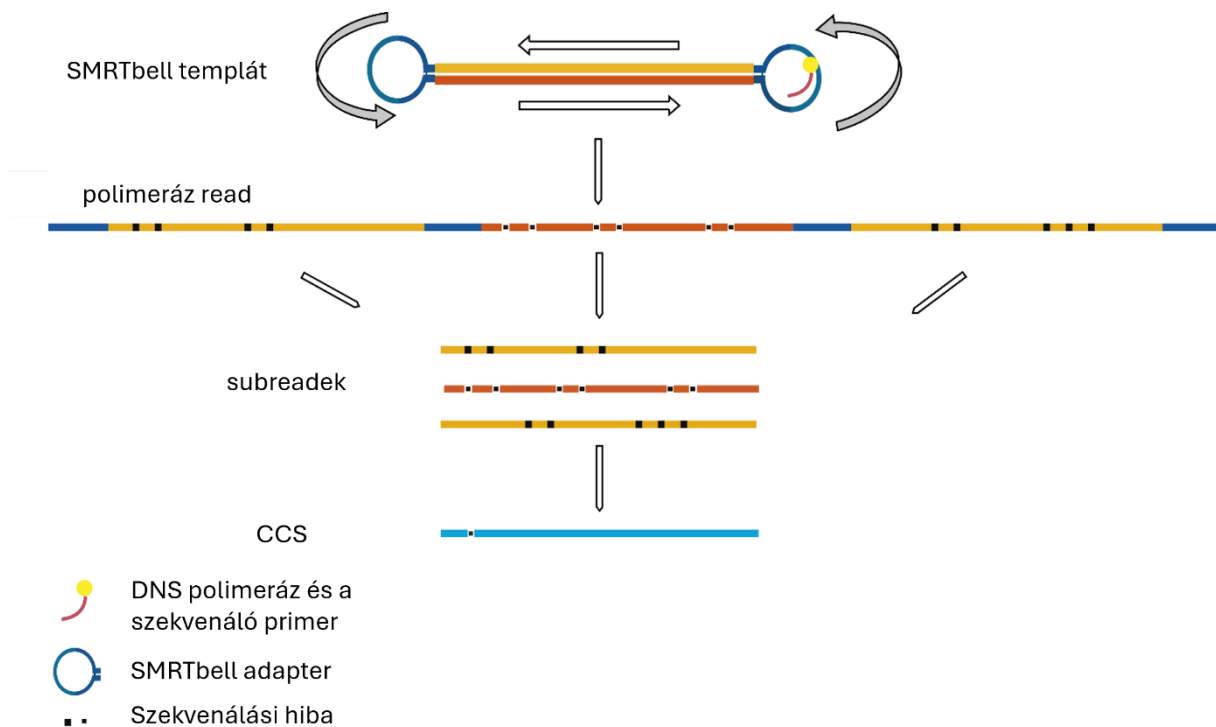
A PacBio szekvenálás egy forradalmi módszer²⁶¹, a DNS polimerizáció közvetlen megfigyelésén alapszik a célmolekulák DNS replikációs folyamata során, fluoreszcensen jelölt dezoxi-nukleotidok (dNTP-k) és egy mérnöki úton módosított DNS polimeráz^{263,264} segítségével. Az első, Real-Time Sequencer (RS) nevű készülékük a 2010-es évek végén nagyon kis közösség számára vált elérhetővé. Az RSII 2013-ban került piacra, amit a Sequel (majd a Sequel IIe) váltott 2015-től, a Revio pedig 2022-ben jelent meg.

A módszer hivatalos neve Single Molecule, Real-Time (SMRT) szekvenálás, mely lehetővé teszi egyetlen molekula valós idejű olvasását anélkül, hogy a mintát amplifikálni kellene. E módszer jelentős előrelépést jelent a genomikában, mivel hosszú leolvasásokat eredményez, nagy pontossággal²⁶¹. A PacBio technológia egyik alapvető specifikuma az úgynevezett Zero-Mode Waveguide (ZMW) Nanowell technológia. A DNS-molekulákat egy chipre viszik, amely az RSII esetén ~ 150 ezer, a Sequelnél és a Revionál több millió pici, ún. "nanowell"-t tartalmaz. Mindegyik nanowell egy-egy ZMW, amely olyan apró, hogy csak egyetlen DNS-molekulát enged be (10^{-21}). Egyetlen DNS-polimeráz enzim – mely a ZMW aljához rögzített – is csatlakozik minden DNS-molekulához, amely a szekvenálási reakciót katalizálja. A technika másik sajátossága a real-time, azaz valós idejű optikai észlelés²⁶¹. Amikor a DNS-polimeráz elkezdi szintetizálni a komplementer DNS-szálat, minden beépített nukleotid saját, jellemző fluoreszcens jelöléssel rendelkezik. A ZMW nanowell-ek lehetővé teszik, hogy csak azok a fotonok legyenek észlelve, amelyek a nanowell alján történő szintézis során – a nukleotid beépülésekor – keletkeznek. Ez azt jelenti, hogy a berendezés valós időben képes észlelni és rögzíteni minden egyes nukleotid beépülését, ami hosszú és pontos szekvenciák meghatározását teszi lehetővé. A PacBio szekvenálásnak számos előnye van más szekvenálási módszerekkel szemben, köztük a hosszú olvasások képessége, amelyek megkönnyítik a genomok összeszerelését és a strukturális variánsok meghatározását. A PacBio platform különösen alkalmas kisméretű, nagy GC-tartalmú és nagy mennyiségű ismétlődő szekvenciát tartalmazó genomok szekvenálására^{4,265}. A PacBio előnye más módszerekkel szemben, hogy nem okoz szisztematikus hibákat, s a nagy konszenzusos pontosságának köszönhetően minden felmerülő hibát tulajdonképpen ki is javít²⁶⁶. A technológia hátránya a nagy költségigény és a viszonylag alacsonyabb átteresztőképesség. A módszer speciális könyvtárkészítést is magában foglal: úgynevezett SMRTbell templátokat készítünk, melyek topológiailag cirkulárisak, míg strukturálisan lineárisak. Mindezt a kétszálú cDNS két végére ligálással rákapcsolt hurokszerű (stem-loop) adapterek biztosítják (7. ábra).



7. ábra A PacBio szekvenálás elve²⁶². A PacBio szekvenálás során a teljes hosszúságú cDNS-t speciális, ún. stem-loop adapterekkel körkörösé alakítják, majd a szekvenálás ZMW-kben történik, ahol a rögzített DNS polimeráz fluoreszcensen jelölt nukleotidokat épít be. A ZMW-k képesek egyetlen nukleotid beépítésének fluoreszcens jelét érzékelni. A polimeráz többször is áthaladhat a cDNS templáton; így a konszenzus leolvasások pontosabbak, mint a nyers olvasatok.

A szintézis, azaz a szekvenálás során a DNS polimeráz egy-egy ilyen SMRTbell templátot (**8. ábra**) sokszor átír, ezáltal egyetlen DNS molekuláról akár több tucat leolvasás is keletkezhet. Ezeket egyenként subreadeknek nevezzük, míg a subreadeket egyben polimeráz readnek^{131,262}. A subreadekből a PacBio szoftvere (ezt alkalmanként változtatják, pl. korábban SMRT Analysis, majd SMRT Link, s ezek újabb és újabb verziói) cirkuláris konszenzus szekvenciákat (CCS) állít elő, melyek nagyon nagy pontosságú leolvasások. [A CCS-t korábban az inszertek leolvasásainak (reads of insert, ROI) nevezték].



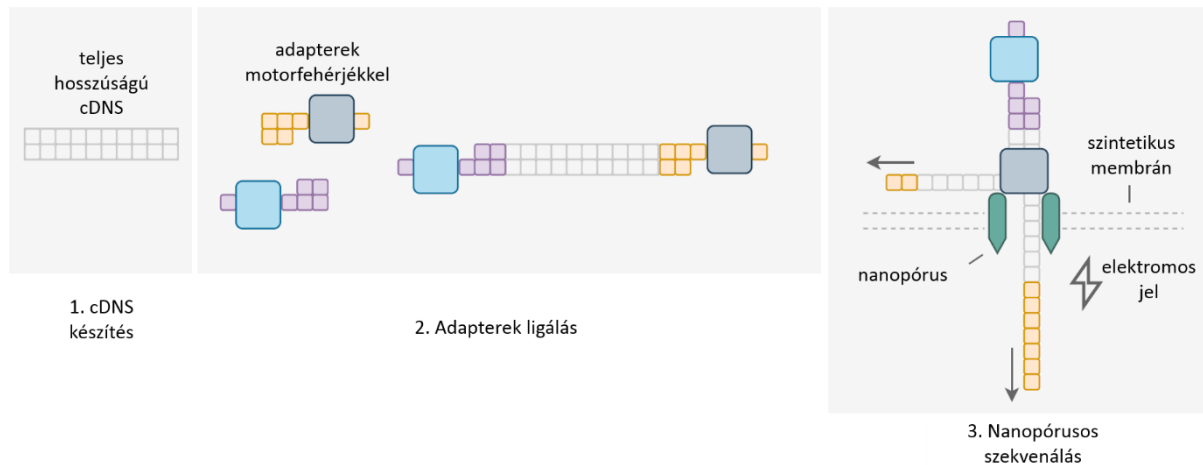
8. ábra Az SMRTbell templát szekvenálásának sematikus reprezentációja¹³¹. Az ábra a szekvenálás folyamatát és a terminológiát mutatja. Látható, hogy az inszertet többször is leolvassa a polimeráz: a teljes szekvenciát, beleértve az adaptereket is, polimeráz olvasásnak nevezzük. Miután az adapter szekvenciákat eltávolítjuk, a szekvencia subreadekre oszlik. A körkörös, vagy cirkuláris konszenzus szekvencia (CCS), amelyet korábban az inszert leolvasásának (reads of insert; ROI) is neveztek, az azonos inszert többszöri leolvasásainak nagyon jó minőségű konszenzus szekvenciája.

3.2.3 A nanopóruson alapuló szekvenálás elve

Az Oxford Nanopore Technologies (ONT) által fejlesztett nanopórusos szekvenálás a PacBio-hoz hasonlóan, lehetővé teszi hosszú DNS- és RNS-szakaszok valós idejű, egyetlen molekula-szintű leolvasását. A technológia alapját a biológiai nanopórusok képezik, melyek integrálódnak egy elektromosan ellenálló membránba (**9. ábra**). A nanopórus egy nanométer nagyságrendű lyuk, ami egy elektrolit oldattal töltött mikroszkopikus kamrában helyezkedik el. A kamra két részre oszlik, amelyeket a nanopórust tartalmazó membrán választ el^{261,262}. Mindkét oldalon ionokat tartalmazó oldatot helyeznek, amely lehetővé teszi az elektromos áram áthaladását a nanopóruson keresztül. A membrán két oldala között feszültséget alkalmaznak, ami ionáramot generál a nanopóruson keresztül. Amikor a DNS vagy RNS molekula áthalad a nanopóruson, akadályt képez az ionáram útjában, ami átmeneti változást okoz az áram erősségében; minden egyes bázis specifikus változást okoz²⁶¹. Az áramban bekövetkező változásokat észleli és elektromos jelekké alakítja a számítógépes rendszer, majd a jeleket számítógépes algoritmusok elemzik, s meghatározzák az

áthaladó nukleinsav molekula bázissorrendjét. Ezt az ONT bázishívó programja, azaz basecallere végzi, amely a nyers szignált nukleotiddá alakítja. Az ONT basecallerek folyamatosan frissülnek, Albacore, Guppy, majd legújabbán Dorado, s ezek újabb verziói.

A nanopórusos szekvenálás előnyei közé tartozik a hosszú leolvasások mellett a valós idejű adatgyűjtés, valamint az előkészítési és amplifikációs lépések szükségességének csökkenése. Ez a technológia különösen hasznos komplex genetikai régiók, mint például nagyméretű ismétlődések vagy szerkezeti variációk vizsgálatához, valamint epigenetikai módosítások, mint például a metiláció közvetlen detektálásához²⁶².



9. ábra Az ONT szekvenálás elve²⁶². A cDNS templátokhoz kapcsolt motor fehérjék az ONT (pl. a MinION készülékkel történő) szekvenálás során a szintetikus membránon lévő nanopórusokhoz kötődnek. A motor fehérjék az egyik DNS szálát áthúzzák a nanopóruson, miközben egy érzékelő méri a membrán két oldala közötti potenciálváltozásokat. Az elektromos jel specifikus a membránon áthaladó nukleotidokra.

3.2.4 A Loop-Seq technológia elve

Az elmúlt két évtizedben az NGS technológia fejlődése jelentősen növelte a kapacitást, s egyre több genomikai adatot nyújt alacsonyabb költségen, azonban továbbra sem versenyképes az LRS módszerekkel számos tekintetben. A módszer működésénél fogva nem tud néhány száz bp-nál hosszabb leolvasásokat produkálni. Ennek a korlátozásnak a leküzdésére kidolgozták az SLRS könyvtár készítési módszert²⁶⁷, melyben molekuláris tag-eket (~ címkék, azonosítók) és korlátozott mértékű hígítást alkalmaztak a könyvtárak előkészítése során, majd a tag-eket használták hosszabb szekvenciák összeállítására. Később a 10X Genomics kereskedelmi forgalomba hozott egy mikrofluidikai eszközt, amely növelte a kapacitást és egyszerűsítette a munkafolyamatot. A LoopSeq módszer tovább egyszerűsítette a technológiát, megkerülve a mikrofluidika szükségességét és kibővítve az alkalmazási lehetőségeket²⁶⁸.

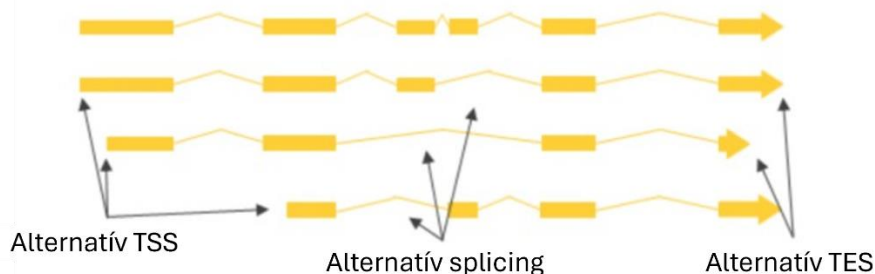
A LoopSeq a hagyományos SRS megközelítéseket alkalmazza a saját barkódolási (barcode) technológiájával kombináltan, ezáltal hosszú-readekkel ekvivalens adatokat produkál az Illumina platformon. Az azonos barcodeokkal rendelkező leolvasásokat egy csoportba gyűjti, majd a cég saját szoftverével hosszabb molekulákká állítja össze. Az egyik fő előnye a LoopSeq technológiának a rendkívül alacsony hibaráta bármely más LRS technológiához képest (ennek elsősorban genom szekvenálásnál van jelentősége). Ezenkívül az SLRS egy független megközelítést nyújt, amely révén alkalmazható más technikák által nyert eredmények megerősítésére.

3.2.5 A hosszú-readek szerepe a transzkriptomikában

Az LRS előnye a transzkriptomikában az ún. izoforma szekvenálás: olyan RNS variánsok feltárását teszi lehetővé, mely a klasszikus NGS technikával és más genomikai módszerekkel [úgy, mint a kvantitatív real-time RT-PCR (qRT-PCR), DNS chip (microarray)] megközelítéssel lehetetlen (10. ábra). Az LRS képes a transzkripció kezdeti (transcription start site; TSS) és végpontjai (transcription end site; TES)], valamint az alternatív splicing által létrejött transzkript izoformák (TI) kombinációinak azonosítására²⁶². Köszönhetően, hogy az LRS lehetővé teszi a teljes RNS molekulák egyetlen leolvasásban történő szekvenálását, így a többszörösen spliceolt izoformák exon kapcsolatai is megbízhatóan azonosíthatók²⁶⁹. Továbbá, e technikával elkülöníthetők egymástól a parallel átfedő transzkriptek (pl. átfedő mono- és poligénus RNS molekulák)²⁶⁹, melynek jelentősége például ezen transzkriptek mennyiségi meghatározásakor felbecsülhetetlen. A legújabb LRS kitek már közvetlenül a natív RNS szekvenálására is alkalmasak²⁷⁰, ezáltal a könyvtárkészítésből származtatható esetleges torzító tényezők is kiküszöbölhetők, s így lehetőséget biztosítanak a transzkriptom precízebb feltérképezésére mind strukturális, mind mennyiségi értelemben. Emellett, a natív RNS szekvenálás lehetővé teszi az epitranszkriptomikai módosulások detektálását is²³.

Mindezen előnyök ellenére az LRS-nek vannak bizonyos technikai (pl. kevésbé pontos leolvasások) és költségbeli korlátai is, de a transzkriptomikai kutatásokban való alkalmazása egyre növekvő trendet mutat az általa nyújtott részletes és átfogó genetikai információk miatt. Megjegyzendő, hogy a leolvasásokat érintő pontatlanság, a transzkriptom kutatásban nem jelent semmiféle hátrányt. Továbbá, az ONT és a PacBio folyamatosan fejleszti szekvenáló enzimeit és reagenseit, a Flow és SMRTCelleket, melyekkel egyre nagyobb precizitást biztosítanak e tekintetben is.

Transzkript izoformák



Rövid-read szekvenálás



Hosszú-read szekvenálás



10. ábra Az LRS megfelelő információkat nyújt az izoformákról²⁶⁹. Az egyes TSS-ek, TES-ek és splice junction-ök SRS-sel is meghatározhatók, azonban ezeknek a jellemzőknek a kombinációját nehéz megkülönböztetni, ha több izoforma található ugyanazon a lokuszon. Az LRS másrészt képes teljes hosszúságú transzkriptek leolvasására, amelyek teljes kontig információt nyújtanak az egyes RNS-ekben szereplő exonokról.

3.3 VÍRUS TRANZKRIPTOMOK ANALÍZISE AZ LRS-ALAPÚ KUTATÁSAINK ELŐTT

A vírusgenomok annotálása, a gének lokációjának meghatározása korábban elsősorban az ORF-ek észlelésével történt²⁷¹. A szekvenálás előtti módszerek korlátozott betekintést engedtek a vírusok transzkripció aktivitásába. A gének funkcionális vizsgálata, a virális génexpresszió analízise az SRS-t megelőzően többnyire indirekt technikákkal történt elsősorban, mint a Northern blot, DNS chip és qRT-PCR. Ezek a módszerek vírus-specifikus próbákat, primereket használnak a vírus RNS-ek kimutatására a mintákban. Habár e technikák specifikusak, azonban korlátozottak az alacsony áteresztőképesség és a magas RNS-igény miatt. Továbbá lévén nem szekvenciafüggetlen módszerek, kizárólag ismert nukleinsavak kimutatására alkalmasak. A Northern blottal szemben a qRT-PCR nagy érzékenységgel és specifitást is biztosít egyes RNS szekvenciák detektálására és kvantitálására. Azonban a teljes transzkriptom széles körű vizsgálatát ez sem teszi lehetővé, és szintén csak előre meghatározott szekvenciák detektálására, vizsgálatára alkalmazható.

Bár ezek a módszerek fontos információkat szolgáltatottak, korlátozottak voltak a transzkriptom teljes mélységének és komplexitásának feltárásában. A szekvenálás előtti módszerekkel a vírus transzkriptomoknak csak egy részét lehetett megismerni, és az adatokat gyakran nehéz volt értelmezni a magas háttérzaj és a technikai korlátok miatt.

PhD munkám elsősorban a qRT-PCR technikára fókuszált. E módszerrel vizsgáltam meg a PRV összes, fehérje kódoló génjének kinetikai profilját, ami újdonságnak számított, hiszen addig teljes vírus transzkriptomot nem vizsgáltak e módszerrel, s az elemzéséhez egy újfajta matematikai megoldást kellett dolgoznunk².

Az SRS technikák megjelenése a vírusok transzkriptomikai kutatását is fellendítette. A modern módszerek lehetővé tették a vírusok és gazdasejtjeik genetikai interakcióinak részletes és átfogó vizsgálatát. Az SRS módszerek a herpeszvírusok analízisében is egyre nagyobb szerepet kaptak, különösen a TSS-ek és TES-ek detektálására alkalmazták. Relatív sok publikáció született, a 2010-es évektől kezdve egyre többen foglalkoztak SRS megközelítéssel, míg csoportunk 2015-ben publikált először SRS szekvenáláson alapuló transzkriptomikai eredményeket^{3,272}.

Az LRS módszerek megjelenésével szinte egyidőben, mintegy úttörőként, bevezettük először a PacBio módszereit (amplifikált és nem amplifikált könyvtárkészítés) a virális RNS profilok elemzésére. Első PacBio publikációnk 2014-ben jelent meg⁴, elsőként szekvenáltunk meg a PRV teljes genomot, majd 2015-ben és 2016-ban elsőként közöltünk LRS-sel meghatározott vírus transzkriptomot^{6,272}. Az ONT különféle megközelítéseit [amplifikált cDNS szekvenálás, direkt (d)cDNS és dRNS szekvenálás] is elsőként alkalmaztuk vírusok esetén^{13,16,273,274}.

Az LRS technikák megsokszorozták a korábban ismert herpeszvírus transzkriptek számát. A virális transzkriptek pontos, teljes hosszúságú annotálása mellett ezek a tanulmányok eddig ismeretlen mRNS-eket, ncRNS-eket, poligénes RNS-eket, valamint különböző transzkript izoformákat azonosítottak, beleértve az alternatív spliceolást, valamint a TSS és TES variánsokat.

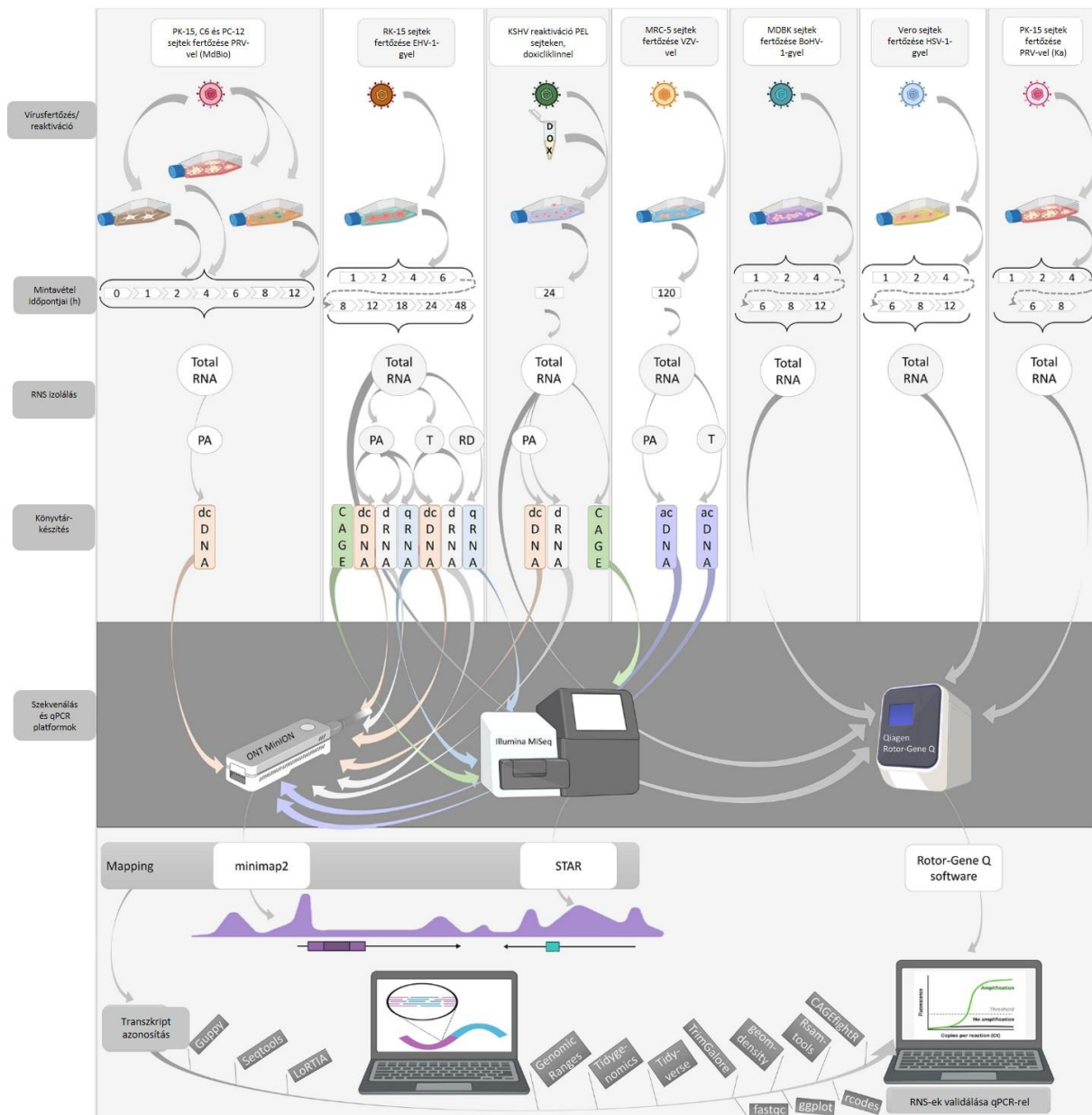
4 CÉLKITŰZÉSEK

A szekvenálási technológiák, a genomika és a bioinformatika térnyerése, egyre növekvő szerepe a molekuláris biológiában az utóbbi évtized(ek)ben megkérdőjelezhetetlen. A virológia területén e módszerek alkalmazása más – humán genomikai, állat- és növénybiológiai, bakterológiai – területekhez képest megkésett. A 2000-es évek első évtizedében elvétve foglalkoztak csak a vírusok összgenomszintű funkcionális vizsgálataival, azokat javarészt klasszikus technikákkal végezték. Csoportunk 2006-tól kezdődően foglalkozik vírusokkal genomikai aspektusból. A csoportunk által megfogalmazott, a transzkriptek összgenom szintű szabályozó szerepével foglalkozó 'Transzkripciós Interferencia Hálózat Hipotézis'²⁷⁵ egyes elemeinek kísérletes bizonyításához elsősorban a PRV-t alkalmaztuk. A vírust egyrészt, mint fertőző ágenszt szeretettük volna minél jobban megismerni, feltárni a molekuláris mechanizmusait, ugyanakkor feltételeztük, hogy a hipotézisünk általános érvényű lehet, tehát a vírust modellként is alkalmaztuk. Fel kellett térképezzük a PRV RNS-ek komplex világát a lehető legteljesebb mértékben. A PRV megismerését követően a további közeli, majd távolabbi rokon vírusok mély, transzkriptomikai elemzése volt a következő fontos állomásunk, mellyel a PRV-nél tapasztalt komplexitás általánosságáról kaphattunk bizonyosságot. Továbbá, a vírusok közti eltérések, egyes vírusokra, vírussaládokra jellemző egyediségek felismerése az alkalmazott technológiákat is validálta. Az évek során tehát a következő célkitűzéseket fogalmaztuk meg:

1. **Technológiai innováció a molekuláris virológiában:** a vírustranszkriptomok újradefiniálása, feltárása kombinált szekvenálási 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, hipereditáltak
3. **Funkcionális genomika**
 - a. olyan mutáns vírusok léterhozása, melyek transzkripciós 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 (*in silico*); 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 a fertőzés leküzdésének stratégiáit
 - a. a vírusfertőzés gazdasejtekre gyakorolt hatásának elemzése
 - b. a sejtvonal a virális génextpresszióra gyakorolt hatásának vizsgálata

5 MÓDSZEREK

A disszertációban bemutatott eredmények alapját képező kísérleteknél számos virológiai, molekuláris biológiai, genomikai és bioinformatikai technikát alkalmaztunk. A metodikák részletes leírása, a pontos protokollok megtalálhatóak az értekezés alapját képező publikációkban. Így e fejezetben a módszerekről csak kevésbé részletesen (pl. a konkrét mennyiségek megnevezése nélkül) írok. Az értekezésben bemutatott adatok/eredmények egy részének alapját a **11. ábrán** vázolt munkafolyamatok szolgáltatták.



11. ábra **Munkafolyamat**⁸². Az ábra az elvégzett szekvenálások egy részét mutatja, a herpeszvírusokra fókuszáltnak. Látható, hogy mely vírusok esetén mely sejtvonalakat alkalmaztuk, milyen könyvtár típusokat készítettünk, mely szekvenáló platformokon végeztük a szekvenálást, illetve a bioinformatikáról és a validálásról is bemutat egy szűk keresztmetszetet. Rövidítések, szövegmagyarázat: Total RNA: össz RNS; PA: poliadenilált RNS; T: Terminátorral kezelt minták; RD: ribodepletált RNS-ek; dcDNA: direkt cDNS-Seq; CAGE: Cap Analysis Gene Expression-Seq; dRNA: direkt RNS-Seq; qRNA: SRS könyvtár NEXTFLEX qRNA-Seq kittel (PerkinElmer); acDNA: amplifikált cDNS-Seq.

5.1 FELHASZNÁLT VÍRUSOK, ALKALMAZOTT SEJTVONALAK

A projektjeink során több mint húszféle vírussal dolgoztunk, melyeket az alábbi táblázatban foglaltam össze (**1. táblázat**). A vírusok szaporításához használt sejtvonalakat, illetve ezek és a vírusok eredetét is feltüntettem, a vonatkozó publikációinkkal együtt.

A projektjeink során alkalmazott vírusok és sejtek, sejtvonalak			
Vírusok	Sejtvonal	Vírus, sejt eredete	Publikáció
Afrikai sertéspestis vírus (African Swine Fever virus, ASFV) ASFV_HU_2018 (ID Number: MN715134)	Porcine alveolar macrophage (sertés alveoláris makrofág; PAM)	AFSV: saját izolátum	20, 276
Autographa californica multiple nucleopolyhedrovirus (AcMNPV) LacZ rekombináns: βgal-AcMNPV	Spodoptera frugiperda (Sf9) sejtvonal	A vírust és a sejtvonalat is a SOLVO Biotechnology-tól kaptuk (Szeged, Hungary)	273, 277
Pseudorabies virus (PRV) Kaplan (Ka) törzs	Porcine kidney-15 epithelial cell (Sertésvese epitél sejtek)	Ka: Vanderbilt University, Nashville, TN, United States PK-15: American Type Culture Collection (ATCC)	6, 7, 13, 19, 272, 278
AyV MdBio törzs	PK-15 Rat (<i>Rattus norvegicus</i>) glial cell (patkány glia; C6) Rat pheochromocytoma PC12	PRV: saját izolátum, Szerbia PK-15, C6, PC12: ATCC	131
AyV Ka, ep0 mutáns	PK-15	PK-15: ATCC	279
AyV Ka, ul54 mutáns	PK-15	PK-15: ATCC	280
AyV Ka, VHS mutáns	PK-15	PK-15: ATCC	281
Bovine alphaherpesvirus 1 (BoHV-1) Cooper törzs	Madin-Darby szarvasmarha vese (Madin-Darby Bovine Kidney, MDBK) Juhvese (OK)	BoHV-1: ATCC MDBK: ATCC OK: ATCC	18, 282- 284
Coxsackie B vírus (Coxsackievirus, group B serotype 5, CVB5)	African green monkey (<i>Chlorocebus sabaeus</i>) kidney (Vero) cells	CVB5: Nemzeti Élelmiszerlánc-biztonsági Hivatal (NÉBIH)	285
Epstein-Barr vírus	Akata, Epstein-Barr virus (EBV)-'termelő' sejtvonal	EBV: ATCC Akata: ATCC	22
Equid alphaherpesvirus 1 (EHV-1)	rabbit kidney (RK-13) epithelial cells	EHV-1: ATCC RK-13: ATCC	25
Humán citomegalovírus (Human Cytomegalovirus, HCMV) Towne VarS törzs	Humán magzati tüdő fibroblaszt (Human lung fetal fibroblast cells, MRC-5)	HCMV: ATCC MRC5: ATCC	9, 21, 286, 287
Herpesz simplex vírus 1-es típus (Herpes simplex virus type 1, HSV-1), KOS törzs	Vero, African green monkey kidney cells (afrikai zöldmajom vese sejtek)	HSV-1: ATCC Vero: ATCC	8, 16, 288- 289
Influenza A vírus A/Puerto Rico/8/1934 H1N1 (IVA)	Madin-Darby Canine Kidney (MDCK) cells	IVA: Országos Epidemiológiai Központ	285
Kaposi's sarcoma-associated herpesvirus (KSHV) TREx	KSHV-positive primary effusion lymphoma (PEL) cell line iBCBL1-3xFLAG-RTA102	KSHV pozitív PEL: JU. Jung (University of Southern California, B. Papp & Z Toth University of Florida)	26 290
Krími-kongói vérzéses láz vírusa (Crimean-Congo hemorrhagic fever virus Kosova Hoti strain, CCHFV)	Vero	CCHFV: izolátum a Balkán- félszigetről (Duh 2008)	285

Majomhimlő vírus (Monkeypox virus, MPXV) MPXV_NRL 4279/2022	African green monkey (<i>Chlorocebus sabaeus</i>) kidney fibroblast (CV-1) cells	MPXV: izolátum, Dr. Jirincova (The National Institute of Public Health, Prague, Czech Republic) CV-1: ATCC	291
Nyugat-nílusi vírus (West Nile virus, WNV)	Vero	WNV: saját izolátum, Szerbia	285
Porcine circovirus-1 (PCV-1) *	PK-15	PK-15: ATCC	178, 292
Porcine endogenous retrovirus (PERV) *	PK-15	PK-15: ATCC	293, 294
Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2)	CV-1	SARS-CoV-2: saját izolátum Vero: ATCC	24
Vaccinia vírus (Vaccinia virus, VACV) Western Reserve törzs	CV-1	VACV: NÉBIH CV-1: ATCC	17, 274, 295-297
Varicella Zoster vírus (VZV) OKA/Merck törzs	MRC-5	VZV: MSD Pharma Hungary Kft MRC-5: ATCC	14, 298
Vesicular stomatitis Indiana vírus (VSIV)	Vero és Human glioblastoma (T98G)	VSIV: Orvosi Mikrobiológiai Intézet Vero, T98G: ATCC	285
Zika vírus MR766 (ZIKV)	Vero	ZIKV: European Virus Archive; Vero: ATCC	285

1. táblázat A disszertáció alapját képező vírusok. A táblázatban bemutatom a vizsgált vírustörzseket, azok eredetét, illetve a szaporításukhoz alkalmazott sejtvonalakat. * A PRV fertőzött PK-15 sejttel kapcsolatos kísérletekből.

5.2 VÍRUSFERTŐZÉS MENETE, IDŐZÍTÉSES KÍSÉRLETEK

A vírusfertőzés folyamatát a PRV Kaplan törzsének példáján mutatom be részletesen, majd a többi vírussal kapcsolatban csak az ehhez képesti eltéréseket jelzem. Azon kísérleteknél, ahol a vírusfertőzés gazdagének expressziójára kifejtett hatását is vizsgáltuk, nem fertőzött, ún. mock-infected sejteket is szaporítottunk, majd ugyanazon módszerrel szekvenáltuk is, mint a kontrollt. A dinamikus kísérleteket időpontoként legalább három biológiai párhuzamossal végeztük. A kivételeket (technikai párhuzamos) jelzem.

A ZIKV, WNV, CCHFV, SARS-CoV-2 és MPXV vírusok szaporítását, az RNS izolálást és könyvtárkészítést a Pécsi Tudományegyetem (PTE) Szentágothai János Kutatóközpont, BSL-4-es szintű laboratóriumában, az ASFV szaporítását és az RNS izolálást az Állatorvostudományi Kutatóintézet (ÁTKI) BSL-3-as szintű laboratóriumában, az EBV-vel kapcsolatos munkát az RT-Europe Research Center laboratóriumában végeztük. Az összes többi vírussal a megfelelő BSL-1-es, vagy BSL-2-es szintű laboratóriumokban dolgoztunk az SZTE-n, valamint a NÉBIH-ben. A KSHV-t RNS formájában floridai kollaborációs partnerunktől, míg a BoHV-1-et Mississippiből kaptuk.

5.2.1 PRV Kaplan törzs

A PK-15 sejteket 37°C-on, 5% CO₂ jelenlétében tartottuk fenn 5% magzati szarvasmarhaszérummal [fetal bovine serum (FBS)] kiegészített Dulbecco-féle módosított Eagle tápoldatban (DMEM; Thermo Fisher Scientific). Az oldathoz gentamicint (Thermo Fisher Scientific) adtunk. Gyorsan növekvő, szemi-konfluens PK-15 sejteket 10 MOI plakk-képző egységnek (pfu) megfelelő PRV-vel fertőztünk, majd 37°C-on 5% CO₂ mellett a teljes citopátiás hatás eléréséig inkubáltuk. A sejteket háromszor fagyasztottuk-olvasztottuk, majd 10,000 g-vel 15 percig centrifugáltuk. Ezt követően

határoztuk meg a titeret. A vírus stock-okat használtuk a további kísérleteinkhez. A kísérletek során nagy titerrel (10 pfu/sejt) fertőztünk sejteket, majd 1h-n át inkubáltuk, a felülúszót eltávolítottuk. A sejteket foszfáttal pufferelt sóoldattal (PBS) mostuk, majd új tápoldatot adtuk hozzájuk és 1, 2, 4, 6, 8, illetve 12 órán át inkubáltuk. A vírusok dinamikus elemzéséhez az egyes időpontokból külön izoláltunk RNS-t, készítettünk könyvtárat és független SMRTCelleken (PacBio), illetve Flow Celleken (ONT és Illumina) szekvenáltuk meg, vagy, amennyiben egyazon futtatásban kerültek megszekvenálásra, akkor előtte barcodeoltuk őket. A statikus elemzéshez az időpontokból RNS keveréket, mixet készítettünk, s ezt vittük végig a teljes procedúrán. A sejteket -80°C-on tároltuk.

Az inkubáció minden vírus esetén 37°C-on, 5% CO₂ mellett történt, ha eltérés volt jeleztem, ugyanígy a fertőzést követő inkubációs időpontok is akkor kerülnek a továbbiakban feltüntetésre, ha a PRV Ka-tól eltérő időpontokat alkalmaztunk. A sejteket mindig -80°C-on tároljuk.

5.2.2 További PRV törzsek

PRV MdBio

A PK-15, C6 és PC12 sejteket 1 MOI PRV MdBio vírussal fertőztük.

PRV Ka u154Δ

A PK-15 sejteket 0,1 MOI *u154* deletált PRV-vel fertőztük, majd 1, 2, 4, 6, és 8 h-n keresztül inkubáltuk. Azért csak 8 h-ig vizsgáltuk, hogy a fertőzött sejtekből kikerülő új vírusok a nem fertőzött sejtekben ne tudjanak új fertőzési ciklust indítani (lévén kistiterű fertőzés).

PRV Ka ep0Δ

A fertőzött PK-15 sejteket 10 pfu/sejt vírussal fertőztük, majd 0, 1, 2, 4, 6, 8, 12, 18 és 24 h-n keresztül inkubáltuk.

PRV Ka VHSA

A fertőzött PK-15 sejteket (10 MOI) 0, 1, 2, 4, 6, 8, 12, 18 és 24 h-n keresztül inkubáltuk.

5.2.3 Egyéb herpeszvírusok

BoHV-1

Az MDBK sejteket 100 U/ml penicillint és 100 µg/ml sztreptomocint, ill. 5% FBS-t tartalmazó MDBK tápoldatban tartottuk fenn; kis titerű fertőzést alkalmaztunk (0.1 MOI) a vírus stock elkészítéséhez. A kinetikai analízishez 5 MOI-t használtunk.

Az OK sejtek fenntartásához 10% újszülött borjú szérumot (NCS) és antibiotikumokat tartalmazó DMEM-et használtunk. A kísérletekhez 1 MOI vírussal fertőztük a sejteket, a szaporítást NCS-mentes DMEM-ben végeztük.

EHV-1

Az RK-13 sejteket 10% magzati borjúszérumot [fetal calf serum (FCS)] és 80 µg/ml of gentamicint tartalmazó DMEM-ben szaporítottuk. A vírus stock-ok elkészítéséhez 0,1 MOI vírussal fertőztük a sejteket, míg a szekvenáláshoz 4 MOI-val. Három technikai párhuzamossal dolgoztunk időpontonként.

HSV-1

A Vero sejteket 10% FBS-t és penicillin-sztreptomycin 10K/10K keveréket (Lonza) tartalmazó DMEM oldatban szaporítottuk. A fertőzéshez 1 MOI (pfu)/sejt vírust alkalmaztunk.

VZV

Az MRC5 sejteket 1% antibiotikum/antimikotikum oldatot (Lonza) és 10% FBS-t (GE Healthcare Life Sciences) tartalmazó DMEM oldatban tartottuk fenn, a kísérletekhez 1 MOI-val fertőztük a sejteket, amit 5 napig inkubáltunk.

HCMV

Az MRC-5 sejteket 10% FBS-sel és penicillin/sztreptomocinnel (Lonza) kiegészített DMEM-ben tartottuk fenn. A sejteket a HCMV Towne VarS törzsével (0,5 pfu/sejt) fertőztük. A fertőzést követően 1, 3, 6, 12, 24, 72, 96 és 120 h-val állítottuk le az inkubációt.

KSHV

A transzgenikus PEL sejtvonalon a litikus reaktivációt doxiciklinnel indukáltuk a litikus transzkriptom vizsgálatához, az RNS-t a reaktivációt követő 24 h-s mintából izoláltuk.

EBV

Az Akata sejteket RPMI-1640 tápoldatban tartottuk fenn, amit 10% FCS-sel és penicillin/sztreptomycin-nel egészítettünk ki. A sejteket mostuk, majd újra szuszpendáltuk 106 sejt/ml koncentrációban RPMI oldatban, amit kecske anti-humán IgG-vel (Jackson ImmunoResearch) egészítettünk ki. Kontrollként normál RPMI-ben is szuszpendáltunk sejteket²⁹⁹. A reszuszenziót követően 100 µl sejtszuszenziót használtunk fel 10 perc, 90 perc, 4, 12, 24, 48 és 72 óra elteltével az RNS izolálásához, melyből qRT-PCR-rel ellenőriztük a transzkripció indukció sikerességét. A BZLF és a GP350 gének aktivitását mértük, referencia génekhez normalizálva.

5.2.4 További nagy, DNS genomú vírusok**VACV**

A CV-1 sejteket 10% FBS-t és antibiotikum-antimikotikum oldatot tartalmazó RPMI 1640 tápoldatban (Sigma-Aldrich) szaporítottuk. A fertőzést 10 MOI vírussal végeztük, majd ezt követően 1, 2, 3, 4, 6, 8, 12 és 16 h-val a fertőzést követően állítottuk le a kísérletet.

MPXV

A vírust egyszer passzáltuk CV-1 sejteken, majd plakk esszével határoztuk meg a titert. A transzkriptomikai kísérletekhez 2 ml MPXV-t [5 pfu/sejt (MOI = 5) vírust] alkalmaztunk, amely (Minimum Essential Medium, Eagle) MEM oldatban került a sejtekhez hozzáadásra. Az 1h-s inkubáció során 10 percenként kevertük (shakeltük) a mintákat. A sejteket mostuk, majd 10 ml MEM oldatot adtunk hozzá, melyet 2% FBS-sel, 2 mM L-glutaminnal és 1% penicillin-sztreptomycin oldattal egészítettünk ki. A sejteket 1, 2, 4, 6, 12 és 24 órán át inkubáltuk.

ASFV

A PAM sejteket az OIE Manual's alapján (Office International des Epizooties, 2019) szaporítottuk RPMI 1640 tápoldatban, mely L-glutaminnal (Lonza) és 10% FBS-sel, (Euro Clone), 1% Na-piruváttal (Lonza), 1% antibiotikum-antimikotikum oldattal (Thermo Fisher Scientific) és 1% nem esszenciális aminosav oldattal (Lonza) volt kiegészítve. A vírus stock titerét immunfluoreszcens assay-vel (esszével) határoztuk meg²⁷⁶.

AcMNPV

Az Sf9 epitél sejteket Sf-900 II SFM tápoldatban tartottuk fenn (Thermo Fisher Scientific). A vírusfertőzött sejteket Corning Spinner Flaskákban (Sigma Aldrich/Merck) szaporítottuk 26 °C-on, a rázógépet 70 rpm értékre állítva. A vírus egy LacZ expresszálo rekombináns vírus, mely gén a *polh* gén promóter régiójába van ültetve. A sejteket 2 MOI pfu/sejt vírussal fertőztük, majd 0, 1, 2, 4, 6, 16, 24, 48, illetve 72 h-ig inkubáltuk.

5.2.5 RNS vírusok**SARS-CoV-2**

A Vero sejteket MEM tápoldatban tartottuk fenn, melyet 10% FBS-sel és 2 mM L-glutaminnal, valamint antibiotikum-antimikotikum oldattal egészítettünk ki (mind Sigma-Aldrich). A vírus mintavevő pálcáról való izolálásához szintén ezt a MEM-FBS-antibiotikum-antimikotikum oldatot alkalmaztuk, 1ml oldatba helyeztük a mintavevő pálcát, majd a mintát tartalmazó oldatot Ministar® 0,22-µm filterrel (Sartorius AG) szűrtük. A szűrleményt Vero sejteket tartalmazó flaskába helyeztük és 1 h-n keresztül inkubáltuk, majd MEM oldatot adtunk hozzá. A sejteken naponta ellenőriztük a citopátiás hatást (CPE), majd a 4. napon (80%-os CPE) a felülúszót begyűjtöttük. Ez jelentette az első passzázst, majd ezt követően alacsony MOI-val további kétszer passzáltuk Vero sejteken (ezt használtuk törzsoldatnak), a titert plakk esszével határoztuk meg.

VSIV

A vírust Vero és T98G sejteken szaporítottuk, MOI=5. A fertőzést 1, 6, 15 és 24 h elteltével állítottuk le.

IVA

A vírus szaporításához 'törzsoldatot', azaz stock-ot készítettünk 10 napos embrionált tojásokon, majd MDCK sejteken plakk esszével határoztuk meg a titert. A sejteket 1ml vírussal fertőztük (MOI=10), majd 1, 4 és 7 h elteltével állítottuk le a fertőzést.

CVB5

Vero sejteken szaporítottuk a vírust, fertőzést követően 1, 6, 15 és 24 h-ig inkubáltuk.

ZIKV, CCHFV, WNV

Mindhárom vírust Vero sejteken szaporítottuk, a transzkriptomikai kísérlethez nem állapítottuk meg a titert. A fertőzést 24 és 72 h elteltével állítottuk le.

5.3 A FEHÉRJESZINTÉZIS GÁTLÁSA

Az MDBK sejteket 10% FBS-t tartalmazó DMEM-ben tenyésztettük, amíg el nem érték a 60–70% konfluenciát. Ezután a tápoldatot szérumentes DMEM-re cseréltük, és 20 vagy 100 µg/mL cikloheximidet (CHX) adtunk hozzá. Egy órás inkubálást követően a tápoldatot 10 MOI-s vírusoldatra cseréltük, ugyanazzal a CHX koncentrációval, és 6 vagy 8 h-n át inkubáltuk.

5.4 PRV VIRION ÉS VÍRUS DNS IZOLÁLÁS

A virionok izolálásához PK-15 sejteket fertőztünk PRV Ka, MdBio és *u154Δ* törzseivel (MOI=0,1 pfu/sejt), s a teljes citopátiás hatás eléréséig inkubáltuk a sejteket. Háromszori fagyasztás-olvasztást követően centrifugálással távolítottuk el a sejttermelékét. A felülúszót begyűjtöttük, a virionokat ultracentrifugálással (70,000 xg, 1h; Sorwall WX Ultra 90, Thermo Scientific) választottuk el 30%-os cukor ágyon. A felülúszó eltávolítását követően a leülepedett vírust puffer oldatban [2% NaCl 5M, 1% Tris 1M (pH=6.7), 0.5% EDTA 0.2 M (pH=7)] ismét felszuszpendáltuk. A fehérjéket SDS (0.5%) és proteináz K hozzáadásával eltávolítottuk, majd a vírus DNS-t tradicionális fenol/kloroform módszerrel tisztítottuk.

5.5 RNS IZOLÁLÁS

Az össz-RNS izolálást minden esetben a NucleoSpin RNA Kittel (Macherey-Nagel), a cég által javasolt protokoll alapján, míg a riboszomális RNS-ek (rRNS) eltávolítását ribodeplációval végeztük. Bár a kit tartalmaz DNázt, mi minden esetben tovább kezeltük az izolált RNS-t az esetleges maradvány genomi (g)DNS eltávolítása céljából, ehhez TURBO DNA-free™ Kit-et (Invitrogen) használtunk.

5.5.1 PoliA szelekció / az RNS-ek 3' végének azonosítása

A poliadenilált RNS frakció kinyeréséhez munkánk során kétféle kitet alkalmaztunk: az Oligotex mRNA Mini Kit-et (Qiagen) és a Poly(A) RNA Selection Kit-et (Lexogen). E módszerrel az mRNS-eket és a poliA farokkal rendelkező egyéb [pl. hosszú nem-kódoló (lnc)] RNS-eket kapjuk meg az izolálás eredményeként, s ezekről tudunk a szekvenáláshoz könyvtárat készíteni.

5.5.2 Ribodepláció / a teljes RNS frakció elemzése

A teljes transzkriptomok (kis RNS-eket kivéve, melyekre nem fókuszáltunk) feltárásához hozzátartozik a nem poliadenilált RNS-ek analízise is. Ennek érdekében az összRNS mintákból ribodeplációval eltávolítottuk az (r)RNS-eket a Ribo-Zero Magnetic Kit H/M/R kittel (Epicentre/Illumina).

5.5.3 Terminátor kezelés / az RNS-ek 5' végének azonosítása

Az intakt RNS-ek szekvenálásához fel kell dúsítani azon RNS-eket, melyek cap struktúrája megtalálható az 5' végükön. Ehhez a Terminator™ 5'-Phosphate-Dependent Exonuclease (Lucigen) enzimet használtuk, amely nem emészt az 5'-cap-pel rendelkező (tehát pl. az érett, intakt) mRNS-eket, de e struktúra hiányában igen. Az RNS-ek TSS-ének azonosítása érdekében e módszer mellett számos könyvtárkészítési (cap-szelekció) és bioinformatikai (LoRTIA) módszert bevezettünk az LRS

esetén, illetve kiegészítő SRS módszereket (CAGE-Seq) is alkalmaztunk. Ezeket a megfelelő fejezetekben tárgyalom.

5.5.4 Az RNS minták tárolása

Az izolált RNS-eket a további felhasználásig minden esetben -80°C -on tároljuk.

5.6 GENOMIKAI MÓDSZEREK

5.6.1 Szekvenálás / könyvtárkészítés

Az egyes vírusok transzkriptomának minél precízebb meghatározása érdekében a szekvenálási technikák elérhető legszélesebb skáláját alkalmaztuk. A relatíve olcsó és nagy lefedettséget biztosító Illumina SRS módszert, s a mind a genom – és transzkriptom pontos felépítéséhez szükséges LRS módszereket: a drágább, de valamelyest pontosabb leolvasást biztosító PacBio-t és az olcsóbb, nagyobb hibarárával szekvenáló ONT-t is.

5.6.1.1 *Genom szekvenálás*

Genomot abban az esetben szekvenáltunk az optimális térképezés érdekében, ha az adott vírustörzs szekvenciája még nem volt elérhető az adatbázisokban.

PRV Ka törzs: a virionokból izolált DNS-ből, korábban közölteknek megfelelően^{300,301}, 6 kb-os és 20 kb-os SMRTbell könyvtárakat készítettünk a PacBio standard protokolljai szerint. A szekvenálást RSII készülékkel, öt SMRTCellen végeztük el a 'P5-C3 chemistry' (P5 DNS polimeráz és C3 reagensek) alkalmazásával.

PRV MdBio törzs: 5kb-os SMRTbell könyvtárat készítettünk^{300,301}, a PacBio 5kb-os protokollját követve. Ehhez $2\ \mu\text{g}$ genomi DNS-t g-TUBE-bal (Covaris) daraboltunk fel, átlagosan 5000 bp-os fragmensekre. A szekvenálást RSII készülékkel végeztük egy SMRTCellen, 'P6-C4' reagenseket használva.

5.6.1.2 *Rövid-read RNS szekvenálás*

Alkalmaztuk az Illumina HiScan²⁷² és MiSeq^{25,276} platformokat a PRV, a BoHV-1, az EHV-1 és az ASFV vírusok transzkriptomának elemzéséhez. A könyvtárkészítés az alábbiak szerint történt:

- a. Illumina random RNS szekvenálás
 - i. **PRV:** a könyvtárat a ScriptSeq v2 RNA-Seq Library Preparation Kittel készítettük (Epicentre), paired-end szekvenáláshoz (100 nt x 2), majd Illumina HiScanSQ platformon szekvenáltuk.
 - ii. **EHV-1:** a könyvtárkészítéshez a NEXTflex® Rapid Directional qRNA-Seq Kitet használtuk (PerkinElmer), majd MiSeq készüléken szekvenáltunk (150 nt x 2)
 - iii. **ASFV:** a könyvtárkészítéshez a NEXTflex® Rapid Directional qRNA-Seq Kitet használtuk (PerkinElmer), majd MiSeq készüléken szekvenáltunk (150 nt x 2)

- b. Illumina poliA RNS szekvenálás
 - i. **PRV:** ún. single-end könyvtárat készítettünk, horgonyzott oligo(VN)T20 primerrel és a ScriptSeq v2 RNA-Seq Library Preparation Kittel, s HiScanSQ-val szekvenáltuk.

- c. CAGE szekvenálás

A CAGE Preparation Kit-et (CAGE™ Prep Kit, DNAFORM), illetve a cég által javasolt protokollt alkalmaztuk. A CAGE könyvtárat 1 µg poliA-szelektált RNS-ből (KSHV), vagy 10 µg össz RNS-ből (többi vírus) készítettük. Az RNS-t és a random RT primert (CAGE™ Prep Kit) összekevertük, majd denaturáltuk. Az első cDNS szálat a SuperScript III enzimmel (Invitrogen) szintetizáltuk. Az RT enzim aktivitásának és specifikusságának fokozása érdekében trehalóz/szorbitol keveréket is hozzáadtunk (CAGE™ Prep Kit). A reakciót 25°C-on 30 másodpercig, majd 50°C-on 60 percig inkubáltuk. Az 5'-cap (és a 3'-vég ribóz) diol csoportjait NaIO₄-tal oxidáltuk, és Biotin (hosszú kar) hirazint kapcsoltunk hozzá. Először az oxidációt hajtottuk végre NaOAc (1M, pH 4.5, CAGE™ Prep Kit) és NaIO₄ (250mM, CAGE™ Prep Kit) a mintákhoz való hozzáadásával, és 45 percig jégen, sötétben történő inkubálással. Ezt követően a mintákhoz 40% glicerolt és Tris-HCl-t (1M, pH 8,5, CAGE™ Prep Kit) adtunk. NaOAc (1M, pH 6,0) és Biotin Hidrazint (10 mM, CAGE™ Prep Kit) kombináltuk a mintákkal, és az oxidált diol maradványokat 23°C-on 2 órán át biotinizáltuk. Ezután az egyszálú RNS-t RNase I (CAGE™ Prep Kit) kezeléssel emésztettük.

A biotinizált, capes RNS mintákat a cap-trapping során a protokollnak megfelelően²⁶ előkezelt Streptavidin gyöngyökhöz kötöttük 37°C-on 30 percig, majd mágnesre tettük. A gyöngyöket kétszer Mosópuffer 1-gyel, majd Mosópuffer 2-vel és végül Mosópuffer 3-mal mostuk (mind a CAGE™ Prep Kit-ből). Ezután a cDNS-eket eluáltuk a gyöngyökről a Releasing Buffer alkalmazásával, mely szintén a kit komponense. Ezt 95°C-on történő 5 perces inkubáció követte, majd a cap-es RNS-t tartalmazó felülúszóval dolgoztunk tovább. A tRNS-Streptavidin beadekhez RNase I puffert (CAGE™ Prep Kit) adtunk, majd ismét mágnesre helyeztük. A cDNS-t tartalmazó felülúszót jégre helyeztük, majd a kit RNase keverékével (RNase H és RNase I) inkubáltuk 37°C-on 15 percre. A potenciális maradék RNS-t RNase I-gyel emésztettük.

5.6.1.3 Hosszú-read RNS szekvenálás

- a. PacBio transzkriptom szekvenálás RSII és Sequel platformokon
 - i. nem amplifikált/ izoforma szekvenálás (IsoSeq)

A PoliA RNS frakcióból cDNS-t írtunk a SuperScript Double-Stranded cDNA Synthesis Kittel (Life Technologies); az első szál szintéziséhez azonban a kit saját enzime helyett a SuperScript III reverz transzkriptázt (Life Technologies) használtuk. Az RT reakcióhoz horgonyzott Oligo(dT)20 primert használtunk (Life Technologies). A cDNS-ből a PacBio DNA Template Prep Kittel (v.2.0), a 'Pacific Biosciences template preparation and sequencing protocol for Very Low (10 ng) Input 2 kb libraries with carrier DNA' protokollja alapján készítettünk SMRTbell szekvenáló könyvtárat. Hordozó (carrier) DNS-ként a pBR322 klónozó plazmid vektort használtuk (Thermo Scientific). A kész SMRTbell könyvtárat a PacBio DNA polymerase binding kittel (XL 1.0; part #100-150-800) kapcsoltuk a DNS-polimerázhoz és a v2 primerekhez. A polimeráz-templát komplexeket a PacBio MagBead Binding Kitjével kapcsoltuk magbeadekhez. A szekvenáláshoz P5-C3 reagenseket használtunk, 180 perces szekvenálási időt (movie length) állítottunk be.

ii. amplifikált IsoSeq

PoliA RNS mintákat használtunk a cDNS íráshoz és a könyvtárkészítéshez kétféle protokollt alkalmaztunk:

- 'PacBio: Isoform Sequencing (Iso-Seq) using the Clontech SMARTer PCR cDNA Synthesis Kit and No Size Selection (for the analysis of short transcripts)'
- 'PacBio: Isoform Sequencing (Iso-Seq) using the Clontech SMARTer PCR cDNA Synthesis Kit and Manual Agarose-gel Size Selection (analysis of long transcripts)'

Előbbit, azaz a méretszelekció nélküli protokollt a rövid (~ < 2kb), utóbbit pedig (manuális gélizolálással) a hosszabb transzkriptek analízise érdekében alkalmaztuk. A cDNS írást a SMARTer PCR cDNA Synthesis Kittel (Clontech) végeztük. Oligo(dT) primerrel (Clontech kit), illetve adapterkapcsolt GC gazdag random primerekkel is szintetizáltunk első cDNS szálát (független reakciókban). A random primereket az Integrated DNA Technologies-szel (IDT) szintetizáltattuk. Az RT-t és a PCR-t a PacBio könyvtárak esetén minden esetben a Bio Rad DNA Engine Tetrad 2 Thermal Cyclerral végeztük.

Méretszelekció nélkül: az sscDNS-t PCR-rel amplifikáltuk, (16 ciklus), KAPA HiFi enzimmel (Kapa Biosystems), majd 500 ng cDNS-ből készítettük a SMRTbell könyvtárakat a PacBio DNA Template Prep Kit 2.0-val.

Manuális méretszelekció agaróz gélből: KAPA HiFi enzimmel amplifikáltuk a cDNS-t, kétféle PCR kondíciót alkalmaztunk: 12 PCR ciklus és 1:45 perc extenzió a 2-3 kb-os transzkriptek feldúsítása érdekében, illetve 15 PCR ciklus és 3 perc extenzió a hosszabbak analíziséhez.

A SMRTbell könyvtárakat a RSII esetén PacBio DNA/Polymerase Binding Kit P6-tal (P/N 100-356-300) kapcsoltuk a DNS polimerázhoz és a szekvenáló v2 primerekhez, majd ezt a komplexet MagBeadekhez (PacBio MagBead Binding Kit). A Sequel platformon való szekvenáláshoz a Sequel Sequencing Kitet és v3 primereket használtunk. A PacBio RS II szekvenálást a DNA Sequencing Reagent 4.0 (P/N 100-356-200) használatával végeztük. SMRTCellenként (8Pac v3) egy 240, vagy 360 perces movie-t állítottunk be. A Sequel futtatáshoz a Sequel SMRT Cell 1M-t használtuk és 600 perces moviet alkalmaztunk. (movie time: Az SMRTCellből történő adatgyűjtéshez megadott idő.)

b. ONT transzkriptom szekvenálás MinION és Mk1C készülékekkel

i. amplifikált cDNS szekvenálás

Az '1D strand switching cDNA by ligation' módszerrel készítettük az amplifikált cDNS könyvtárakat az ONT Ligation Sequencing kitjével (SQK-LSK108). A poliA RNS frakcióból kiindulva horgonyzott oligo(dT) primerrel ((VN)T20) és SuperScript IV reverz transzkriptázzal írtuk az első cDNS szálát, a strand switchinget (szálváltás, visszaírás) elősegítendő, három O-metil-guanin RNS bázissal 5' adapter szekvenciát adtunk hozzá. Az amplifikációt Kapa HiFi DNS-polimerázzal (Kapa Biosystems) végeztük el, a kit primereivel. Az end-repairhez (végek javítása: az 5' túlnyúló végek feltöltése, illetve 3' túlnyúló végek eltávolítása) és az A-tailinghez (adeninek hozzáadása) a NEBNext End repair/dA-tailing Module-ját használtuk, ezt követően adaptert ligáltunk a mintánkhöz a NEB Blunt/TA Ligase Master Mixszel. A mintákat R9.4 SpotON Flow Cellekre vittük (loadoltuk). Az ONT szekvenálások esetén minden RT-hez és PCR-hez Veriti Cyclert (Applied Biosystems) használtunk.

ii. direkt cDNS (dcDNS) szekvenálás

A könyvtárakat poliA szelektált RNS mintákból, amplifikáció nélkül készítettük az ONT Direct cDNA Sequencing Kitjével (SQK-DCS109), az ONT leírása alapján. Az RNS-ből az ONT VN primereivel és a Maxima H Minus Reverse Transcriptase enzimmel (Thermo Fisher Scientific) írtuk az első cDNS szálat, a reakcióelegy az ONT Strand-Switching primerét is tartalmazta. Az RNS:cDNS hibridekről az RNS-t RNase Cocktail Enzyme Mix-szel (Thermo Fisher Scientific) emésztettük le. A második cDNS szál szintézisét LongAmp Taq Master Mix [New England Biolabs (NEB)] használatával végeztük, valamint az ONT PR2 primerével. Ezt követően az end-repairt és a dA-tailing lépést a NEBNext End repair/dA-tailing Module (NEB) reagenseket alkalmazva végeztük el. Az adaptereket a NEB Blunt/TA Ligase Master Mixszel (NEB) ligáltuk a DNS-hez. Az egyes könyvtárakat az ONT Native Barcoding (12) Kitjével barcodeoltuk (címkeztük), majd az ONT R9.4.1 SpotON Flow Celljeire vittük fel a mintákat (200 fmol/Flow Cell).

iii. direkt RNS (dRNS) szekvenálás

Mivel a dRNS szekvenálás könyvtárkészítő kitje (egyelőre) nem alkalmas a barcodeolásra, így az e technikával készített könyvtárakat nem használtuk a génexpresszió dinamikus elemzéséhez, csak a 'statikus' elemzéshez. Ehhez az adott vírus esetén a különféle időpontokból származó poliA-szelektált RNS mintáiból mixet készítettünk, majd az ONT T10 adaptert tartalmazó oligo(dT) primerével megírtuk az első cDNS szálat. A mintához az ONT RNS CS mixét is belekevertük, mellyel a szekvenálás minőségét tudjuk monitorozni, továbbá, hozzáadtuk a puffert (NEBNext Quick Ligation Reaction Buffer; NEB) és a T4 DNS ligázt (T4 DNA ligase; NEB) és szobahőn (~22°C) inkubáltuk. Ezt követően hozzáadtuk az első szál szintéziséhez szükséges puffert (5× first-strand buffer; Invitrogen) és a Dithiothreitol (DTT; Invitrogen), a dNTP-mixet (NEB) és az UltraPure nukleázmentes vizet. Végül a SuperScript III enzimet (Thermo Fisher Scientific) adtuk a mintához, az RT reakció 50°C-on 50 percig zajlott. A NEBNext Quick Ligation Reaction Bufferrel és T4 DNA ligase-zal ligáltuk az ONT RNS adapterét az RNS:cDNS hibridhez. A kész könyvtárakból 100 fmol-t vittünk fel, azaz loadoltunk minden egyes Flow Cellre.

iv. Cap-szelekció módosított könyvtárkészítéssel

A TeloPrime Full-Length cDNA Amplification Kitet (Lexogen) használtuk abból a célból, hogy feldúsítsuk a valóban full-length (teljes hosszúságú), 5' cappelt és poliA-s RNS-eket a könyvtárkészítés során. Ehhez az RT reakciókat a kit saját oligo(dT) primereivel indítottuk. (A nem poliadenilált RNS-ek azonosítása céljából független mintaként random primereket is használtunk.) Az 5' cap struktúrára specifikus adaptert ligáltunk a cDNS végére, majd PCR-ral amplifikáltuk a mintát, a kit saját Enzyme Mixét és a Second-Strand Mixet alkalmazva. A kit saját szilikamembrános oszlopos módszerével tisztítottuk a cDNS-t. A minták specifitásának ellenőrzését qPCR-ral végeztük el. A qPCR kísérletekhez minden esetben Rotor-Gene 6000 (Corbett), vagy – a későbbi projektjeink során – Rotor-Gene Q készüléket (Qiagen) és ABsolute qPCR SYBR Green Mixet használtunk (Thermo Fisher Scientific). A teszthez saját tervezésű, génspecifikus primereket alkalmaztunk, melyet az IDT-től rendeltünk. A szekvenciák az adott publikációkban elérhetők. A megfelelő specifitású cDNS-t ezt követően az ONT 'Regular' protokollja szerint¹⁵ end-repair kezelésnek és adapter ligálásnak vetettük alá, majd az ONT R9.4 SpotON Flow Cell-jein szekvenáltuk.

v. Cap-szelekció Terminátor-kezeléssel

Terminator exonukleáz (Epicentre/Lucigen) enzimmel kezeltük az RNS-t annak érdekében, hogy csökkentsük az inkomplett 5'-UTR (nem leolvasódó) szakaszokkal rendelkező transzkriptek mennyiségét. Az enzim szelektíven degradálja a nem cap-es és nem poliadenilált RNS-eket. A kezelést követően az RNS-ből ONT amplifikált cDNS könyvtárakat készítettünk.

5.6.1.4 *Szintetikus hosszú-read szekvenálás*

A LoopSeq™ Transcriptome 3x8-plex Kit-et (Loop Genomics) használtuk az SLRS könyvtár elkészítéséhez. A könyvtár egy RNS mixből készült, amely a fertőzés utáni 2 és 12 h-s mintákat tartalmazta három biológiai ismétlésben. PoliA RNS mintából indultunk ki. Az első szál szintéziséhez az RNS-hez a kit barcodeolt RT primer mixét (BC RT Mix P) és RT enzimet adtunk. Egy SPRIselect tisztítást követően (a kit tartalmazza) a mintákat felamplifikáltuk, majd ismét SPRIselect tisztítás következett. Ezt követően Distribution Mixet és Enzimet adtunk a mintához (mindkettő a kit komponense).

A cDNS mintákat 320-szorosan hígítottuk, majd a kit reagenseit alkalmazva (Amplification Mix R és Amplification Additive) felamplifikáltuk a mintát. A PCR termékeket mosást követően a Loop Genomics kit enzimeivel (Distribution Mix és Distribution Enzyme) kombináltuk, a cDNS minták barcodeok szerinti eloszlathoz. A barcode aktiválását LoopSeq Activation Mix és Activation Enzyme alkalmazása tette lehetővé. A cDNS-t Fragmentation Mixszel fragmentáltuk, majd hővel inaktiváltuk az enzimet. Ezt az end-repair és az A-tailing követte a Ligation Mix és a Ligation Enzyme alkalmazásával. A tisztított mintákat Index Master Mixszel és Index Primer P6-tal (szekvencia: GAATTCGT) amplifikáltuk. MiSeq készülékkel szekvenáltuk a mintákat (150 bp x2).

5.6.2 **Biszulfit konverzió**

Az RNS metilációt, az 5-metil citozin (5mC) pozíciókat biszulfit konverzióval határoztuk meg. Ehhez az EZ RNA Methylation Kitet (Zymo Research) használtuk, a kit leírásának megfelelően. Össz RNS-ből indultunk ki, ehhez adtuk a kit konverziós reagensét (RNA Conversion Reagent), mely a nem metilált citozinokat uracillá alakítja, az inkubáció 54°C-on történt. Ezt követően oszlopos tisztítást (Zymo-Spin IC Column) végeztünk a kit RNS-kötő pufferét alkalmazva (RNA Binding Buffer), majd végül etanolos tisztítással nyertük vissza a tiszta, biszulfit konvertált RNS-t.

5.6.3 **A könyvtárak tisztítása**

Az ONT cDNS és dcDNS, valamint az Illumina könyvtárakat minden enzimátikus lépést követően AMPure XP Beads-szel (Beckman Culter) mostuk, míg a PacBio könyvtárakat az AMPure PB Beads-szel (PacBio). A mintákat minden esetben UltraPure nukleázmentes vízben (Invitrogen) eluáltuk.

Az ONT dRNS könyvtárak készítése során az RNAClean XP Beads-t (Beckman Coulter) használtuk az enzimátikus lépéseket követően.

A LoopSeq SLRS könyvtárkészítés során minden enzimátikus lépést beades mosás követett, melyhez a LoopSeq kit saját, SPRIselect Cleanup komponensét használtuk.

5.6.4 A nukleinsavak minőségi és mennyiségi ellenőrzése

A fertőzött sejtekből izolált DNS és RNS, illetve a szintetizált könyvtárak mennyiségi ellenőrzését Qubit, Qubit 2.0 és 4.0 készülékkel (Invitrogen™) végeztük. Az alábbi Qubit kiteket használtuk:

- Qubit™ dsDNA HS Assay kit (cDNS könyvtár mennyiségi ellenőrzése)
- Qubit™ 1X dsDNA HS Assay kit (cDNS könyvtár mennyiségi ellenőrzése)
- Qubit™ RNA HS Assay kit (poliadeniláció, ribodepláció után az RNS mennyiségi ellenőrzése)
- Qubit™ RNA BR Assay kit (össz RNS mennyiségi ellenőrzése)

A minőségi ellenőrzést az Agilent Technologies 2100 Bioanalyzer készülékkel, később a TapeStation 4150-nel végeztük, az alábbi reagensekkel:

- BioAnalyzer
 - o Agilent High Sensitivity DNA Kit
 - o Agilent RNA 6000 Pico Kit
- TapeStation
 - o Genomic DNA ScreenTape
 - o DNA ScreenTape
 - o High Sensitivity DNA ScreenTape
 - o High Sensitivity RNA ScreenTape

5.6.5 A könyvtárak méretszelektálása

A különféle LRS technikák más-más okok miatt, de kedveznek a bizonyos mérettartományba eső fragmenseknek, melyek relatíve feldúsulnak a leolvasások közt. Ez egyrészt a kvantitatív értékeléseknél gond, másrészt bizonyos RNS-ek teljesen kikerülhetnek az LRS „látószögéből”. Emiatt, különféle labortechnikai módszereket alkalmaztunk, az alulbecsült RNS-ek feldúsítására.

5.6.5.1 *Manuális gélizolálás*

Az ONT amplifikált cDNS szintézisen alapuló könyvtárkészítése során, egyes PCR termékeket UltraPure Agarose (Thermo Fischer Scientific) gélen elválasztottunk, s az 500 nt-nál hosszabb fragmenseket izoláltuk vissza a Zymoclean Large Fragment DNA Recovery Kittel (Zymo Research). Így a PCR okozta rövid fragmensek feldúsulását ki tudtuk küszöbölni.

5.6.5.2 *BluePippin™ méretszelekció*

A BluePippin (Sage Science) rendszer festékmentes, előre öntött gélkazettákat használ az agarózból kivánt DNS fragmentum méret-tartományok szétválasztására és kinyerésére. Ez lehetővé teszi a fragmentum hosszak 90 bp-tól 50 kbp-ig történő méret szerinti kiválasztását. A pulzáló mezős gélelektroforézis programmal képes nagy DNS fragmentumok szétválasztására.

A cDNS-t (PacBio könyvtárkészítés közben, az end-repair és adapter ligálást követően) agaróz gélen futtattuk meg a BluePippin készülékkel, így egyrészt a hosszabb readeket ki tudtuk szelektálni, s frakcionálni tudtuk hossz alapján a cDNS-eket. Ezt követően az egyes intervallumokba eső cDNS poolokat az ennek megfelelő, optimális PacBio könyvtárkészítési módszernek vetettük alá (pl. a HSV-1 esetén 0.8-2kb-s, 2-3 kb-s, 3-5 kb-s és 5 kb+ -os könyvtárakat készítettünk és szekvenáltunk meg egymástól függetlenül).

5.7 BIOINFORMATIKA

Az elmúlt, több mint egy évtizedben, amióta szekvenálás-alapú transzkriptomikával foglalkozunk, rengeteg programot alkalmaztunk. Ennek oka egyrészt, hogy az újabb alkalmazott technológiai módszerek más adatfeldolgozást igényelnek. Másrészt, az újabb, felmerülő kérdéseknek megfelelő programokat kerestünk/írtunk, illetve, nem utolsósorban, a molekuláris biológia/genomika térhódítása részben a bioinformatika extrém gyors fejlődésének köszönhető, mi pedig nem csak a laborban, hanem az adatelemzés terén is a legmodernebb megközelítéseket választottuk. Minden újabb projektünk során a legújabb scripteket, programokat, vagy a korábbiak újabb verzióit alkalmaztuk, továbbá saját, ún. *in-house* programcsomagot és scripteket is. A következő szekcióban felsorolásszerűen tárgyalom ezeket, technikák szerint kategorizálva, az alkalmazásuknak megfelelő időrendi sorrendben.

5.7.1 Genom összeállítás és annotáció

A **PRV Ka** genom összeállítása⁴ PacBio RSII nyers readek felhasználásával történt, melyet a BLASR (<https://github.com/PacificBiosciences/blasr>) és SMRT Analysis (<https://github.com/PacificBiosciences/SMRT-Analysis/wiki/SMRT-Pipe-Reference-Guide-v2.0>) programokkal illesztettük a referencia genomra. A protein kódoló gének predikciójához a GATU³⁰² (<https://4virology.net/virology-ca-tools/gatu/>) programot alkalmaztuk.

A **PRV MdBio** genomot a Ka törzs alapján állítottuk össze¹³¹.

A **PRV uI54Δ** genom összeállítása szintén a Ka törzs alapján történt²⁸⁰.

A **PCV-1 Szeged izolátum** genom összeállításához¹⁷⁸ referencia genomot használtunk (GenBank azonosító: NC_001792.2), a PacBio RSII readek térképezését a SMRT Analysis és a GMAP³⁰³ programokkal végeztük.

A **PERV Szeged törzs**²⁹⁴ A PacBio RSII readeket a SMRT Analysis és GMAP³⁰³ programokkal illesztettük a referenciagenomra (GenBank azonosító: AY099323). A törzs sejtvonálhoz alkalmazkodását is analizáltuk, ehhez összehasonlításképp egy sertésből izolált PERV törzs genomját is felhasználtuk (GenBank azonosító: EU789636).

A **VACV WR törzs**²⁹⁵ A genomot transzkriptekből állítottuk össze, PacBio RSII és ONT MinION szekvenálást alkalmaztunk ehhez. A PacBio nyers readeket BLASR hosszú-read térképező programmal mappeltük a referencia genomhoz (GenBank azonosító: NC_006998). A megfelelő illesztés érdekében a GMAP programot is alkalmaztuk. A MinION leolvasásokat az Albacore szoftverrel térképeztük a referencia genomra. Az ORF-eket a Geneious programmal³⁰⁴ prediktáltuk.

A **SARS-CoV-2** genomot²⁴ a Shasta programmal (<https://github.com/paoloshasta/shasta>)³⁰⁵ állítottuk össze a 20000 bp-nál hosszabb readeket felhasználva. A filogenetikai analízist és a mutációk elemzését a Nextstrain³⁰⁶ (<https://nextstrain.org/>) programmal végeztük el a VirStrain³⁰⁷ eredményeként kapott genomon (<https://github.com/liaohherui/VirStrain>). A kihívást az jelentette, hogy megkülönböztessük a virális gRNS-ből származó RNS-eket az sgRNS-ektől. Ehhez a readeket, melyeket eredendően az eredeti vuhani genomhoz (NC_145512.2) mappeltünk, újra elemeztük saját scriptekkel³⁰⁸, melyhez a tidyverse³⁰⁹, az RSamtools³¹⁰, a GenomicAlignments³¹¹, a tidygenomics³¹² és a dplyr³¹³ programcsomagokat is használtuk.

5.7.2 A transzkriptom analízishez használt legfontosabb programok

Ebben a fejezetben szekvenálási technikák szerint lebontva, felsorolásszerűen ismertetem a virális transzkriptomok elemzéséhez alkalmazott bioinformatikai programcsomagokat. Az egyes programok verziószámait nem jelöltem, mindig az aktuálisan elérhető legfrisebb verziókkal dolgozunk. Az erre vonatkozó adatok a publikációinkban megtekinthetők.

5.7.2.1 *PacBio RSII és Sequel transzkriptom szekvenálás*

- a. **PacBio Iso-Seq** protokoll **SMRT Analysis**³¹⁴ az alap elemzésekhez, térképezéshez
- b. **Biopython**³¹⁵ **Bio.motifs** programcsomag a potenciális polimeráz II kötőhelyek és poliA (PA) szignálok azonosításához - <https://biopython.org/docs/1.76/api/Bio.motifs.html>
- c. **JASPAR PolII** adatbázis³¹⁶ <https://jaspar2020.genereg.net/> polimeráz II kötőhelyek azonosításához
- d. **PolyApred**³¹⁷ <https://bio.tools/polyapred> - PA szignálok predikciójához
- e. **GPMiner**³¹⁸ <https://bio.tools/gpminer> - TATA box szekvenciák predikciójához
- f. **SMRT Link** <https://www.pacb.com/support/software-downloads/> - Sequel CCS readek generálásához

5.7.2.2 *ONT transzkriptom szekvenálás*

- a. **Albacore** (ONT) <https://nanopyes.readthedocs.io/en/latest/index.html#parallel-basecalling-with-ont-s-albacore> - a leolvasások bázishívásaihoz
- b. **Guppy** (ONT) https://timkahlke.github.io/LongRead_tutorials/BS_G.html - a leolvasások bázishívásaihoz
- c. **GMAP**³⁰³ <https://manpages.ubuntu.com/manpages/trusty/man1/gmap.1.html> - a térképezéshez
- d. **minimap2**³¹⁹ <https://github.com/lh3/minimap2> - a térképezéshez
- e. **ReadStatistics** Python script <https://github.com/moldovannorbert/seqtools> - a statisztikai elemzésekhez, poliA szignálok azonosításához
- f. **LoRTIA** <https://github.com/zsolt-balazs/LoRTIA> - a TSS-ek, TES-ek és intronok, illetve maguk a transzkriptek megállapításához
- g. **metablastr**³²⁰ <https://github.com/drostlab/metablastr> - Annotációk genomra illesztéséhez
- h. **Liftoff**³²¹ <https://github.com/agshumate/Liftoff> - Annotációk genomra illesztéséhez
- i. **Samtools**³²² <http://www.htslib.org/> fájlok átalakításához (sam -> bam)
- j. **R** programcsomagok a transzkriptek összehasonlításához, a kódoló potenciáljuk elemzéséhez, Kozak szekvenciák kalkulálásához, ORF predikcióhoz, BLAST alapú összehasonlításokhoz és vizualizációhoz

- ORFik³²³ <https://bioconductor.org/packages/release/bioc/html/ORFik.html>
- Gviz³²⁴ <https://bioconductor.org/packages/release/bioc/html/Gviz.html>

5.7.2.3 *Illumina RNA-Seq*

- FastQC** <https://github.com/s-andrews/FastQC> - readek minőségének ellenőrzéséhez
- Tophat**^{325,326} <https://ccb.jhu.edu/software/tophat/index.shtml> térképezéshez
- Cutadapt**³²⁷ <https://cutadapt.readthedocs.io/en/stable/> - trimmeléshez
- Bowtie2**³²⁸ <https://github.com/BenLangmead/bowtie2> - PA-Seq adatok térképezéséhez
- HOMER**³²⁹ (Heinz 2010) <http://homer.ucsd.edu/homer/ngs/analyzeRNA.html> - szálspecifikusság azonosításához és PA csúcsok detekciójához
- Artemis Genome Browser**³³⁰ <https://www.sanger.ac.uk/tool/artemis/> - annotációhoz
- Integrativ Genomics Viewer (IGV)**³³¹ <https://igv.org/> - annotációhoz és vizualizációhoz

5.7.2.4 *Illumina CAGE-Seq*

- fastqc** <https://www.bioinformatics.babraham.ac.uk/projects/fastqc>; readek minőségi ellenőrzéséhez
- TrimGalore** <https://github.com/FelixKrueger/TrimGalore>; leolvasások trimmeléséhez
- STAR**³³² a leolvasások genomhoz való térképezésére
- CAGEfightR**³³³ <https://bioconductor.org/packages/release/bioc/html/CAGEfightR.html> - TSS-ek azonosításához

5.7.2.5 *Illumina LoopSeq*

- Loop Genomics Pipeline Software** <https://analysis.loopgenomics.com/accounts/login/> – a rövid Illumina leolvasások hosszú-readekké való összeillesztéséhez
- minimap2** a térképezéshez

5.7.2.6 *Vizualizáció*

- SnapGene Viewer** <https://www.snapgene.com/snapgene-viewer>
- Artemis**^{330,334} <https://www.sanger.ac.uk/tool/artemis/>
- IGV**³³¹ <https://www.igv.org/>
- Geneious** <https://www.geneious.com/>
- ggplot tidyverse**³⁰⁹ <https://ggplot2.tidyverse.org/>
- Bioconductor**³²⁴ <https://www.bioconductor.org/>
- WebLogo**³³⁵ <https://weblogo.threeplusone.com/>
- ComplexHeatmap**³³⁶ <https://bioconductor.org/packages/release/bioc/html/ComplexHeatmap.html>

i. R – saját scriptek

- <https://github.com/Balays/Rlyeh>.
- workflow: https://github.com/Balays/KSHV_RNASeq

5.7.2.7 A transzkript izoformák funkcionális elemzése

- a. **Coding-Potential Assessment Tool**³³⁷ (CPAT) – <https://bio.tools/cpat> a kódoló potenciál kiértékeléséhez
- b. **isoformswitchanalyser**³³⁸
<https://bioconductor.org/packages/release/bioc/html/IsoformSwitchAnalyzeR.html> - az izoformák doménjei időbeli változásának analíziséhez
- c. **SignalP**³³⁹ szignálpeptidek jelenlétének ellenőrzéséhez
- d. **IUPred2A**³⁴⁰ – IDR -ek (intrinsically disordered proteins/protein regions, rendezetlen fehérjék/fehérjeszakaszok) ellenőrzéséhez
- e. **Coding Capacity Assessment Tool** - <http://lilab.research.bcm.edu/index.php> – a kódoló potenciál kiértékeléséhez

5.7.3 Statikus transzkriptom analízis

A vírusok RNS profilalkotása mind a labortechnikai megközelítések, mind a bioinformatika területén rengeteget változott, fejlődött az utóbbi évtizedben, részben eredményeinknek köszönhetően. Ebben a szekcióban bemutatom a fő bioinformatikai munkafolyamatot, amelyet kisebb változtatásokkal a legtöbb esetben (pl. a HSV-1 és az AcMNPV transzkriptom annotálása során) alkalmaztunk az RNS-ek strukturális variánsainak feltérképezéséhez.

5.7.3.1 Térképezés és adatelemzés

A leolvasások referenciagenomhoz (**2. táblázat**) való térképezését a minimap2 programmal végeztük az alábbi beállítással: `-ax splice -Y -C5 -cs`. Ezt követően a LoRTIA programmal határoztuk meg az intronokat, illetve a transzkriptek TSS és TES pozícióit, meghatároztuk a teljes hosszúságú readeket. A feltételezett intronokat úgy definiáltuk, hogy az alábbi konszenzus szekvenciáknak kellett határolni: GT/AG, GC/AG, AT/AC. Emellett, csak akkor fogadtuk el a splice helyeket, ha azokat a dRNS-Seq technika is igazolta.

A leolvasásokon az 5'-végi adapterek és a poliA farok szekvenciáit szintén a LoRTIA programmal azonosítottuk, a Smith-Waterman illesztési pontszámok alapján. Ha az adapter vagy a poliA szekvencia az illesztéstől legalább három nukleotiddal (nt) downstream végződött, az adaptert elvetettük, mivel template-switching révén kerülhetett oda.

Térképezéshez használt genomok GenBank-i azonosítói	
Vírusok és sejtvonalak	Azonosító
AcMNPV	KM667940.1
ASFV	MN715134.1
BoHV-1	JX898220.1
MDBK (BoHV-1 sejt)	GCF_002263795.1
CCHFV	NC_005300.2
CVB5	AF114383
CV-1 (VACV, MPXV sejt)	GCF_015252025.1

EBV	KC207813.1
EHV-1	NC_001491.2
HCMV	LT907985.2
HSV-1	X14112.1
IVA	CF_001343785.1
KSHV	GQ994935.1
MDCK (IVA sejt)	GCA_000002285.4
MPXV	ON563414.3
OK (BoHV-1 sejt)	GCA_002742125.1
PERV	KY484771
PRV	KJ717942.1
PCV	KX816645
RK (EHV-1 sejt)	GCF_000003625.3
SARS-CoV-2	MT560672.1
Sf9 (AcMNPV sejt)	PRJNA380964
T98G (VSIV sejt)	GCA_000001405.28
VACV	LT966077.1
Vero (HSV-1, ZIKV, WNV, CCHFV, CVB5, VSIV, SARS-CoV-2 sejt)	GCA_000409795.2
VSIV	NC_001560.1
VZV	NC_001560.1
WNV	NC_001563.2
ZIKV	NC_012532.1

2. táblázat A readek mappeléséhez használt referenciagenomok GenBank-i azonosítói.

A transzkript jellemzőit, mint az intronok, a TSS-ek és a TES-ek, csak akkor annotáltuk, ha legalább két leolvasásban (ezt az értéket később emeltük) és a régiós lefedettség 0,1%-ában kimutathatók voltak. Az RNS-degradáció hatásainak csökkentése érdekében csak azokat a TSS-eket annotáltuk, amelyek jelentős csúcsot adtak egy ± 50 nt hosszú ablakaikban, a Poisson-eloszlás szerint (Bonferroni korrekcióval). A kis abundanciájú readeket, amelyeket egyetlen kísérletben mutattunk ki, akkor fogadtuk el transzkriptként, ha más transzkriptek is használták ugyanazokat a TSS-eket és TES-eket. A LoRTIA *Transcript Annotator* modulja segítségével annotáltuk a TI-kat. A hosszú, kis abundanciájú readek 5'-végeit egyenként ellenőriztük az IGV segítségével.

Az RT és a PCR során keletkező potenciális fals termékekből származó readeket kizártuk az elemzésből. Ezek a template switchingből, illetve az oligo(dT) és a PCR primerek nem specifikus bekötődéséből származhatnak. Az oligo(dT) primerek az A-gazdag genomi/transzkript régiókba is beköthetnek, ezáltal fals readeket generálhatnak.

Egy lehetséges TSS-t vagy TES-t valószínűleg fogadtunk el, ha a LoRTIA a hosszabb izoformák esetében legalább három, a rövidebb változatok esetében legalább öt független mintában, azonosította. A rövidebb izoformák szigorúbb választási kritériumának oka, hogy ezek az RNS fragmentáció, töredezettség eredményei is lehetnek. Csak azokat a readeket fogadtuk el transzkriptként, amelyek tartalmaztak egy fentebb említett módon annotált TSS-t és TES-t. A nagyon hosszú transzkriptek ($\geq 3,000$ bp) esetében egy másik szabályt alkalmaztunk: egy leolvasást transzkriptként fogadtunk el, ha hosszabb volt minden annotált átfedő RNS-nél, még akkor is, ha nem volt hozzárendelhető, annotált TSS-se és alacsony volt a lefedettsége.

5.7.4 Dinamikus transzkriptom analízis

A transzkriptek kvantitálását és a vírusfertőzés során bekövetkező mennyiségi változásait az LRS módszerek előtt precízen nem lehetett mérni, hiszen más technikák nem teszik lehetővé az egymással átfedő transzkriptek elkülönítését.

5.7.4.1 A Pseudorabies vírus (PRV) gének kinetikai besorolása PacBio RSII-vel

Az első, LRS alapú vírus génextpressziós tanulmány az általunk a PacBio RSII módszerrel végzett, a PRV-re vonatkozó elemzés volt⁷. A mappelt readokat TI-khoz rendeltük hozzá, ha a leolvasás koordinátái nem haladták meg az izoformák 5' vagy 3' végét több mint 7 nukleotiddal. A readok irányát a poliA-szekvencia alapján határoztuk meg. Ha a read több mint egy izoformához illeszthető volt, akkor a legrövidebbhez rendeltük. A readszámot normalizáltuk a mitokondriális (mt)DNS-re térképeződő leolvasásokhoz (**A** értékek), valamint a PRV genom relatív kópiaszámához is (**An** értékek). Minden egyes TI gyakoriságát (**F** érték) úgy számítottuk ki, hogy az adott transzkript readszámát elosztottuk a mintában lévő össz vírusread számával. A relatív expressziós értékeket pedig úgy számítottuk ki, hogy az egyes transzkriptek F értékeit normalizáltuk az adott transzkript átlagos F értékeihez (**Fx** értékek).

Az F , F_x , A , An , A_6 és An_6 értékeket az alábbi egyenletek szerint számítottuk ki:

$$F_{tr1t1} = \frac{R_{c_{tr1t1}}}{\sum_{i=tr1}^{tr147} R_{c_{it1}}},$$

$$F_x_{tr1t1} = \frac{F_{tr1t1}}{\sum_{i=t1}^{t6} F_{tr1t1}},$$

$$A_{tr1t1} = \frac{R_{c_{tr1t1}}}{R_{c_{mit1}}},$$

$$An_{tr1t1} = \frac{A_{tr1t1}}{R_{DNA_{t1}}},$$

$$A_6_{tr1t1} = \frac{A_{tr1t1}}{A_{tr1_6h}},$$

$$An_6_{tr1t1} = \frac{An_{tr1t1}}{An_{tr1_6h}}$$

R_c a transzkripthez (tr) rendelt readok száma, t egy adott időpont a fertőzést követően, R_{cmit} a mitokondriális RNS-ekre eső readok számát jelöli. R_{DNA} a mért DNS relatív mennyiségét jelenti.

A PRV TI-kat három klaszterbe csoportosítottuk a relatív expressziós értékei szerint k-means klaszteranalízissel, Euklideszi távolsággal, a Gene Cluster 3.0 program használatával³⁴¹. Azon transzkriptek csoportját, amelyek a fertőzés utáni első órákban magas relatív expressziós értékekkel rendelkeztek, korai (E) klaszterbe soroltuk, a későbbi fertőzési szakaszokban magas relatív expressziós értékekkel rendelkezőket kései (L2) klaszterbe, míg a klasztereken kívüli transzkripteket szivárgó-késői (L1) kategóriába soroltuk.

A transzkriptek A és An értékeinek számításához a PRV readokat a mitokondriális readszámmal normalizáltuk. Ehhez a readokat a *Sus scrofa* 10.2 MT kromoszóma szekvenciájához térképeztük. Az alfaherpeszvírusok fertőzése során bekövetkező RNS degradáció számos problémát felvetett a kvantitatív analízis során, problémát jelentett a megfelelő referencia RNS kiválasztása. A qRT-PCR projektjeinkhez a 28S rRNS-t használtuk e célból, ez azonban a PoliA- szelekción (is) alapuló LRS módszerek esetén nem alkalmazható. Egy tanulmány a HSV-1 fertőzte Vero sejtek esetén

kimutatta, hogy az mtDNS mennyisége csökken a vírus ul12.5 génje következtében³⁴², ez a gén azonban a PRV-ből hiányzik.

5.7.4.2 *A Vakcinia vírus (VACV) gének relatív expressziós szintjeinek meghatározása PacBio Sequel és ONT MinION szekvenálás alapján*

A VACV transzkriptom dinamikus profilozásához kombináltan alkalmaztuk a PacBio és az ONT módszereit. A readeket a LoRTIA programcsomaggal elemeztük, s csak a program szűrőin átjutott transzkriptekkel dolgoztunk a kalkulációk során. A további szigorításunk az volt, hogy kizárólag a legalább öt független, teljes hosszúságú readdel reprezentált transzkripteket használtuk az analízishez. A posztreplikatív gének a kései fertőzési stádiumokban alulreprezentáltak, extrém magas TSS számuk miatt¹⁷. Az adott időpontbeli normalizált relatív expressziós értéket (**Rt**) az alábbi képlettel számoltuk:

$$Rt = \left(\frac{c}{\sum_{i=1}^n c} \right)^{-h}$$

, ahol n = az adott mintában az összes TI-k száma, C= az adott időpontban az adott transzkript leolvasásainak száma, és a $h = \sum_{i=1}^k C$ pedig az adott transzkript C értékei átlagát adja a k= 1, 2, 4, 6 és 8 óra tekintetében. Az Rt értékeket az Instant Clue statisztikai programmal ([GitHub - hnolCol/instantclue: Instant Clue - Interactive Data Analysis](https://github.com/hnolCol/instantclue)) hierarchikus klaszterezettük, így meghatároztuk a transzkript izoforma klaszterek számát. Az euklideszi metrikát használtuk a sorok csoportosítására az adatkeretben, majd ezt követően k-means hierarchikus klaszterezést végeztünk. A klaszterek átlagait a BioVinci programmal ábráztuk. A VACV proteinek ismert transzkript izoformáinak funkcionális analíziséhez az UniProt adatbázis leírásait használtuk (<https://www.uniprot.org/>).

A VACV ORF-ek dinamikus expressziós elemzéséhez az **Rt** értékeket az alábbi módosításokkal kalkuláltuk: minden olyan transzkript **Rt** értékét, mely az adott ORF-et teljesen átírta, összeadtuk, kivéve azokat a poligén transzkripteket, amelyek az adott ORF-et downstream pozícióban tartalmazták (utóbbiak ugyanis nem olvasódnak le).

5.7.4.3 *A Bovine alphaherpesvirus-1 (BoHV-1) gének expressziós analízise*

A virális génexpresszió kinetikai elemzéséhez az ONT dcDNS-Seq adatainkat használtuk, s csak azokat a readeket, melyek a LoRTIA program beállításainak megfeleltek, s transzkriptként annotálta őket a program. Ezt követően kizártuk azokat a transzkripteket, melyeket kevesebb mint négy read erősített meg, majd a normalizációt a DESeq2³⁴³ szoftverrel végeztük el, a 'Median Ration Normalization' módosított verzióját használtuk. A biológiai párhuzamosok átlag normalizált leolvasásai számait tekintettük az adott TI readszámának.

5.7.4.4 *A herpesz szimplex vírus-1 (HSV-1) gének expressziós analízise*

A transzkripteket a LoRTIA programmal azonosítottuk, elsőként a TSS-ek, TES-ek és intronok analízisével és validálásával. Az intronokat csak akkor fogadtuk el, ha a Depledge és munkatársai által közölt³⁴⁴ dRNS-Seq adatszettet felhasználva is visszaigazoltuk. A dinamikus elemzést a LoRTIA által megadott, adott transzkriptre eső readszámokat felhasználva végeztük. A TSS és TES

dinamikát az IGV-vel, míg az egyes transzkripteket dinamikáját a Geneious programmal ábrázoltuk.

5.7.4.5 *A Vesicular stomatitis Indiana virus (VSIV) sejtspecifikus, dinamikus transzkriptom elemzése*

A vírus gének expressziós értékeit a Salmon program³⁴⁵ segítségével becsültük meg, és az eredményül kapott értékeket az ImpulseDE-vel³⁴⁶ értékeltük ki [amely a DESeq2) normalizálási módszerét alkalmazza], hogy elemezzük minden gén kifejeződése mértékét az idő függvényében (a minták közötti génextpressziós különbségeket).

5.7.5 RNS módosulások analízise

A dRNS szekvenálás során adatokat kapunk az RNS-ek különféle módosulásairól is, amik detektálásához a nyert fast5 readeket és a Tombo (<https://github.com/nanoporetech/tombo>) szoftvert használtuk.

5.7.6 Vírusfertőzés a gazdagének expressziójára gyakorolt hatásának analízise

A readek gazdagenomra térképezéséhez a minimap2 programot használtuk.

5.7.6.1 *BoHV-1 – B. taurus sejtvonal*

Kizártuk a mapping quality (MAPQ), azaz a térképezés minőségére vonatkozó MAPQ = 0 értékeket. Csak azokkal a readekkel számoltunk, amelyek megfeleltek a referenciagének exon szerkezetének (± 5 bp ablakkal).

Differenciális expressziót (DE) kalkuláltunk a mock sejtekhez képest, 0,01-es hamis felfedezési arány (FDR) küszöböt alkalmazva. A p-értékeket a Benjamini & Hochberg eljárással igazítottuk. A DE gének normalizált readszámai mediánját az edgeR-ből exportáltuk. A génextpressziót értékeket normalizáltuk a maximum expressziós értékekhez, majd klaszter analízissel hasonlítottuk egymáshoz. Az alábbi programokat alkalmaztuk:

- a) EdgeR-t³⁴⁷ (DGEList funkció) - a biológiai párhuzamosok csoportosításához
- b) R (<https://www.r-project.org/>), amap_0.8-16 csomag, Kmeans funkció, Euklideszi távolság számítás – klaszterezéshez. A Calinski-kritériumok alapján az optimális klaszterszám a hat volt.
- c) ggplot2, geom_smooth function – a klaszterek vizualizálásához
- d) PANTHER³⁴⁸ – overreprezentációs analízishez
- e) Gene Ontology (GO) biológiai folyamatok és GO molekuláris funkciók annotációs adatkészleteit használtuk az overreprezentációs elemzéshez
- f) WebLogo3 (<http://weblogo.threeplusone.com/>) – TSS és TES szekvenciák elemzéséhez/ábrázolásához
- g) Inkscape (<https://inkscape.org/>) – ábrák készítéséhez
- h) IGV – ábrák készítéséhez

5.7.6.2 VACV - *C. saebaues* sejtvonala

Kizártuk a mapping quality (MAPQ), azaz a térképezés minőségére vonatkozó MAPQ = 0 értékeket, csak azokkal a readokkal számoltunk, amelyek megfeleltek a referenciagének exon szerkezetének (± 5 bp ablakkal). A génekre eső readszámokat normalizáltuk a gazda genomhoz térképezett összreadokkal, hogy azonosítani lehessen azokat az abundáns háztartási géneket, amelyeket nem befolyásolt a vírusfertőzés. A top 16 háztartási gén geometriai átlagát számoltuk ki. A normalizáláshoz egy $< 0,2$ -es variációs együtthatót használtunk. A statisztikai elemzés egy kritériuma volt, hogy a normalizált összes readcount > 60 legyen a hat időpontban. A gének expressziós profiljainak elemzéséhez a normalizált értékekből relatív expressziót számoltunk, ahol a legmagasabb expressziós időpont értéke 1,0 (100%) volt.

Az alábbi programokat használtuk:

- R csomag, k-means algoritmus (<https://cran.r-project.org/bin/windows/base/old/3.5.0/>) – a klaszterezéshez
- R csomag, cascade KM algoritmus a Calinski-Harabasz kritériumok³⁴⁹ alkalmazásához, amellyel az optimális klaszterszámot adtuk meg
- R csomag, heatmap – a klaszterek ábrázolásához
- PANTHER³⁴⁸ – overrepresentációs analízishez
- Gene Ontology (GO) biológiai folyamatok és GO molekuláris funkciók annotációs adatkészleteit használtuk az overrepresentációs elemzéshez
- ggplot2 tidyverse³⁰⁹ <https://ggplot2.tidyverse.org/reference/ggplot.html> - ábrák készítéséhez
- ComplexHeatmap (<https://jokergoo.github.io/ComplexHeatmap-reference/book/>) - ábrák készítéséhez
- Inkscape - ábrák készítéséhez

5.8 KVANTITATÍV REAL-TIME PCR

Kvantitatív RT-PCR technikát (qRT-PCR) alkalmaztunk a mutáns vírusok esetében a mutációk globális expresszióra gyakorolt hatásának elemzéséhez, valamint az LRS technikával azonosított új transzkriptek validálására és az új RNS-ek kinetikai kategóriáinak megerősítésére. Ezenkívül a technikával is mértük a gazdagének vírusfertőzés által okozott expresszióbeli változásait.

Az RT reakciót 70 ng össz RNS mintából, Superscript III enzimmel (Life Technologies) és génspecifikus primerekkel végeztük. A primereket az IDT-től rendeltük, a szekvenciák az adott publikációkban megtalálhatók. A qRT-PCR reakciókhoz Absolute QPCR SYBR Green Mix-et (Thermo Scientific) használtunk, míg 28S riboszómális (r)RNS-t használtuk referencia génként.

Az amplifikációs görbét a qPCR készülék programjával Rotor-Gene Q software version 2.3.5 (Qiagen) (<https://www.qiagen.com/ie/resources/resourcedetail?id=9d8bda8e-1fd7-4519-a1ff-b60bba526b57&lang=en>) készítettük. A program Comparative Quantitation modulja kalkulálta az amplifikáció hatékonyságot (Efficiency, E) is, valamint a küszöbfluoreszcencia (Cycling threshold, Ct) értékeket is ez a szoftver állította be.

A relatív expressziós arányokat (R) egy korábbi tanulmányunknak megfelelően², a következő képlettel számítottuk ki (a párhuzamosok átlagára vonatkozóan):

$$R = \frac{(E_{sample\ 6h})^{Ct_{sample\ 6h}}}{(E_{sample})^{Ct_{sample}}} : \frac{(E_{ref\ 6h})^{Ct_{ref\ 6h}}}{(E_{ref})^{Ct_{ref}}}$$

A '*sample*' (minta) a vizsgált vírus RNS-re, a '*ref*' pedig a 28S rRNS-re (belső kontroll) vonatkozik. A kinetikai kategorizálásnál a fertőzést követő 6 h-s mintát vettük a legtöbb esetben maximum értéként egy korábbi publikációnk alapján².

Az mRNS szint 'nettó növekedését' (R_{Δ}) és a 'változás mértékét' (R_a) az alábbiak szerint számoltuk:

$$R_{\Delta} = R_{(t+1)} - R_t$$

$$R_a = R_{(t+1)} / R_t$$

A kiütött gén által kódolt fehérje a vírusgének expressziójára gyakorolt hatását az R_r érték számításával határoztuk meg, mely a mutáns és a *wt* vírus R értékei közti arányt jelenti:

$$R_r = R_{gén\Delta} / R_{wt}$$

ahol az $R_{gén\Delta}$ és az R_{wt} az adott gén, adott időpontjában mért R értéke a mutáns vírusban és a *wt*-ban. Az E, E/L és L gének átlagos R_r értékeit is kalkuláltuk:

$$\bar{R}_r = \bar{R}_{gén\Delta} / \bar{R}_{wt}$$

Pearson-féle korrelációt alkalmaztunk a génexpresszió dinamikája vizsgálatához²:

$$r = \frac{\sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y})}{(n-1)S_x S_y}$$

Továbbá, ún. R_x értékkel is számoltunk, ahol normalizációs faktorként az adott mintában az össz vizsgált transzkript (*other*) átlagos expressziós értékét vettük. A '*sample*' pedig az adott transzkriptre utaló értékeket jelenti. A módszer egy korábban leírt metódust vesz alapul³⁵⁰, azonban a Ct értékek mellett mi a hatékonyságot is figyelembe vettük.

$$R_x = \frac{E_{sample}^{Ct_{sample}}}{E_{other}^{Ct_{other}}}$$

5.9 NORTHERN BLOT

5.9.1 Hagományos Northern blot

A Northern blot kísérleteket Ausubel és mtsai.³⁵¹ által leírtak szerint hajtottuk végre a következő módon: vírusr fertőzött és nem fertőzött sejtekből izolált RNS-t 1%-os formaldehid-agaróz gélben választottuk el, majd kapilláris blotting módszerrel vittük át pozitívan töltött nylon membránra, UV keresztkapcsolással. A próbákat amplifikáltuk, majd radioaktív jelöléssel láttuk el. A teljes éjszakán át tartó [overnight (ON)] hibridizációt 42°C-on végeztük hibridizációs oldatban. A hibridizáció után a filtereket mostuk.

5.9.2 Mikro RNS (miRNS) Northern blot

Két különböző PCR próbát használtunk. A forward primert T7 promóter szekvenciával, a reverz primert pedig SP6 promóter szekvenciával kapcsoltuk össze. A mintákat denaturáló (8M ureát tartalmazó) 12%-os poliakrilamid gélekben fracionáltuk, majd kapilláris módszerrel membránra vittük át és UV keresztkapcsolással rögzítettük. A hibridizációt 37 °C-on végeztük. Az [α -32P]UTP-vel jelölt RNS próbát használtunk a hibridizációhoz. A membránokat kétszer 10 percig mostuk.

A Northern blot kísérletekhez használt reagensek		
Alkalmazott anyagok	Pontos összetétel, ill. alkalmazott kit (cég)	
	Tradicionalis	miRNS
Nylon membrán	<ul style="list-style-type: none"> Hybond-N (Amersham) Nytran N (Schleicher & Schuell BioScience) 	Nytran N (Schleicher & Schuell BioScience)
Radioaktív jelölő kit	DecaLabel DNA Labeling Kit (Thermo Scientific)	
Radioaktív jel	α -32P dCTP	α -32P]UTP
PCR	AccuPrime GC-Rich DNS polimeráz (Life Technologies)	
ON hibridizációs oldat	<ul style="list-style-type: none"> 0,1% SDS, 5x Denhardt-oldat, 30% dezionizált formamid, 5 mM EDTA, 0,9 M NaCl, 50 mM Na₂HPO₄, 100 μg/ml fragmentált hering sperma DNS, α-32P jelölt próba 0,5% SDS, 5x SSPE, 5x Denhardt-oldat és 20 μg/ml fragmentált denaturált lazac sperma DNS 	<ul style="list-style-type: none"> 50% formamid, 0,5% SDS, 5x SSPE, 5x Denhardt-oldat és 20 μg/ml fragmentált, denaturált lazac sperma DNS
Oldat a filter mosáshoz	<ul style="list-style-type: none"> 0,2xSSC, 0,1% SDS, 65°C (3x) SSC 0,1% SDS 68 °C (10 perc), 0,5x SSC, 0,1% SDS (10 perc), 0,1x SSC 0,1% SDS (10 percig) 	<ul style="list-style-type: none"> 2x SSC, 0,1% SDS (10 perc, 2x)
Elemzés	Typhoon™ FLA 9000 képalkotó és ImageQuant 5.0 szoftver	

3. táblázat A Northern blot kísérletek során használt anyagok. A potenciálisan kódoló (mRNS), a hosszú-nem kódoló (lnc) RNS és a miRNS kimutatásra eltérő kiteket és anyagösszetételt alkalmaztunk, illetve évek során is változtattunk ezen, így több verziót is feltüntettem.

5.10 MIKROSKÓPOS TECHNIKÁK

A vírusfertőzéshez használt sejtek állapotának ellenőrzéséhez Telaval (Carl Zeiss Jena) inverz fénymikroszkópot használtunk.

A zöld fluoreszcens proteint (GFP) expresszáló mutáns PRV-k szelekciójához az Olympus IX71 inverz fluoreszcens mikroszkópot használtuk.

5.11 GÉNTÉCHNOLÓGIAI MEGKÖZELÍTÉSEK

5.11.1 Mutáns vírusok előállítása

Különféle, a vírus transzkripcióját szabályozó (*ep0*, *us1*), vagy egyéb kulcsfontosságú szereppel bíró (többek között az *ul54*, *ul41*, *ul23*, *us7*, *us8*) géneket homológ rekombinációval kiütöttünk annak érdekében, hogy e gének a virális életciklusra, az összgenom-szintű génexpresszióra gyakorolt hatásáról, funkciójukról képet kapjunk. Ezekkel a knock-out (KO) vírusokkal a vad típusú (*wt*) törzshöz hasonlóan időzítéses fertőzést végeztünk, majd a génexpresszió változásait vizsgáltuk egyrészt a vírushoz fertőzés egyes időpontjaiban, másrészt a *wt* vírushoz viszonyítva. Korábbi projektjeink során a qRT-PCR-os megközelítést alkalmaztuk, e disszertációban az *ul41*²⁸¹, az *ep0*²⁷⁹ és az *ul54*²⁸⁰ gendeletált vírusokkal kapcsolatos eredményeinket foglalom össze.

5.11.1.1 PRV *ul41Δ*

A *XhoI* DNS fragmens az egész *ul41* gént tartalmazza. A gén a VHS fehérjét kódolja. Ezt a fragmens a pRL494 vektor *SalI* helyére építettük be. E DNS szegmens egyedi *NruI* hasító helyét *EcoRI* linkerrel helyettesítettük, mely így egy kereteltolódásos (frameshift) mutációt eredményezett az *ul41* génben (pUL41-RI konstrukció). Következően egy *EcoRI* linkerekkel körülvett lacZ expressziós kazettát (mely a HCMV IE1 promóterét, lacZ gént és SV40 poliA szignált, valamint terminációs szekvenciákat tartalmaz) ültettünk a pUL41-RI *EcoRI* helyére, s ezt a targeting vektort használtuk a transzfekció során a tisztított PRV Ka DNS-sel együtt. A mutáns vírusokat a kék plakk fenotípus alapján szelektáltuk, számos plakkisztítási lépés során. Ezt követően a lacZ expressziós kazettát kivettük a mutáns PRV-ből a Boldogkői és munkatársai által leírtaknak megfelelően³⁵². A lacZ-tartalmú vírus DNS-t *EcoRI*-gyel hasítottuk, majd ligáltuk és kotranszfektáltuk. A lacZ gén kazetta hiányos mutáns vírusokat fehér plakk fenotípus alapján szelektáltuk. A mutációt az *EcoRI* fragmenst tartalmazó PCR fragmens szekvenálásával validáltuk. A VHSΔ vírus terjedési tulajdonságai (lassabb) egérmodellen voltak tesztelve, az erre vonatkozó adatokat egy korábbi munkánkban mutattuk be¹.

5.11.1.2 PRV *ep0Δ*

A mutáns vírust Boldogkői és munkatársai³⁵³ által korábban publikált módszer szerint készítettük. Röviden, az alábbiak szerint: a PRV genom *KpnI*-F fragmensét beklónoztuk a pRL425 plazmid *KpnI* helyére³⁵⁴. Az eredményül kapott rekombináns plazmidot használtuk a *BamHI*-14 DNS fragmens eltávolítására, amely majdnem az egész *ep0* gént tartalmazta, ezt követően pedig egy lacZ génexpressziós kazettát beillesztettük az eltávolított gén helyére. Ezt a 'targeting' plazmidot használtuk az *ep0* gén PRV genomból való deletálására, homológ rekombináción alapuló technikával. A mutáns vírusokat a kék plakk fenotípus alapján választottuk ki többszöri plakk tisztítási lépéseken keresztül. A mutáció jelenlétét Southern blot és PCR technikákkal ellenőriztük. Annak megerősítésére, hogy az eltávolított gén egyedül felelős a vírus megváltozott fenotípusáért, a mutációt kijavítottuk és ezt a revertánst használtuk kontrollként, a növekedési tulajdonságait elemeztük.

5.11.1.3 PRV *ul54Δ*

A teljes *ul54* gént tartalmazó *Bam*HI fragmenst agaróz gélben izoláltuk, majd a pRL525 plazmidba klónoztuk³⁵⁴. Az ennek eredményeként született rekombináns plazmidot használtuk templátként a PRV célzott genomi régiójával homológiát biztosító két határoló (flanking) szekvenciájának PCR amplifikációjához. Az egyedi *Eco*RI helyet az *ul54* génbe inszertáltuk, egy 1017 bp-os helyre (2901 és 3919 bp között) a PCR során. Ezt követően az *Eco*RI helyekkel körülvevett, GFP-t kifejező kazettát (pEGFP-N1 vector, Clontech) inszertáltuk a targeting szekvencia *Eco*RI helyeire. Az eredményként kapott rekombináns plazmidot használtuk, mint transzfer konstrukt a génkiütött (KO, vagy *ul54Δ*) vírus előállításához. A linearizált targeting plazmidot a vad típusú Ka törzsből izolált DNS-sel együtt transzfectáltuk a PK-15 sejtekbe. A rekombináns vírus homológ rekombinációval készült, majd inverz fluoreszcens mikroszkóppal elemeztük és a fluoreszcencia alapján izoláltuk a rekombináns vírusokat. Az első izolátumot plakk-tisztítottuk, s ezt addig ismételtük, míg az esetleges kontamináció (vad vírustörzs) el nem tűnt.

5.12 ENGEDÉLYEK

PRV: NÉBIH (02.1/774//1/2014)

A további vírusokkal való munkát a kollaborációs partnerek laborjaiban végeztük, az engedélyeket amennyiben szükséges volt, a publikációkban feltüntettük.

6 EREDMÉNYEK

Ebben a munkában bemutatom a hosszú read-szekvenálási technológiák potenciálját a vírus transzkriptomok elemzésében a csapatunk által tanulmányozott vírusokon keresztül. A virális transzkriptomokat felhasználva szemléltetem a harmadik generációs szekvenálási technológiában rejlő lehetőségeket vírus RNS-ek izoformái sokféleségének feltárására, a transzkriptom szerkezeti összehasonlítására. A PRV, EHV-1, BoHV-1, VACV és VSIV esetében a vírus génexpresszió elemzésében való alkalmazhatóságukat fogom hangsúlyozni, míg az AcMNPV kapcsán az RNS-módosítások detektálásának képességére koncentrálok. Bár ezek lesznek a konvergencia pontjaim, az izoformák detektálását és annotálását vizsgáló fejezetben, a szemléltetés kedvéért, további vírusokkal kapcsolatos tanulmányaik egyes részeit is felhasználom. Bemutatom, hogy a virális transzkriptom komplexitás az egészen pici cirkovírusokra (PCV), illetve az RNS genomú vírusok egy-egy képviselőjére (VSIV, PERV) is igaz. Bemutatom, hogy az LRS módszerek miképpen alkalmazhatók a vírusok kinetikai osztályainak, illetve génexpressziós dinamikájuknak a meghatározására (PRV, BoHV-1, HSV-1, EHV-1, VACV, VSIV).

További példákon keresztül illusztrálom, hogy az általunk feltárt, a virális transzkriptomra jellemző komplexitás bár általános érvényű az összes vírusra, melyet az elmúlt évtizedben elemeztünk, a különféle jellegzetességek jobban, vagy kevésbé hangsúlyosak egyes víruscsoportok esetén.

Betekintést nyújtok a különféle vírusok, gazdasejtek expressziójára gyakorolt hatásával kapcsolatos eredményeinkről, valamint az általunk leírt legfontosabb víruscsoport, a replikáció asszociált (ra)RNS-ek típusairól a herpeszvírusok közt, illetve potenciális funkciójukról.

6.1 TECHNOLÓGIAI ÚJÍTÁSOK: „WETLAB”, BIOINFORMATIKA

Amellett, hogy elsőként vezettük be az LRS módszert a virális transzkriptomok vizsgálatára, a kutatásaink során felmerülő problémákra újításokat kellett foganatosítanunk, egyedi megoldásokat kellett kidolgoznunk mind a laboratóriumi munkafolyamatok, mind pedig a bioinformatika területén.

6.1.1 Új technológiai megközelítések

Elsőként vezettük be a PacBio-t és az ONT-t is a vírus transzkriptom kutatásba. Mindkét cég mindegyik módszere okoz valamilyen problémát, ezért kombinált megközelítéseket alkalmaztunk, valamint számos kiegészítő technikát: LRS [amplifikált és nem amplifikált (ún. direkt cDNS) szekvenálás, direkt (d)RNS szekvenálás], SRS, CAGE-Seq, cap-szelekció, Terminator kezelés, más, független adatszettek alkalmazása.

6.1.1.1 *Transzkript annotálás*

- Transzkripció start helyek (TSS) megállapítása
- Transzkripció vég helyek (TES) azonosítása
- Splice helyek, intronok megállapítása
- RNS molekulák annotálása

Az alábbi, az annotálás során felmerülő problémákat laboratóriumi módszerekkel oldottuk meg:

- a) Az ONT dRNS szekvenálás a readok 5' végét nem határozza meg, melynek okai pl: a motorfehérje leállása, idegen feszültségcsúcsok, elakadt pórusok feloldása³⁵⁵
 - megoldás: cDNS szekvenálás
- b) A PacBio Iso-Seq módszere nem tudja meghatározni az 5' végek kémiai természetét (cappelt-e, 5'-monofoszfát, vagy 5' hidroxil csoportot tartalmaz), mivel a template switching akkor történik meg a könyvtárkészítés során, amikor reverz transzkriptáz eléri az RNS molekula 5' végét³⁵⁶⁻³⁵⁹. Ugyanez igaz az ONT cDNS szekvenálási módszereire is.
 - megoldás: szelekció a cap-es, azaz intakt mRNS-ekre (Terminátor kezelés, cap-szelekció a könyvtárkészítés során és CAGE-Seq)
- c) Az RT és a PCR bizonyos gyakorisággal fals intronokat generálnak
 - megoldás: ONT dRNS szekvenálás
- d) Az ONT és a PacBio RNS szekvenálásra alkalmas könyvtárkészítő kitjei mind oligo(dT) primer alkalmazásával működnek, tehát az első lépés, az RT szűr a poliA farokkal rendelkező RNS-ekre. Így egyrészt elveszítjük a hosszú, nem poliadenilált nem-kódoló (nc)RNS-eket, illetve nem kapunk meg 5' végeket valós gyakorisággal.
 - megoldás: random primerek alkalmazása az RT-hez

Az optimális, ha minden létező módszert kombináltan alkalmazunk.

6.1.1.2 *Dinamikus transzkriptom vizsgálata*

A gének expressziós dinamikájának, kinetikájának meghatározásához alapvető, hogy a különféle transzkriptek mennyiségileg összehasonlíthatók legyenek, azonban az összes technika okoz valamilyen elhajlást (bias) az eltérő hosszúságú transzkriptek tekintetében.

Az alábbi problémákkal szembesültünk

- a) Szekvenálási bias: mind a PacBio, mind az ONT módszerre jellemző, hogy ún. „fragment length bias”-t okoznak, azaz az eltérő hosszúságú DNS (vagy RNS) szálakat eltérő valószínűséggel szekvenálják meg. Ennek oka, hogy a hosszabb nukleinsavak kisebb valószínűséggel kerülnek be egy-egy ZMW-be, illetve nanopórusba. Itt a valóban hosszú, több kbp-os leolvasások alulreprezentációja jelenti a problémát. Ez a probléma a genom szekvenálás esetén kevésbé jelentős (bár ez esetben is a minél hosszabb leolvasások az optimálisak az illesztések megkönnyítése érdekében), ahol egyrészt nem cél az egyes leolvasások kvantitálása, másrészt különféle módszerekkel (pl. g-TUBE alkalmazása) a fragmensek mérete egységesíthető. A génexpressziós vizsgálatok esetén azonban egyrészt alapvető jelentőségű a mennyiségi elemzés, másrészt a hosszú transzkriptek esetleges, a technikai hiba miatti elvesztése fals eredményekhez vezethet.
 - megoldás: gélből izolálás manuálisan, vagy SAGE és BluePippin készülékkel, eltérő könyvtárkészítési protokoll a PacBio szekvenáláskor, melyekkel még a szekvenálást megelőzően fel tudtuk dúsítani a hosszú transzkripteket
- b) Mintafelviteli bias: a PacBio szekvenáláskor az ún. MagBead-es mintafelvitel (loading) nem kedvez a ~ 800 bp-nál rövidebb fragmenseknek, ezek kimosódnak (eliminálódnak) a mintából.
 - megoldás: ONT dcDNS módszer alkalmazása, mely a 200-800 bp-os mérettartományba eső átiratok azonosításához optimálisabb¹³.

- c) PCR bias: a PCR okozta mennyiségi eltérés a különböző fragmens hosszak között: adott kondíciók mellett a PCR során a rövidebb fragmensek szelektíve jobban feldúsulnak
- megoldás: eltérő ideig tartó annealing és eltérő ciklusszám alkalmazása^{360,361}, vagy nem amplifikált IsoSeq a PacBio esetén,
 - dcDNS szekvenálás az ONT esetén

6.1.2 Bioinformatika

A szekvenálás/adatelemzés során számos olyan probléma is felmerült, melyek bioinformatikai megoldást igényeltek. Egyrészt, amikor elkezdtünk LRS transzkriptomikával dolgozni, egyáltalán nem volt olyan program, mely a transzkriptek annotálására alkalmas lett volna. Emiatt saját feldolgozási metódusokat kellett kidolgozzunk, mely kezdetben a meglévő programok kombinált alkalmazását jelentette, majd a csoportunk munkatársai írtak megfelelő scripteket és programcsomagokat. A LoRTIA programcsomag³⁶² az, amely a potenciális, a template switchingből származó fals, alternatív TES-eket és intronokat kiszűri²⁸⁸.

6.1.3 Génkiütött vírusok transzkriptomának analízise RT²-PCR-rel

Kidolgoztunk egy saját matematikai megközelítést (lásd. **5.8. fejezet**), mellyel a vírusgének expressziója kvantitálható és a fertőzés különböző időpontjaiban mért adatok egymással összehasonlíthatók.

6.2 VÍRUS GENOMOK ANNOTÁLÁSA: DNS ÉS RNS SZEKVENCIÁKBÓL

Az RNS molekulák azonosításhoz szükség van referencia genomokra. Érdekes 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 genomot⁴ (GenBank: KJ717942.1). A Ka törzs mellett egy általunk generált géndeletált törzset²⁸⁰, illetve egy szintén általunk, újonnan izolált PRV-t¹³¹, valamint a PCV-1¹⁷⁸ genomot is megszekvenáltunk. A PERV²⁹⁴ és a VACV²⁹⁵ genomot pedig a transzkriptek 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 leolvasásokból²⁴.

6.2.1 Pseudorabies vírus (PRV)

6.2.1.1 PRV Kaplan törzs

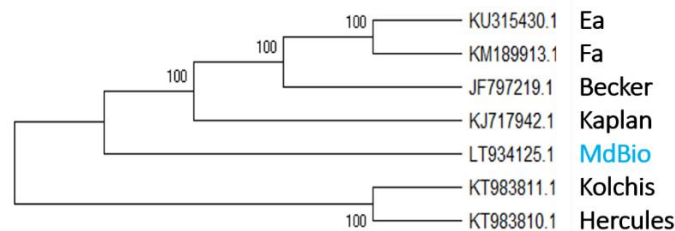
A PRV eredeti genomszekvenciája hat vírustörzsből³⁶³ előállított mix volt, amelyet a hagyományos Sanger-módszerrel határoztak meg (Becker, Indiana-, Funkhauser, Ka, NIA-3, Rice, and TNL). A PRV Ka törzs és más törzsek teljes genomját Sanger kapilláris szekvenálás³⁶⁴ és Illumina, nagy lefedettséget biztosító "deep" szekvenálás^{129,365,366} segítségével határozták meg. A PacBio technika nagy előnye, hogy a hosszú leolvasások révén a genomokat *de novo* is rendkívül könnyű összeállítani, illetve a magas G+C tartalmú, ripitekben gazdag genomok is pontosan meghatározhatók. A PRV közel 74%-os G+C tartalmú, ismétlődésekben gazdag duplaszálú (ds)DNS

vírus (~143 kbp), ezért nehéz szekvenálni akár a Sanger, akár a második generációs rövid leolvasási technikákkal.

A PRV Ka szekvenálás eredményeként extrém nagy lefedettséget (1200x) kaptunk, melyet a nyert 78111 ROI eredményezett. Az így összeállított virális dsDNS genomot 143423 bp alkotja, a G+C tartalom 73,59%. Hetven protein kódoló, két látencia-asszociált gént annotáltunk, továbbá 16 miRNS-t prediktáltunk⁴.

6.2.1.2 PRV MdBio törzs

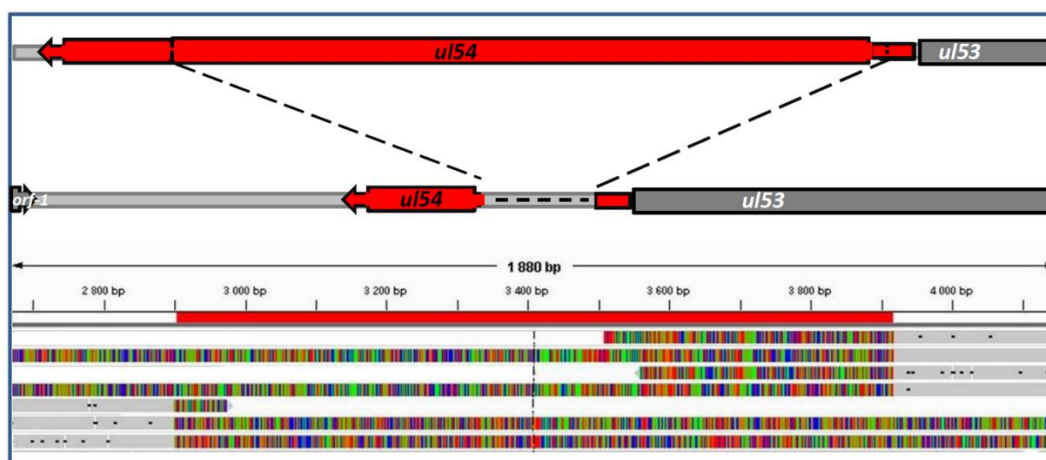
A genomot RSII szekvenátorral határoztuk meg, mely összesen 4450 kiváló minőségű ROI readet eredményezett. Ez összességében átlagosan 70x lefedettséget biztosított a teljes genomon. Az összeállított vírusgenom 142922 bp-os, 73,56%-os G+C tartalmú dsDNS. Két hosszabb inszerciós/deléciós variációt, valamint két C-T tranzíciót azonosítottunk¹³¹. Számos PRV törzzsel összevetve, úgy találtuk, hogy az általunk izolált MdBio törzs a Kínában izolált Ea törzstől (GenBank: KU315430) áll filogenetikailag a legtávolabb, míg a Ka törzshöz a legközelebb (**12. ábra**).



12. ábra **Evolúciós törzsfá**¹³¹. Az általunk izolált MdBio törzs, a szintén általunk megszekvenált Ka laboratóriumi törzs, valamint öt további PRV törzs (Ea, Fa, Becker, Kolchis, Hercules) teljes genomszekvenciáit felhasználva végeztük az evolúciós elemzést a maximum likelihood (legnagyobb valószínűség) módszerrel. Az ábrán a nevek mellett a GenBank azonosítók szerepelnek.

6.2.1.3 PRV ul54Δ

A deléció jelenlétét PacBio szekvenálással igazoltuk (**13. ábra**). A wt vírussal való összehasonlítás kimutatta a tervezett deléciót, s mindösszesen ez az egy eltérés mutatkozott a két vírustörzs szekvenciája között.



13. ábra **Az ul54 deléciós mutáns PRV genomi régiója**²⁸⁰. Az ábra felső blokkja az érintett deléciós génszakaszt sematikus ábrázolja, míg lent a PacBio readek IGV vizualizációja látható.

6.2.2 Porcine circovirus 1 (PCV-1)

A PCV-1 szekvenálás eredményeként átlagosan 1170 nukleotid (nt) hosszúságú ROI readeket kaptunk¹⁷⁸, melyek átlagosan 48,37x-es (42-57x) lefedettséget eredményeztek a referencia genomon. Az általunk meghatározott PCV-1 törzs cirkuláris, ssDNS genomja 1759 bp, az átlagos G+C tartalma 48,44%, s mindösszesen két fehérje kódoló gént tartalmaz. A referencia szekvenciától kilenc pontmutációban, egy inszerciós mutációban tér el. Az NCBI nukleotid adatbázisban elérhető DNS szekvenciákkal való összehasonlítás a sejtkultúrához és az alkalmazott kondíciókhoz való genetikai alkalmazkodásra utal, mivel a nem konzervatív változások aránya a csendes mutációkhoz viszonyítva viszonylag magas (KN/KS = 1,89). Ezenkívül heterogenitást azonosítottunk három genomi lokációban: a 67. pozíciónál a C-G szubsztitúció (arány: 33%); az 1105. pozíciónál C-T csere (arány: 43%), míg az 1503. pozíciónál szintén C-T csere (arány: 32%).

6.2.3 Porcine Endogenous Retrovirus (PERV)

A szekvenálás eredményeként 17544 ROI readet kaptunk, mely 3238x-es átlagos lefedettséget jelentett a teljes genomon. A ROI-k átlag hossza 1555 nt. A PERV Szeged törzsének genomja 8673 bp hosszú és három fehérjekódoló gént tartalmaz. G+C tartalma 49,3%. A sertésből izolált törzshöz képest 323 pontmutációt azonosítottunk, az AT-GC mutációk száma (179 pozíció) a fordított, GC-AT cserékhez képest (110 pozíció) 1,62x magasabb. A polimorfizmus analízise során megfigyeltük, hogy az *env* gén körül van a legváltozatosabb régió (165 eltérő nukleotid), míg a gag-pol gén 46 változó nukleotidot tartalmaz²⁹⁴.

6.2.4 Vaccinia vírus (VACV)

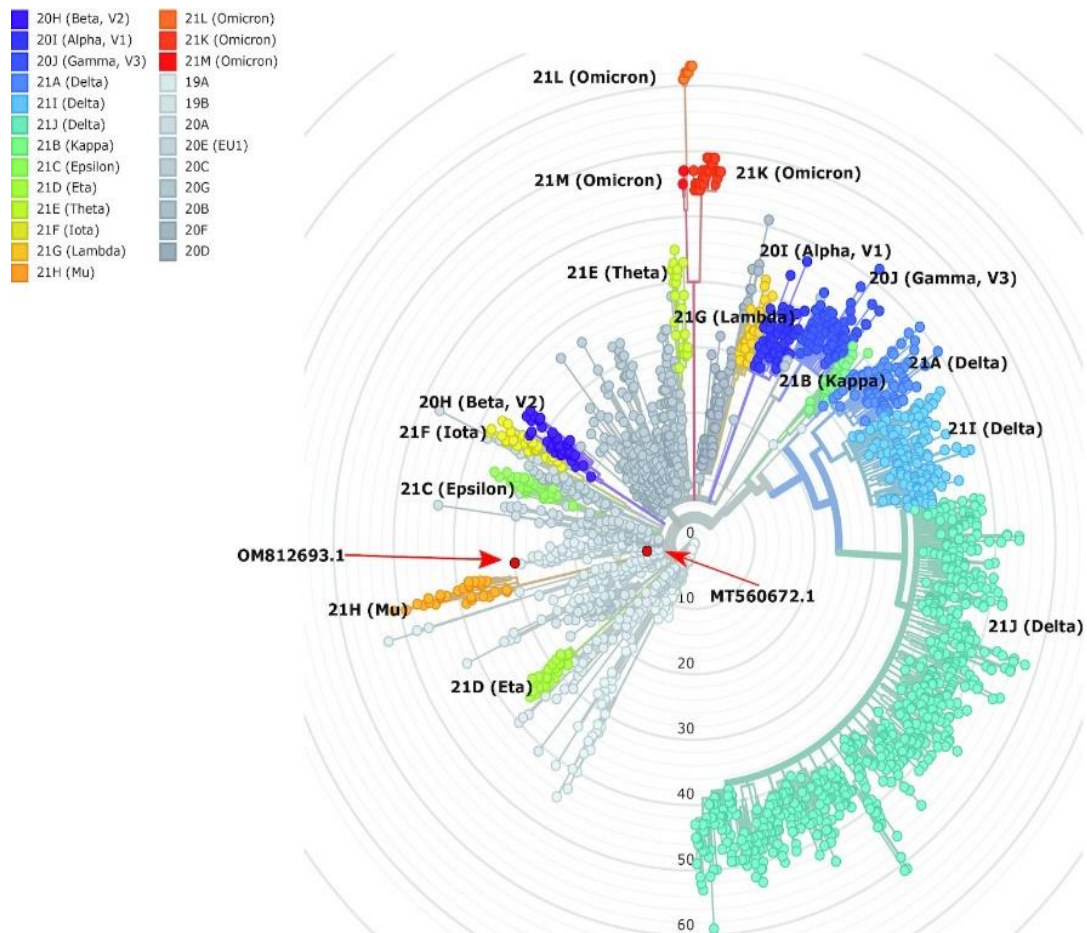
A VACV WR törzse esetén a PacBio amplifikált cDNS szekvenálás módszerét a transzkriptom analízisre alkalmaztuk²⁹⁵. Az elemzés során megfigyeltük, hogy a genom teljes egészében az elsőtől az utolsó nukleotidig transzkripcionálisan aktív. A nagy lefedettség és a PacBio CCS szekvenciák (ROI) adta precizitás lehetővé tette, hogy összeállítsuk a teljes VACV genomot²⁹⁵; az elemzésbe az ONT MinION adatokat is bevontuk. Összesen 59154 PacBio és 16175 ONT readet térképeztünk. Mindkét szekvenálás eredményezte readek a teljes vírusgenomot lefedik. A PacBio 289,3x-es, míg a nanopórusos megközelítés 43,6x-es lefedettséget, előbbi 953 bp-os, míg utóbbi 525 bp-os átlagos read hosszt adott. Az általunk megszekvenált WR törzs 194888 bp, 33,3%-os G+C tartalommal. A dsDNS genom 218 protein-kódoló gént tartalmaz. A GenBank referencia szekvenciához képest 163 pontmutációt azonosítottunk, melyek közül 58 funkcionális (nonsilent) mutáció. Tizennyolc inszerciót és négy deléciót is detektáltunk. A pontmutációk közül a C-T csere a leggyakoribb. A legtöbb aminosav cserét a C9L génben azonosítottuk. Egy új (potenciálisan) kódoló gént (A54L) annotáltunk. A nukleinsav-változások többsége (összesen 21 esemény) a vírusgenom úgynevezett F régiójában csoportosul. A mutációk főként olyan vírusgénekben történtek, amelyek szerepet játszanak a gazda védekezésének modulálásában, a transzkripció szabályozásában és a vírus replikációjában.

6.2.5 SARS-CoV-2

A SARS-CoV-2 RNS genomú vírus, így az izolált RNS, melyet a könyvtárkészítéshez használtuk keverten tartalmazta a genomi RNS-t és az mRNS-eket, illetve egyéb transzkripteket.

Az sgRNS-eket úgy határoztuk meg, mint azok az RNS-ek, amelyek átfednek valamely szubgenomikus ORF-fel és történt bennük egy template switching, mely összeköti ezt a térképezett (mappelt) régiót a genom 5' leader szakaszával (a referenciagenom 55–85 pozíciójában). A gRNS-eket pedig úgy definiáltuk, hogy ezek azok az RNS-ek, melyek átfednek az ORF1ab-vel (legalább 10 nt-dal) és nem tartoznak a szubgenomikus kategóriába. Az összes többi transzkriptot „unclassified”, azaz osztályozatlan RNS-ként kategorizáltuk.

A saját izolátumunk genomja (OM812693.1) összeállításához a 20 kbp-nál hosszabb readeket használtuk. Mindösszesen 109 db e kategóriába illő leolvasásunk volt, melyek elegendőnek bizonyultak egy 29782 bp-os kontig összeállításához (SARS-CoV-2_Hun-1_GenomeDraft_v1). A nyers (draft) genom szekvenciát összehasonlítva az eredeti, vuhani törzs szekvenciájával, összesen 30 mutációt detektáltunk, közülük 3 frame shift mutáció. A szekvenciánk az MT560672.1 genomhoz áll legközelebb, mindkettő a 20A (EU1) kládba tartozik²⁴ (14. ábra).



14. ábra A filogenetikai fa bemutatja a szekvenált SARS-CoV-2 törzseket, a vírus meghatározott kládjai szerint²⁴. Az általunk izolált törzset pirossal jelöltük, és egy piros nyíl mutatja a pozícióját (OM812693.1). A felhasznált referenciagenomot (MT560672.1), szintén egy piros nyíllal jelöltük. A fát a Nextstrain pipeline-nal generáltuk. Az összes variánst a nekik rendelt klád szerinti nomenklatúra alapján színeztük.

6.3 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^{6,272}. Ezt követte számos további vírus, úgy, mint a HSV-1⁸, HCMV^{9,286}, VACV¹⁷, PCV-1²⁹² és a PERV²⁹³, ugyanezen módszerrel. Majd az újabb technikák megjelenésével alkalmazni kezdtük a PacBio Sequel II-t [PRV¹⁹, HSV-1²⁸⁸, HCMV²⁸⁷, VACV²⁷⁴, AcMNPV²⁷⁷] és az ONT MinION [PRV¹³, VZV¹⁴, EBV²², HSV-1²⁸⁸, AcMNPV²⁷³, VACV²⁹⁷, ASFV²⁰, HCMV²⁸⁷, BoHV-1²⁸², PERV²⁹³, IVA, CVB5, VSIV, WNV, CCHFV és ZIKV²⁸⁵, SARS-CoV-2²⁴, EHV-1²⁵, MPXV²⁹¹, KSHV²⁶] 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^{22,26,289}. Sok esetben pedig, főként azon vírusok esetében, ahol korábban még SRS platformot sem használtak transzkriptomikai céllal, ott mi ezt is elvégeztük (PRV³, EHV-1²⁵, ASFV²⁰, PCV¹⁷⁸), és bevezettünk egy SLRS²⁸² módszert is. 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 fejeződnek ki, illetve, hogy az adott génről leíró izoformák közt figyelhető-e meg időbeli eltérés.

A következőkben először általánosságban ismertetem a virális transzkriptomra jellemző jelenségeket, melyeket a kutatásaink során tártunk fel, majd ezeket egy-néhány vírus példáján keresztül szemléltetem. E disszertációban a hangsúlyt elsősorban a herpeszvírusokra helyezem, hiszen munkám egy jelentős részét az e vírusokkal való munka jelentette. Azon jellegzetességeket, melyek más víruscsaládokra tipikusak, azokat az adott család egy-egy képviselőjével mutatom be.

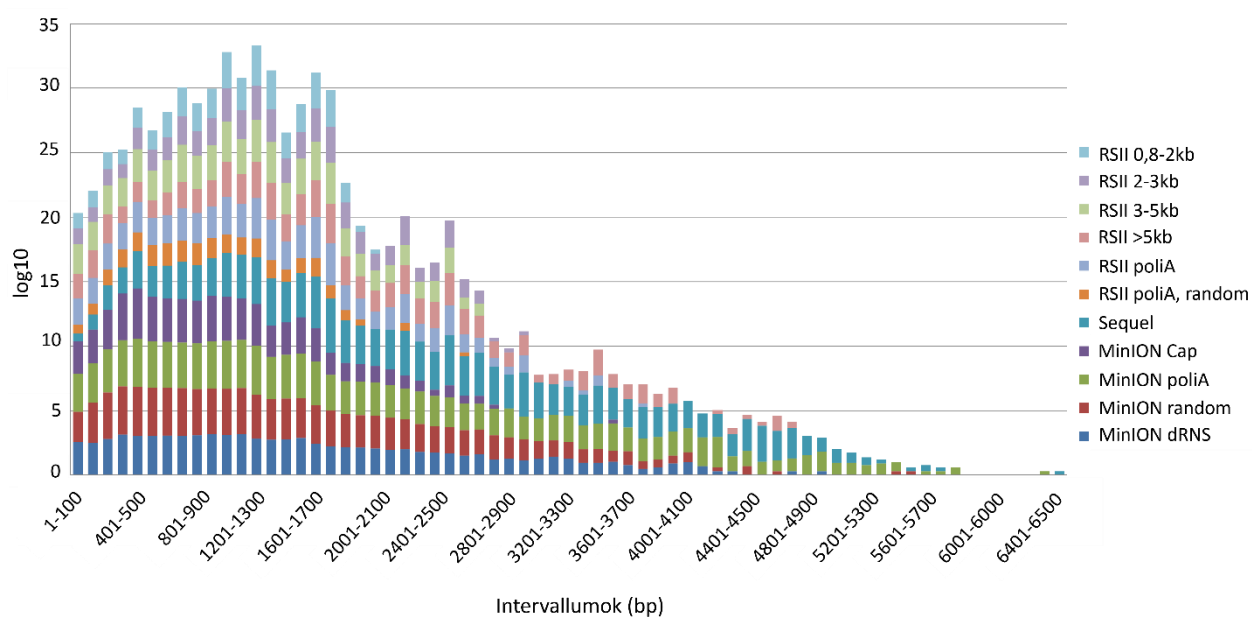
6.3.1 Read-hosszak

Az alábbi, **4-es táblázat**ban foglaltam össze a vírus transzkriptom szekvenálás statisztikai adatait. A jóval részletesebb táblázatok a publikációkban, azok kiegészítő anyagaiként megtekinthetők.

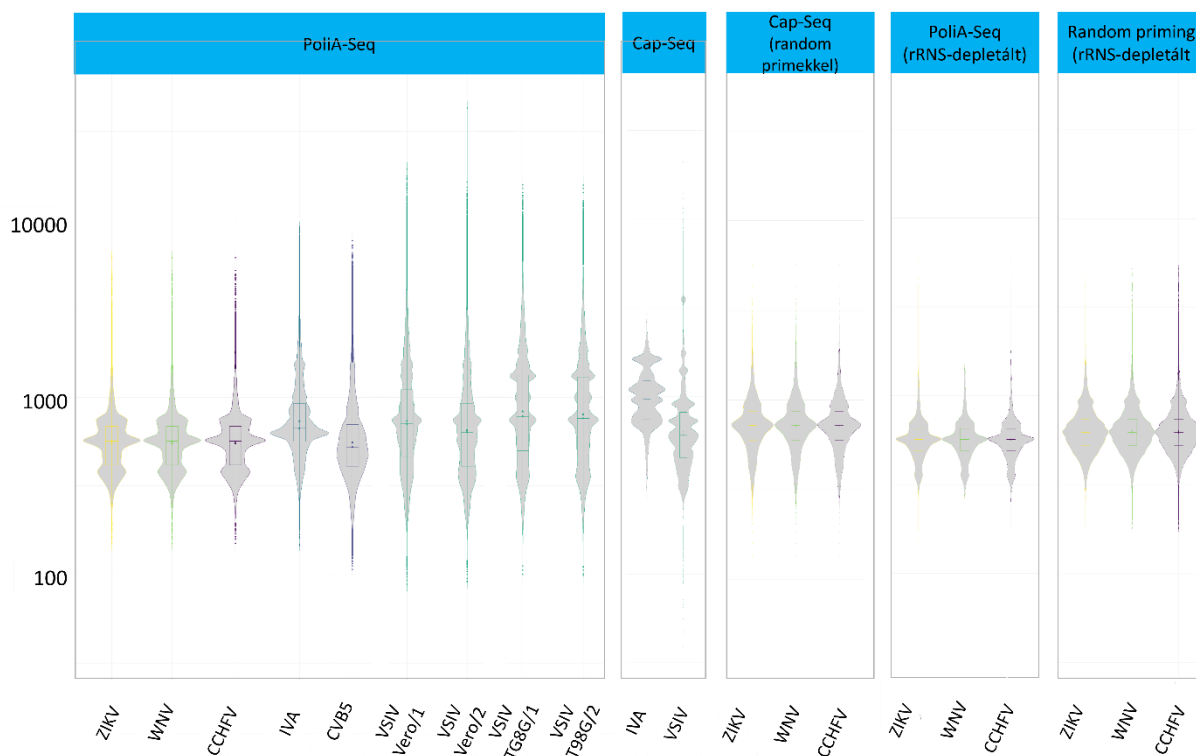
A különféle könyvtárkészítési és szekvenálási metódusok eltérő átlagos readhosszakat eredményeznek. A leolvasások átlag hosszát az adott vírusra jellemző ORF méretek (pl. VACV-nál, az ASFV-nél általában jóval rövidebbek, mint a HV-oknál), valamint az is befolyásolja, hogy a poligén transzkriptek mennyire gyakoriak. Továbbá, a technikák is okoznak eltéréseket (pl. ONT vs. PacBio; amplifikált vs. nem amplifikált cDNS; dcDNS vs. dRNS), a PacBio technikák az 1000 bp-nál hosszabb fragmenseknek kedveznek (a Sequel még a 4 kbp körüli tartományban is sok readet eredményezett), a MinION esetében nincs kifejezett méretpreferencia, az igen rövid mérettartományban is nagyszámú leolvasást kaptunk. A **15-ös ábra** a HSV-1-nél alkalmazott technikák eredményezte readhosszakat mutatja, míg a **16-os ábra** hat, RNS genomú vírus erre vonatkozó adatait foglalja össze.

Szekvenálás	Vírusra térképezett leolvasások száma	Átlagos readhossz (bp)
AcMNPV 1D cDNS-Seq	103133	1053
AcMNPV dRNS-Seq	2430	614
ASFV ONT	139711	953
ASFV Illumina	69068	2 x 150
BoHV-1 ONT	3450627	~ 1000
BoHV-1 LoopSeq	12618 (kontig)	150
CCHFV	528	683
CVB5	1508	630
EBV ONT	76629	999
EHV-1 ONT dcDNS	133838	910
EHV-1 ONT dcDNS	319379	1019
EHV-1 ONT dRNS	53625	722
EHV-1 ONT dRNS	40001	827
EHV-1 ONT dcDNS A+B+C	3706773	659
HCMV RSII	33086	1168
HSV-1 PacBio RSII	38972	1369
HSV-1 PacBio Sequel	80061	1924
HSV-1 ONT 1D cDNA-Seq	1505848	967
HSV-1 ONT Cap-Seq		683
HSV-1 ONT dRNA-Seq		823
HSV-1 ONT Terminator		873
IVA	10916	812
KSHV	1089861	714
MPXV dcDNS	1475053	800-1000
MPXV dRNS	318802	800-1000
PCV Illumina	197645	100
PCV PacBio	766	1170
PERV	17544	1555
PRV PacBio RSII	215417	1244
PRV PacBio Sequel	12555	1763
PRV ONT Cap-Seq	194232	810
PRV ONT dcDNS		786
PRV ONT dRNS		909
SARS-CoV-2 dcDNS	1527249	2209
SARS-CoV-2 dRNS	281418	1652
VACV PacBio RSII	86728	1098
VACV PacBio Sequel		1157
VACV ONT	568994	792
VSIV (Vero: 1, 2)	193820 / 273882	805 / 954
VSIV (T98G: 1, 2)	289807 / 255610	1021 / 1122
VZV	66455	1349
WNV	21790	491
ZIKV	592	634

4. táblázat A szekvenálásonként kapott readszámok (a teljesség igénye nélkül). A táblázatban minden esetben az oligo(dT) primerrel készített könyvtárakból származó eredményeket tüntettem fel, a random primerrel készített cDNS-ekről átlagosan rövidebb leolvasásokat kaptunk.



15. ábra A HSV-1 szekvenálása során kapott nyers readok hossz-eloszlása a különféle szekvenálási technikákat alkalmazva¹⁶. A technikák szerint felosztott oszlopdiaagram a log₁₀-es skálán mutatja a leolvasások számát, 100 bp intervallumokra bontva.



16. ábra A violin plotok a ZIKV, WNV, CCHFV, IVA, CVB5 és VSIV szekvenálása során kapott readhossz eloszlásokat mutatják, technikánként²⁸⁵. Az ONT 1D poliA szekvenálás esetén 566 és 1369 bp között, a Cap-Seq módszernél 725 és 1092 bp közötti értékeket kaptunk.

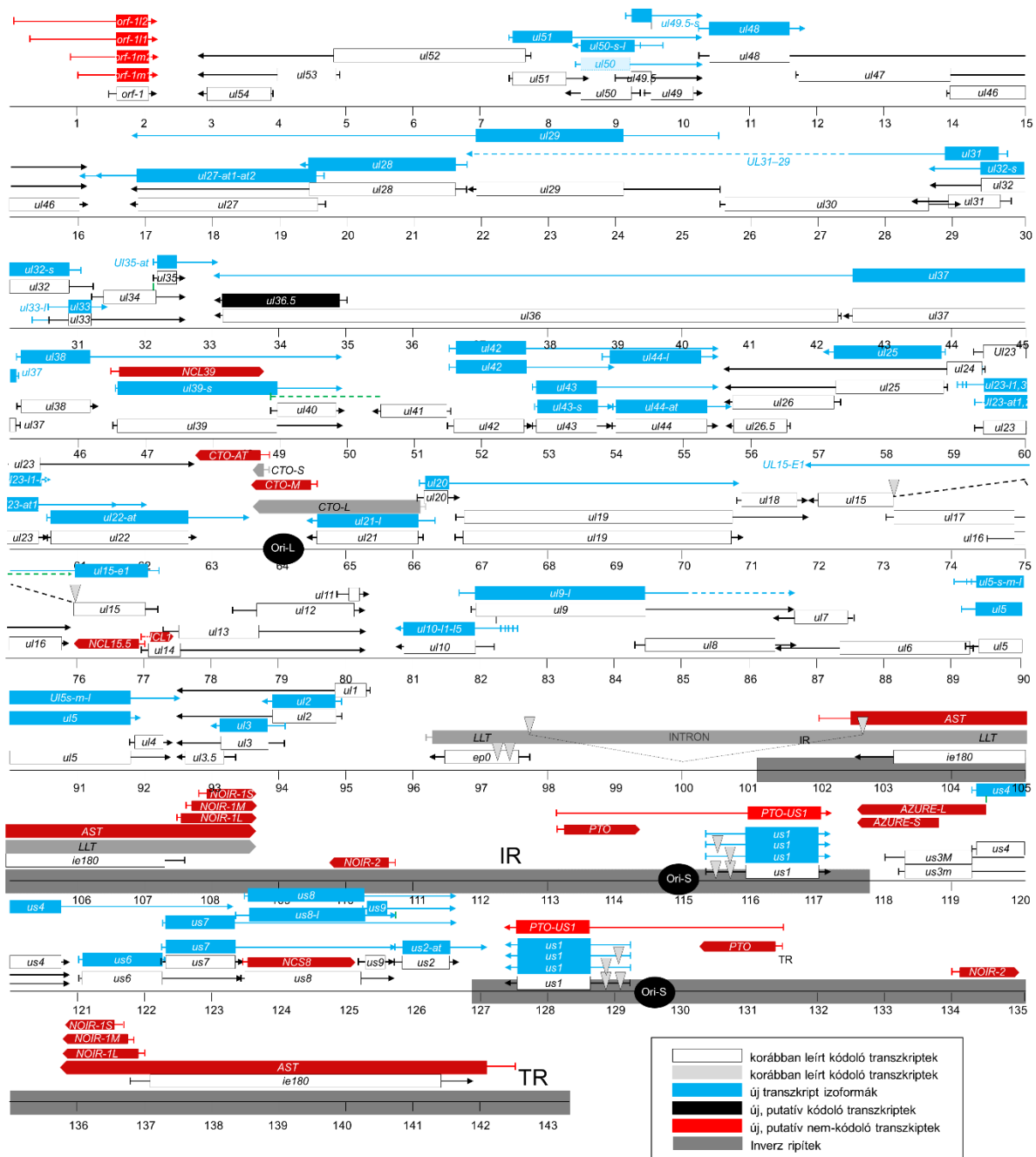
6.3.2 A virális transzkriptom rejtett komplexitása

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^{6,8,14,273}. A herpeszvírusok replikációs origóit átfedő, nem kódoló transzkripteket azonosítottunk, amelyek szerepet játszhatnak a DNS-szintézis szabályozásában^{6,22,25,272}, 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 génexpressziót¹⁷. Meghatároztuk mostanra számos humán- illetve állati patogén, így öt α -, egy β - és két γ 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 virális transzkriptom komplexitást, jellegzetességeit a következőkben a PRV, a HSV-1, az EHV-1, a KSHV, a VACV és a PCV-1 példáján szemléltetem.

6.3.2.1 PRV

A PacBio RSII alkalmazásával, ami a PRV RNS profiljának feltérképezésére irányult, úttörőkké váltunk a molekuláris virológiában^{6,272}. A módszer által egy korábban elképzelhetetlen komplexitást tárult elénk. Eredményeink megmutatták, hogy az LRS módszer alkalmas a teljes hosszúságú RNS molekulák 'deep', azaz megfelelő mélységű, szekvenálására, mind amplifikált, mind nem amplifikált minták esetén. Vizsgálataink eredményeként lényegében újraértelmeztük a PRV transzkriptomot. Kimutattuk, hogy a PRV jelentősen nagyobb genetikai komplexitást mutat (**17. ábra**), mint az az *in silico* ORF-alapú genom annotációk és a gél alapú esszék alapján várható volt. A PRV transzkriptek átfogó, genom-szintű átfedési mintázatának, valamint az Ori-átfedő RNS molekuláknak az azonosítása felvetette egy genom-szintű hálózat létezésének lehetőségét, amely együttesen irányítja a gének kifejeződését és a replikációt.

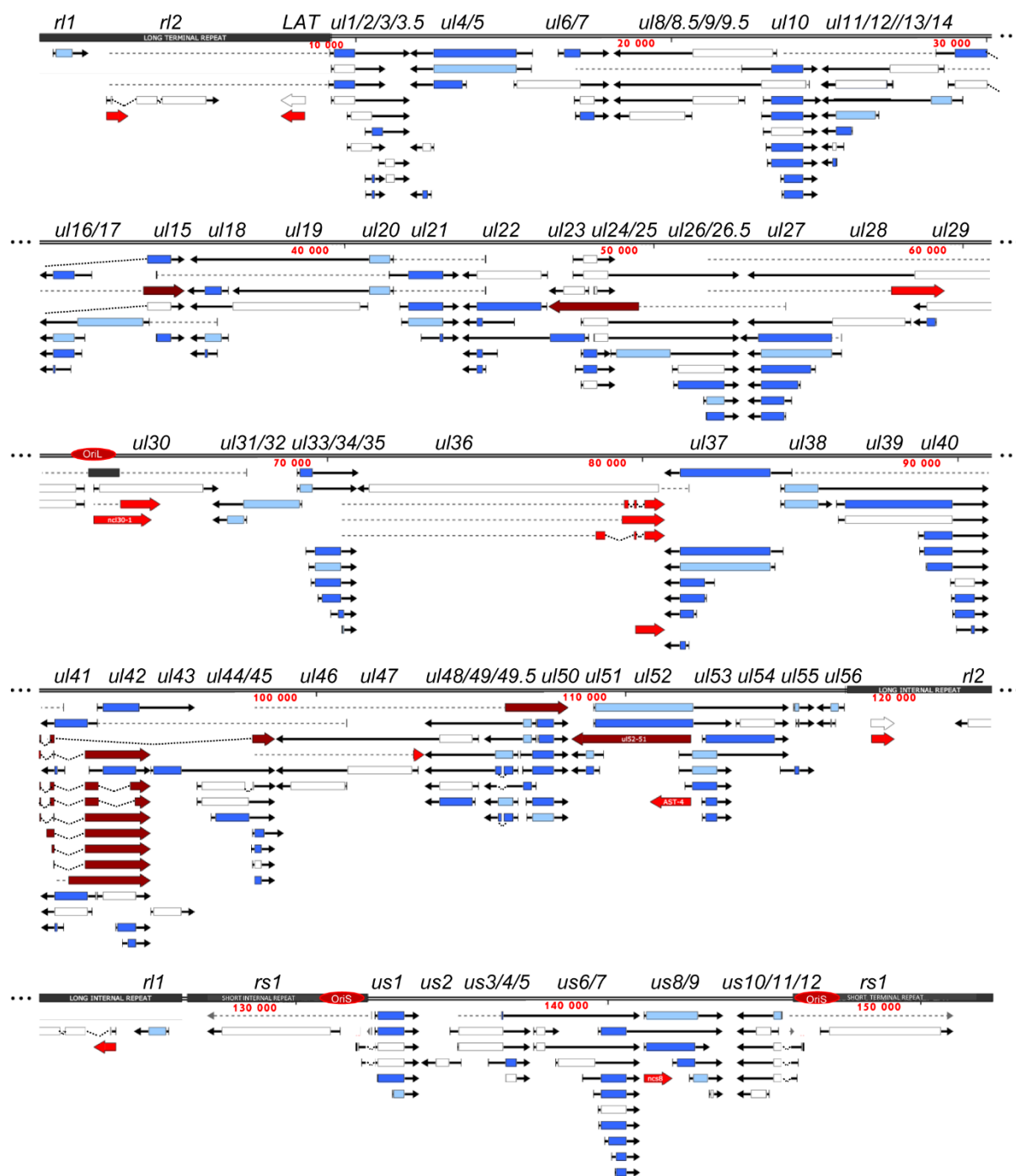
A PacBio Sequel és az ONT MinION megjelenésével a PRV transzkriptomot többször update-eltük^{13,19,278}, pontosítottunk korábban nem meghatározott TSS-eket, illetve az egyre nagyobb lefedettségük révén újabb transzkripteket fedeztünk fel. A **17. ábra** az RSII szekvenálás adataira épülő PRV transzkriptomot, a legabundánsabb RNS-eket mutatja.



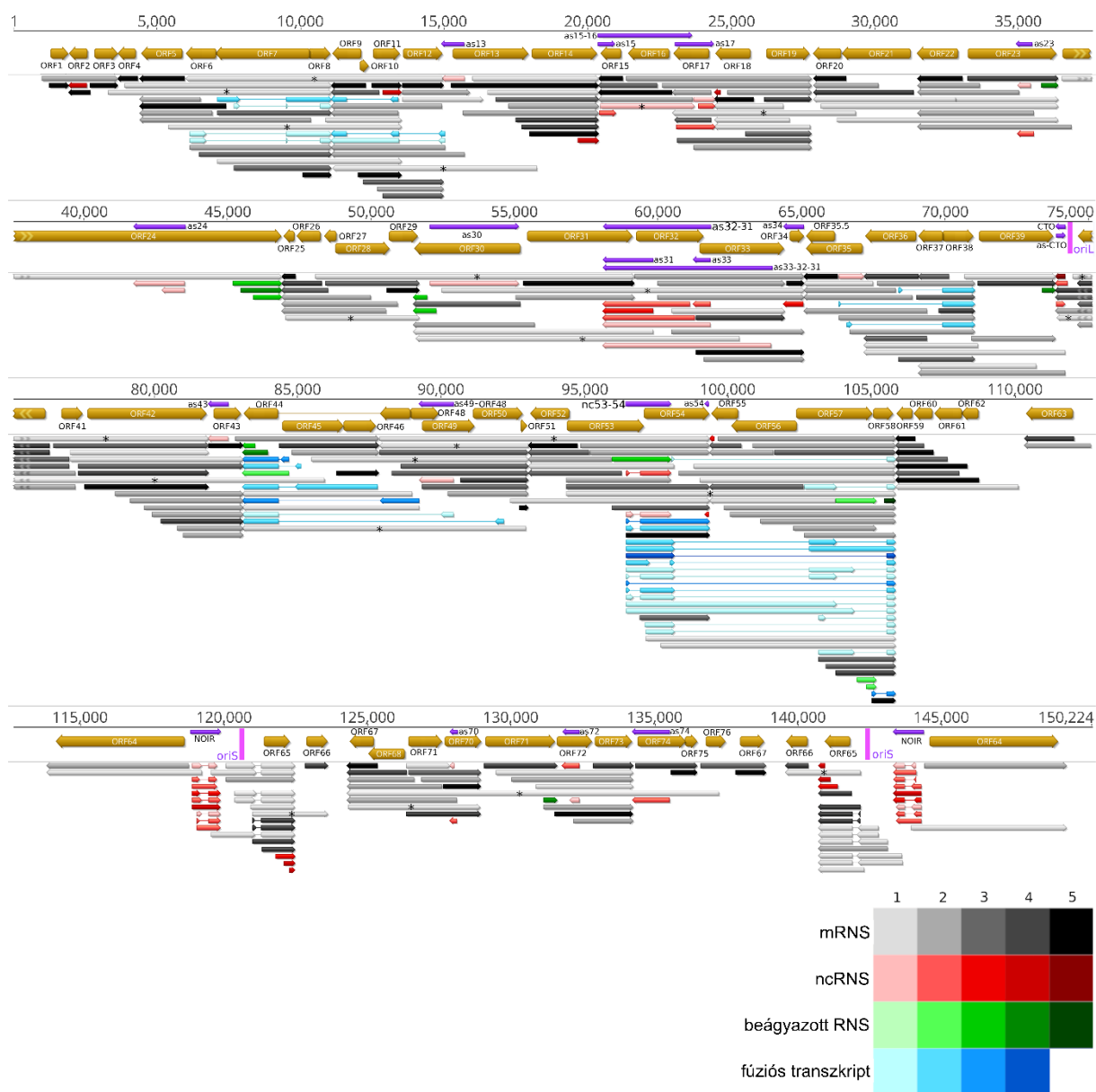
17. ábra A PRV transzkriptom⁶. Az első virális LRS transzkriptom szekvenálás (PacBio RSII) eredményezte RNS profil sematikus ábrázolása (Power Pointtal készült ábra).

6.3.2.2 Egyéb herpeszvírusok

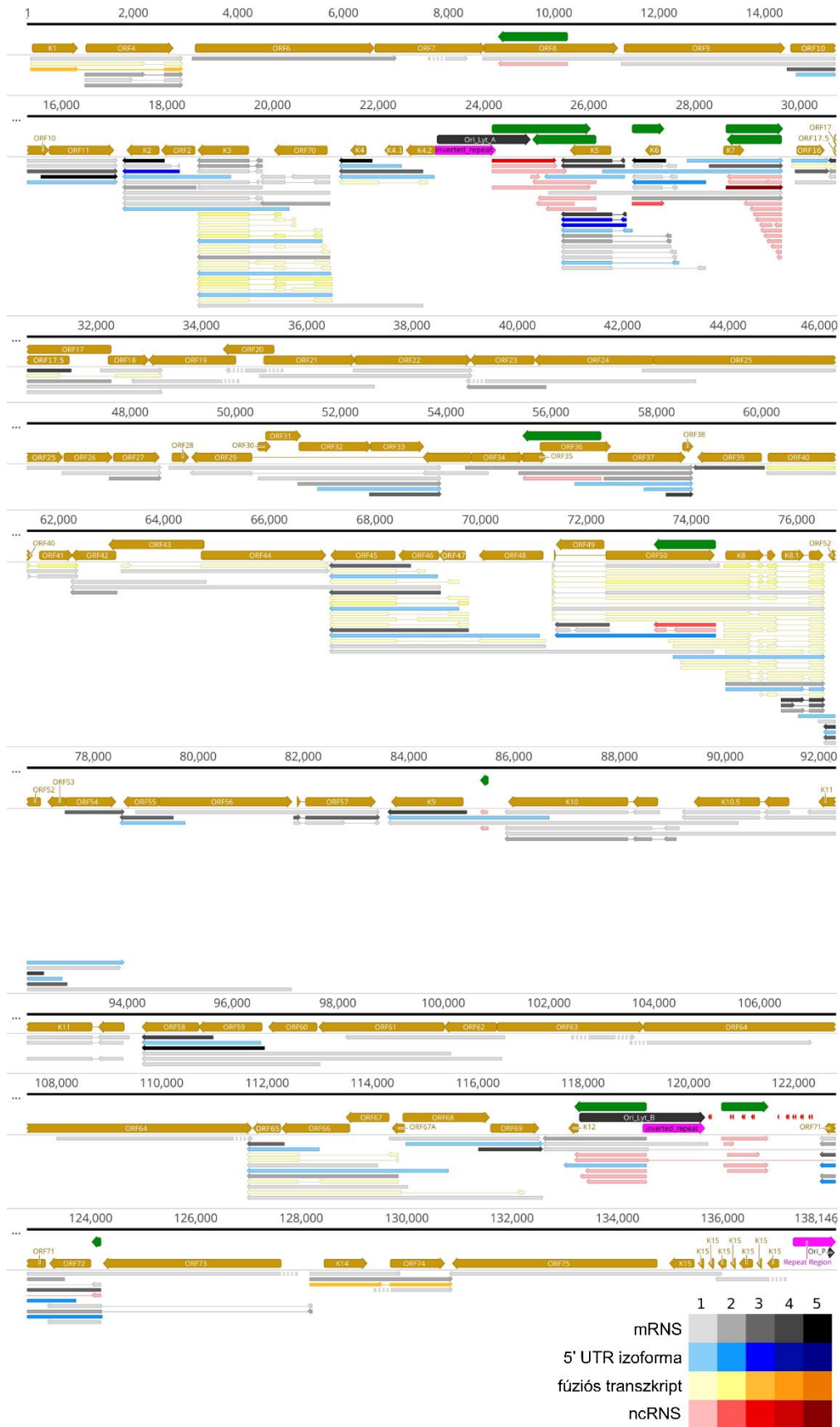
Mind az α -, β - és μ -herpeszvírusok tagjai RNS profiljának LRS-alapú analízise^{6,8,9,22,25,26} azt mutatja, hogy a bonyolult transzkripció átfedés-hálózat általános jelenség a herpeszvírusoknál (**18-20. ábrák**).



18. ábra **HSV-1 transzkriptom**⁸. Az RSII-t a HSV-1 transzkripció profilja meghatározására is alkalmaztuk. A korábban azonosított kódoló RNS-eket azok a nyilak jelölik, melyeken az ORF-eket fehér téglalappal ábráztuk. Az annotált TSS és TES pozícióval nem rendelkező, korábban ismert transzkripteket világoskékkel jelöltük, míg az új, potenciálisan fehérjekódoló géneket sötétkékkel. A már korábban azonosított lncRNS-eket sötétszürke nyilakkal, míg az új, feltételezett lncRNS-eket piros nyilakkal ábráztuk. A sötétvörös téglalap alakú nyilak olyan komplex transzkripteket mutatnak, melyek ellentétes orientációjú ORF-eket is átírnak. A hosszú fekete dobozok a HSV genom ismétlődő régióit jelölik; a piros oválisok a három Ori-t jelölik. Az RNS-ek nevében található rövidítések: S, rövid TSS variáns; L, hosszú TSS variáns; AT, alternatív TES variáns. Az ábra a SnapGene Viewer-rel készült.



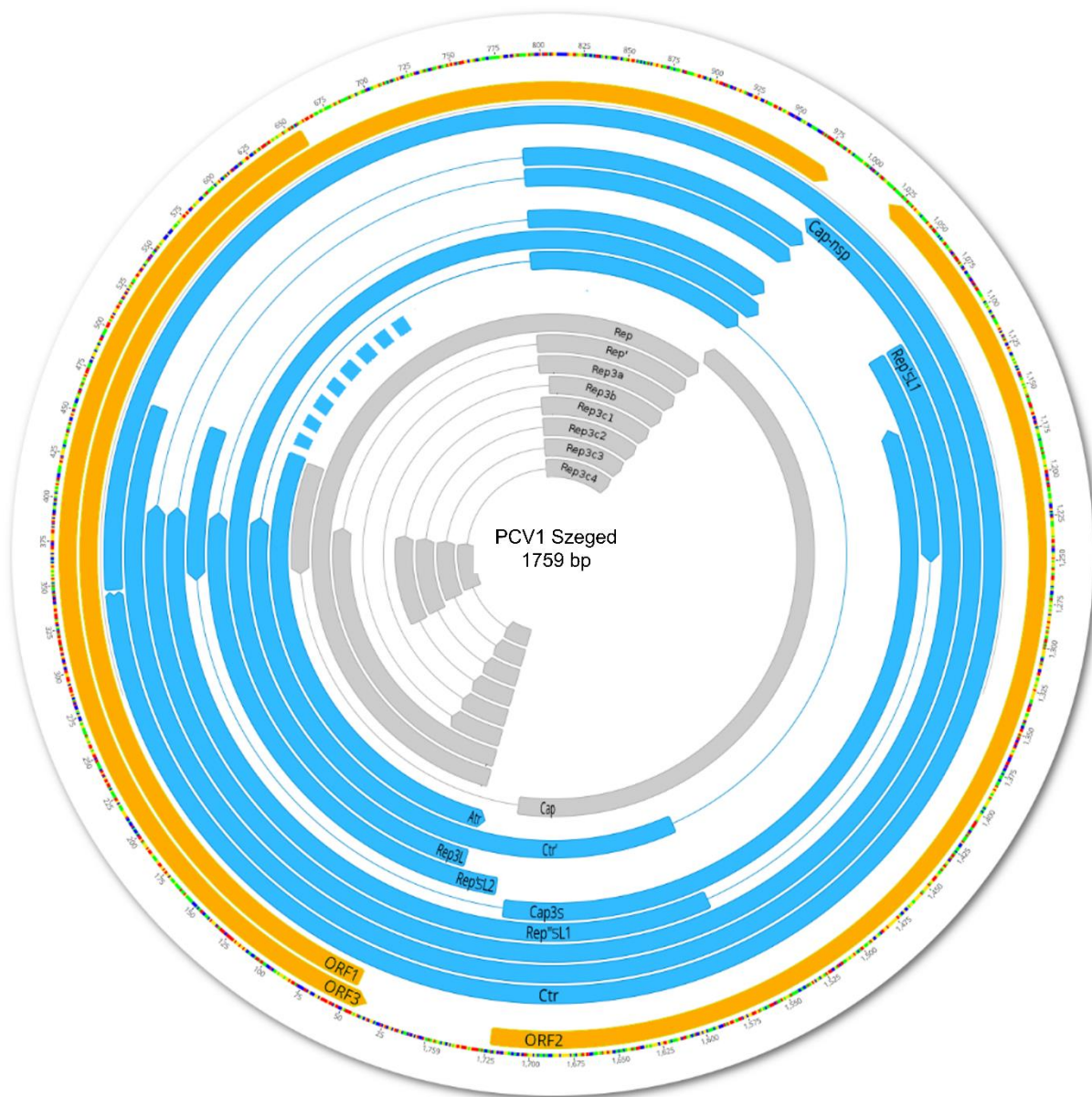
19. ábra EHV-1 transzkriptom²⁵. Az ONT MinION szekvenátorát alkalmaztuk az EHV-1 genomról leíró RNS-ek komplexitásának feltárásához. Az ábrán a LoRTIA program által azonosított transzkriptok kerültek feltüntetésre. Csak azok a nagyon hosszú RNS molekulák kerültek ábrázolásra, amelyeket a LoRTIA programcsomaggal azonosítottunk a dcDNS szekvenálás adatait felhasználva. A nagyon ritka transzkriptok esetében, ahol a TSS megállapításra nem volt lehetőség, csillaggal jeleztük, hogy a TSS-e csak becült pozíció. A színárnyalatok abundanciára utalnak. Az ábra a Geneious programmal készült.



20. ábra A KSHV transzkriptom²⁶. A KSHV transzkriptomot ONT MinION, valamint Illumina CAGE-Seq módszerekkel térképeztük fel. A kanonikus mRNS-ek az adott ORF-eknek a legabundánsabb RNS izoformái. A színárnyalatok a relatív gyakoriságra utalnak: 1: 1–9 leolvasás (read), 2: 20–49 read, 3: 50–199 read, 4: 200–999 read, 5: >1,000 read.

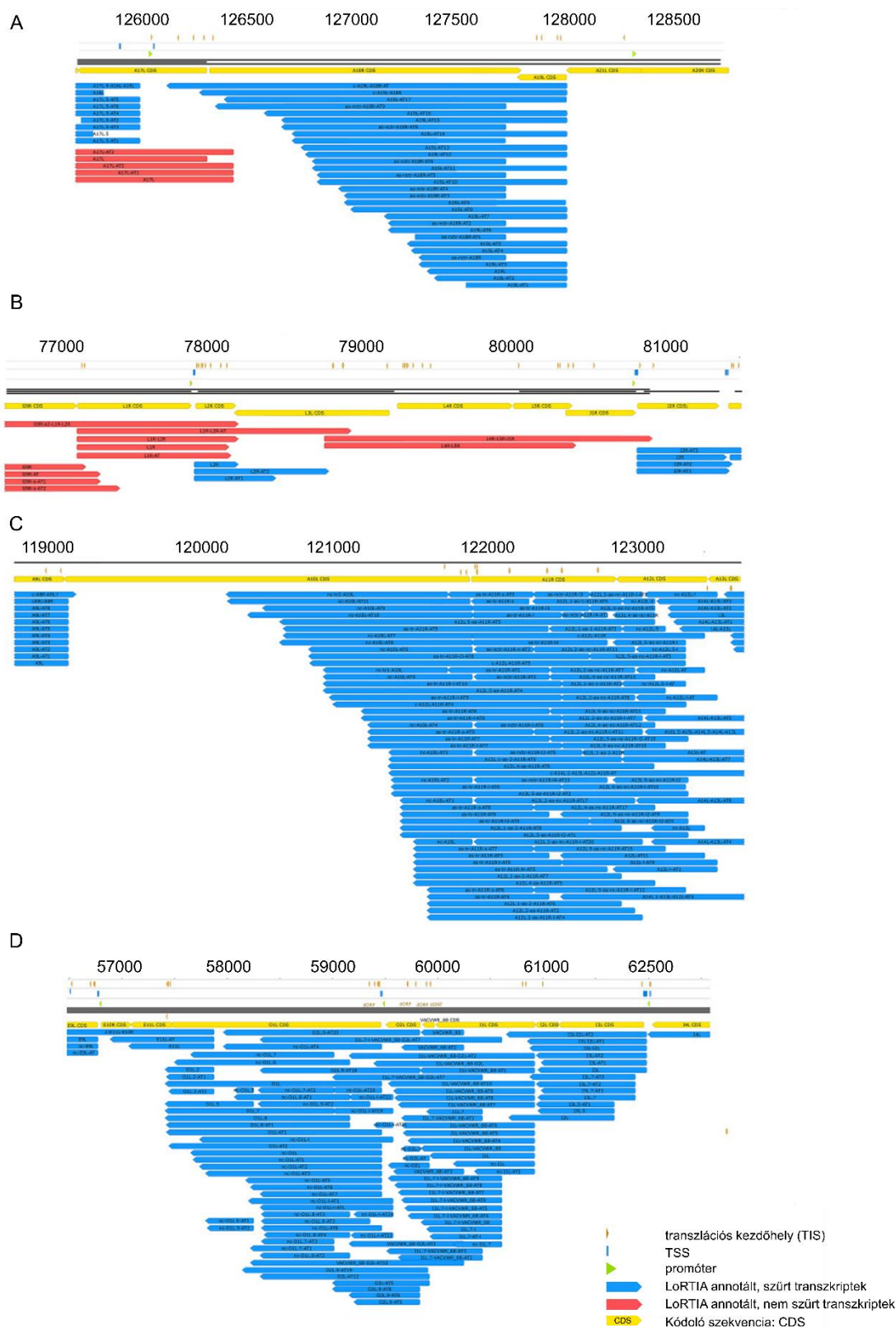
6.3.2.3 Egyéb vírusok

A herpeszvírusoknál tapasztalt összgenomszintű expresszió, a leíró RNS-ek rendkívüli sokfélesége az összes általunk vizsgált DNS genomú vírusra igaz, így még a legkisebb, alig 1,8 kb-os genomú PCV-1 esetén is fedeztünk fel számos új RNS-t, illetve TI-t (**21. ábra**).



21. ábra **PCV-1 transzkriptom**²⁹². A PCV-1 cirkuláris genomon a korábban ismert és az általunk leírt új transzkriptok pozíciója látható. Narancssárga nyilak: ORF-ek; szürke nyilak: már ismert RNS-ek; kék nyilak: új splice izoformák, hosszvariánsok, valamint hosszú, komplex transzkriptok.

A **22-es ábra** a VACV komplexitását mutatja, bizonyos genomi régióra fókuszáltnak. E vírusról ismert volt, hogy a közel 200 kbp-os genom középső régiójában elsősorban az L gének vannak, s hogy ezek változatos végződésűek, azonban a korábban alkalmazott technikák (CAGE, ribosome profiling, NGS) nem tették lehetővé a konkrét transzkriptok azonosítását.

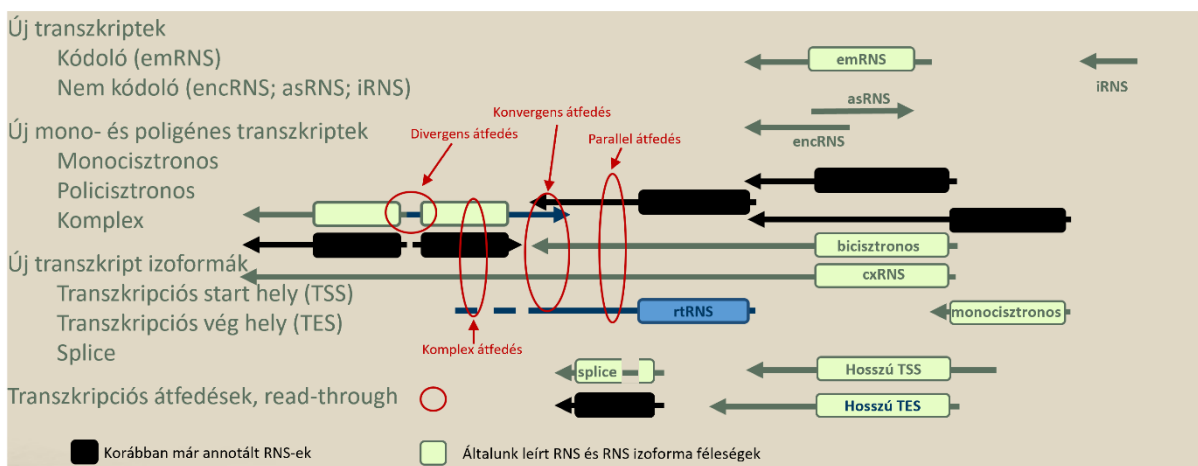


22. ábra A VACV 'kaotikus régiója' komplexitása¹⁷. Ezt az extrém komplexitást (A) antiszensz RNS-ek pl. az A18R lokuszban; (B) az L gének ORF-jein belüli TSS-ek és TES-ek extrém nagy száma; (C) valamint az asRNS-ek mellett további nem kódoló transzkriptek – pl. az A10L-A12L régióban –; biztosítják. (D) Az O1L–I3L genomi szegmensen is extrém nagyszámú ncRNS és TI került azonosításra. A transzkripteket mind a LoRTIA programmal azonosítottuk. Az ábra a Geneious programmal készült.

Az LRS technikák alkalmazása nyújt erre megoldást. Mind a PacBio RSII és Sequel platformjait, mind az ONT MinION szekvenátorát használtuk, többféle könyvtárkészítési módszert alkalmaztunk: pl. PacBio méreletszelektált és nem méreletszelektált, poliA-szelektált és random primingolt IsoSeq módszert; ONT poliA-szelektált, poliA + Cap-szelektált cDNS- és dRNS-szekvenálást alkalmaztunk a minél pontosabb analízis érdekében. A VACV transzkriptom a herpeszvírusokéhoz képest is összetettebb; többszörös transzkripciós átírásokkal teli, amelyek összetett, átfedő RNS-hálózatot hoznak létre. A TES-ek nagyon változatosak, s jellemző az alternatív promóterek széleskörű használata is, mely nagyfokú diverzitást eredményez a TSS-ek terén. Bevezettük a 'szabályos' (regular) és a 'kaotikus' (chaotic) régiók fogalmát. A szabályos genomi régiókból leíró RNS-ek viszonylag pontos TSS-ekkel és TES-ekkel rendelkeznek (herpeszvírus-szerű). Azonban a „kaotikus” helyeken a transzkriptek nagy heterogenitást mutatnak a TSS-ek és TES-ek pontos pozícióit illetően. A fertőzés késői időpontjaiban az I és L gének által kódolt RNS-ek rendkívül változatos TSS-ekkel és TES-ekkel rendelkeznek (például az O1L, O2L, I1L, I2L, I3L, A10L, A11R, A12L, A18R gének), ami jelentős nehézséget okoz az egyes transzkriptek azonosításában (**22. ábra**).

6.3.3 Ami a strukturális komplexitás mögött rejlik

Az előző fejezetben bemutattam, hogy az egyes vírusok RNS profilja rendkívül komplex, hogy egy-egy génről nem csupán egy, hanem esetenként tucatnyi RNS-féleség íródik le. Mindezt kimutattuk számos herpeszvírus esetén^{6,7,18,25,26}, továbbá poxvírusoknál^{17,274,291,297}, az afrikai sertéspestis vírusánál^{20,276}, egy rovarvírusnál^{23,273,277}, cirko²⁹²- és retrovírusoknál²⁹³. 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 transzkriptek, az átfedő transzkriptek és RNS-izofomák, 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, leírtunk. Szintén leírtunk egy speciális transzkript osztályt, melyet replikációs origó-asszociált (ra)RNS-eknek^{82,262,395,398} nevezünk. Leírtunk számos új transzkripciós átfedést mind a szomszédos (lehet parallel, 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). Ezek a jelenségek a **23. ábrán** láthatóak, a következő szekciókban pedig részletesen bemutatom az összes kategóriát.



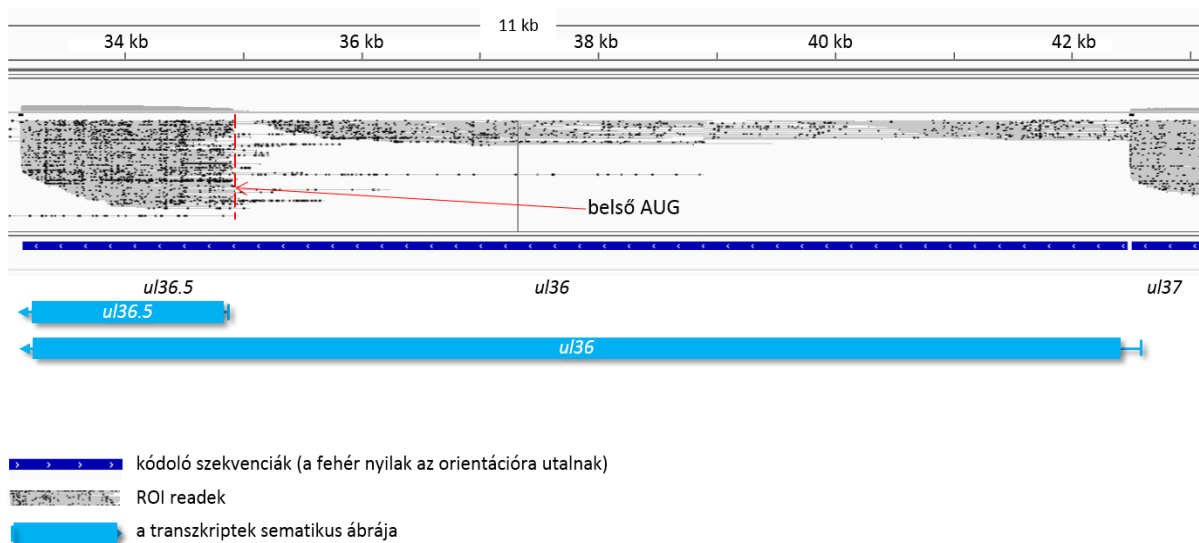
23. ábra Az általunk leírt transzkript féleségek és a kanonikus RNS-ek sematikus ábrája. A feketével jelölt, már korábban annotált RNS-ekhez képest számos új transzkript és transzkript izoforma variánst írtunk le, ezeket zölddel ábrázoltam. Az ábra az egyes transzkript hosszvariánsok következtében kialakult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

6.3.3.1 Új kódoló gének, emRNS-ek

Az LRS technikák rendkívül hatékonyak azoknak az 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-terminálisban 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^{6,8,12,13}, valamint egy bakulovírusban is²⁷³. A **19-es**, a teljes EHV-1 transzkriptomot bemutató **ábrán** zöld színnel jelöltük az eRNS-eket. A következőkben bemutatok néhány emRNS példát:

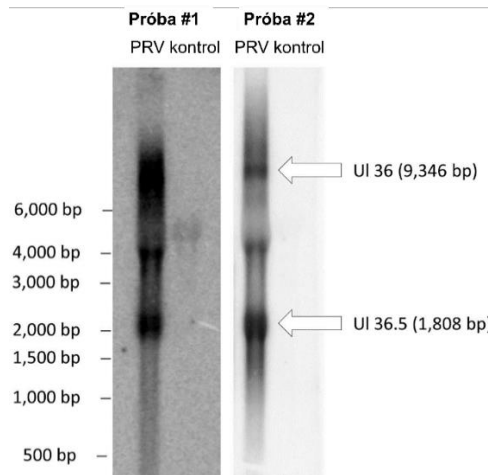
a. PRV *ul36.5*

A PRV esetén (csakúgy, mint az összes herpeszvírusnál), az emRNS-ek száma magas, abundanciájuk változó. Leírtunk egy, az *ul36* génben azonosított, nagymértékben expresszálandó, potenciális fehérje kódoló gént, az *ul36.5*-öt⁵. Az azonosított iFORF az UL36 fehérje rövidebb verzióját kódolhatja. Ez a transzkript alternatív transzkripció iniciáció révén íródik le, s egy 1808 bp-os hosszú, nested, beágyazott RNS-t eredményez az általunk meghatározott KJ717942.1 genom 34934-33127-ig terjedő régiójában (a koordinátákat orientáció alapján: 5' -> 3' adtuk meg, **24. ábra**). Az *ul36* és az *ul36.5* transzkriptek 3'-vége közös, koterminálisban végződnek. Az *ul36.5* RNS-ről potenciálisan leolvasódó fehérje 467 aminosavból áll. Putatív TATA boxot a PRV genom 34959–34964 közötti régiójában azonosítottunk.



24. ábra Az *ul36.5* – új, potenciális fehérje kódoló gén⁶. Az ábrán a PRV egy 11 kbp-os genomi régiója, s az ide térképeződő PacBio RSII readek IGV vizualizációja, valamint sematikus reprezentációjuk látható.

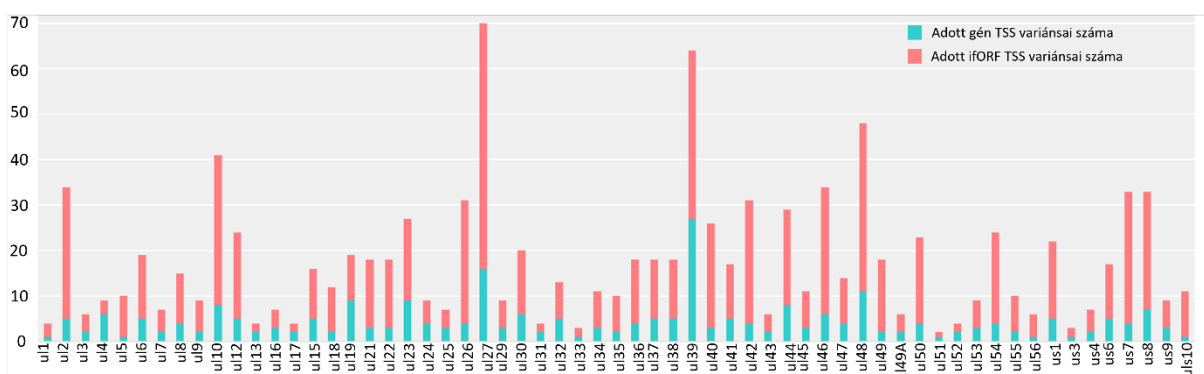
Az *ul36.5* transzkriptet az LRS módszer mellett Northern blottal is kimutattuk (**25. ábra**).



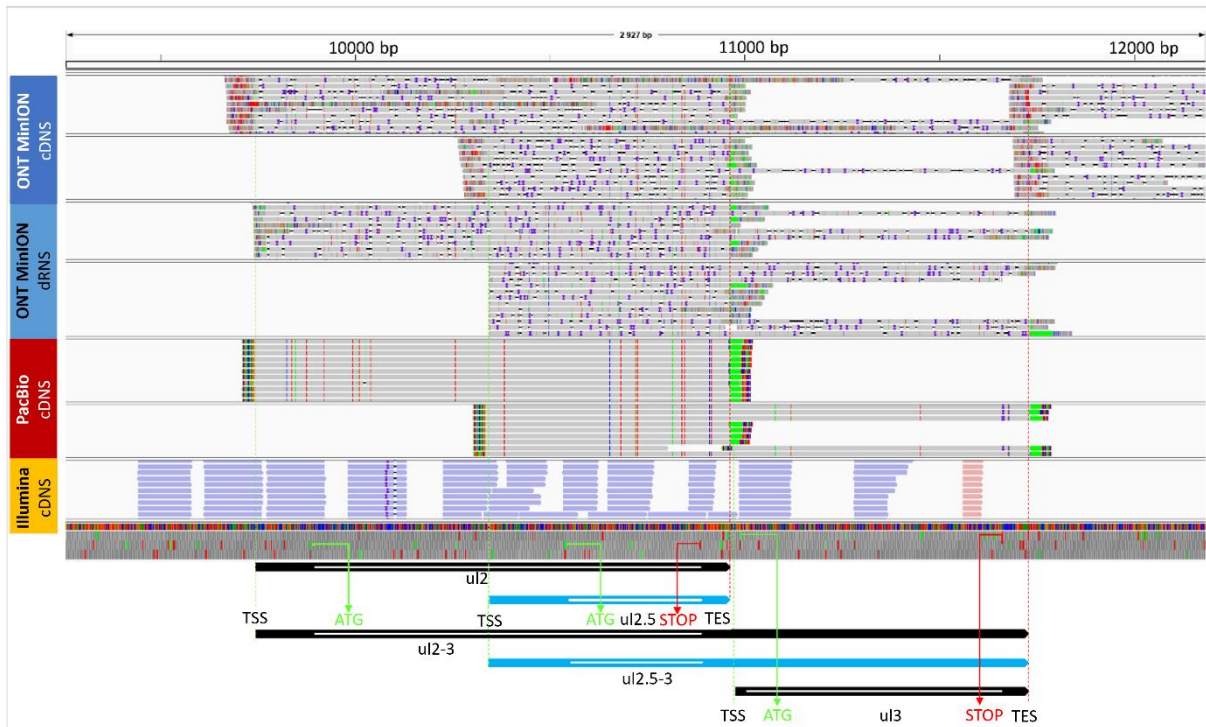
25. ábra Az *ul36.5* transzkript kimutatása Northern blot analízissel⁶. Két, egymást nem átfedő, radioaktívan jelölt DNS-fragmentumot használtunk próbaként, mindkettő az *ul36.5* génen belül helyezkedik el. Mindkét próbával három fragmenst kaptunk: az első megfelel a 9,436 bp hosszú *ul36* transzkriptnek; a második megfelel az 1,808 bp hosszú *ul36.5* transzkriptnek és egy ~ 4,000 bp hosszú fragmens, ami egy lehetséges új, alacsony abundanciájú transzkriptre utalhat. Kontrollként PK-15 sejtekből izolált RNS-t használtunk.

b. A HSV emRNS-ek

A HSV-1 transzkriptom minél precízebb annotálása érdekében a saját PacBio RSII, Sequel, valamint ONT cDNS, dRNS és cap-szelektált szekvenálásunk mellett bevontunk az analízisbe minden elérhető HSV-1 transzkriptom adatot, melyek egy kivétellel mind Illumina SRS szekvenálások voltak³⁶⁷⁻³⁷⁰. Az egy kivétel egy extrém nagy lefedettségű ONT dRNS szekvenálás volt³⁴⁴. Ez a metaanalízis-szerű tanulmányunk²⁸⁹ alátámasztotta, hogy gyakorlatilag az összes HSV-1 gén tartalmaz legalább egy rövidebb, 5' végén csonkolt variánst, iFORF-fel (**26. ábra**). Ezeknek az emRNS molekuláknak a funkcióját nem vizsgáltuk. A HSV-1-ben a saját eredményeink alapján 63 emRNS-t azonosítottunk, ezek közül a Depledge és munkatársai által publikált³⁴⁴ ONT dRNS adatszettel 61-et validáltunk. Egy ilyen validált iFORF-et tartalmazó RNS-t és az ezt megerősítő különféle technikákból származó leolvasásokat a **27. ábrán** szemléltetem.



26. ábra A HSV-1 emRNS-ek²⁸⁹. Az oszlopdiaagram az egyes génekhez tartozó beágyazott ORF-ek számát, illetve az ezekhez tartozó transzkript hosszvariánsok számát mutatja.



27. ábra A HSV-1 *ul2.5* és az *ul2.5-3* eRNS-ek²⁸⁹. 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.

c. Beágyazott gének a bakulovírus genomban

A bakulovírus transzkriptom elemzése során négy új, potenciálisan fehérje kódoló gént azonosítottunk, mindnek közös PA szignálja van a 'host' génjével. A négy közül három putatív mRNS (BV/ODV-C42.5, V-CHAT.5, P80.5) a kanonikus LIS TAAG szekvencia második nukleotidjától indul, ugyanaz az iniciátoruk, mint a nem-csonkított transzkriptokké, mely arra utal, hogy ezek is L transzkriptek. Ezt a kanonikus TATA promóterek hiánya is indikálja. A POLH.5 TSS-e egy ACAGG motívumban található, ami hasonlít a korábban már leírt arthropoda iniciátor szekvenciához, az ACAGT-hez^{371,372}, kivéve, hogy ezt nem egy kanonikus TATA motívum előzi meg. Ennek az iniciátornak a jelenléte jelzi, hogy egy 819 bp hosszú, 5'-ről csonkított POLH-t a gazda RNS-polimeráza írhatja le. A TES-ekkel kapcsolatban, minden új feltételezett fehérjekódoló transzkript tartalmaz egy vagy több PAS-t a 3' végüktől upstream, átlagosan 16,2 bp távolságra.

6.3.3.2 Nem-kódoló RNS-ek

Sokáig élt az az elképzelés, hogy a DNS és a fehérjék játszanak alapvető szerepet a különféle sejtfunciók szabályozásában, míg az RNS-ekről úgy vélték, hogy csak kiegészítő szereppel bírnak. Az utóbbi egy-másfél évtized transzkriptomikai tanulmányai azonban hatalmas számú, változatos ncRNS-t írtak le, melyek széleskörű funkciókkal rendelkeznek, szerepet játszanak többek között az epigenetikai, transzkripciós és poszt-transzkripciós génkifejeződés szabályozásában³⁷³, továbbá egyre többet tudunk az ncRNS-ek DNS replikáció szabályozásban betöltött szerepéről is, a rövid miRNS-ektől³⁷⁴ az lncRNS-ekig³⁷⁵.

Az ncRNS-eket, amelyeket a vírusokban azonosítottunk, az alábbiaknak megfelelően, helyük és orientációjuk alapján kategorizáltuk (**23. ábra**):

a. Intergenikus transzkriptek

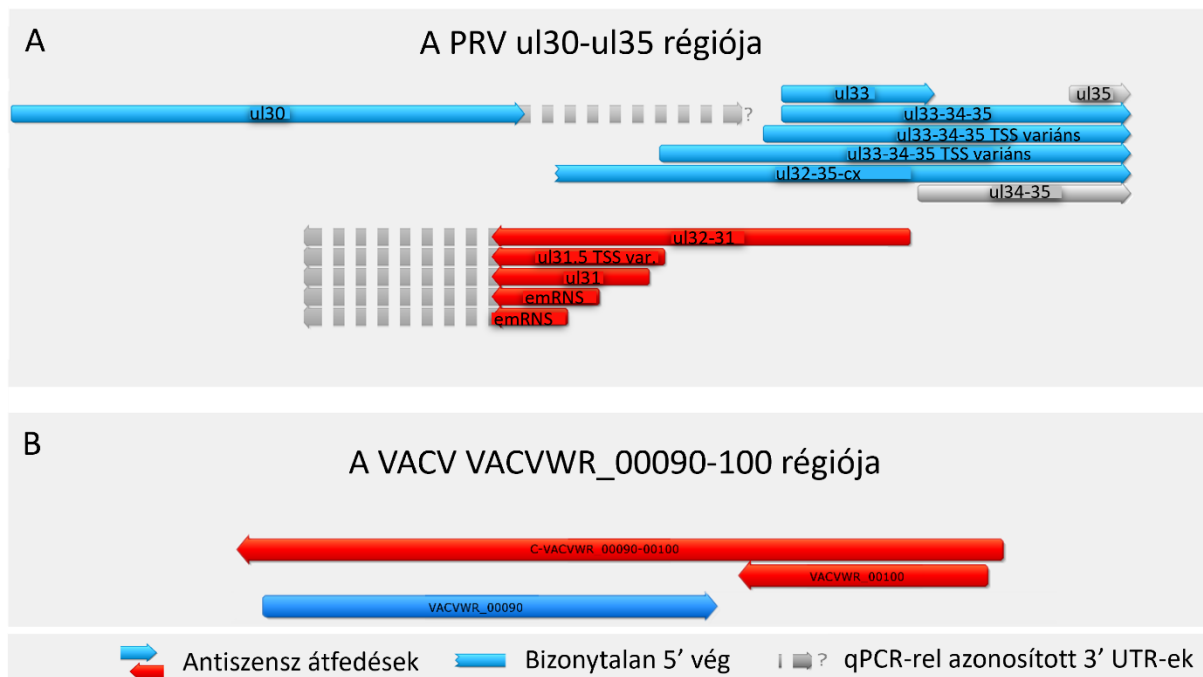
Ezek az RNS molekulák két kódoló szekvencia között helyezkednek el, nincs átfedés közöttük. LRS módszerekkel azonosításra került számos iRNS, habár az ilyen típusú transzkriptek azonosításában az LRS technikák alkalmazása nem jár különösebb előnnyel az SRS-hez képest.

b. Beágyazott, nem kódoló RNS-ek

Az encRNS-ek mRNS-ekben, vagy ncRNS-ekben is elhelyezkedhetnek. A legtipikusabb mRNS-átfedő encRNS-ek az 5'- és 3' végeken csontolt transzkriptek. Az encRNS-eknek általában közös TSS-e, vagy TES-e van más RNS-ekkel, amelyekkel átfednek.

c. Antiszensz transzkriptek

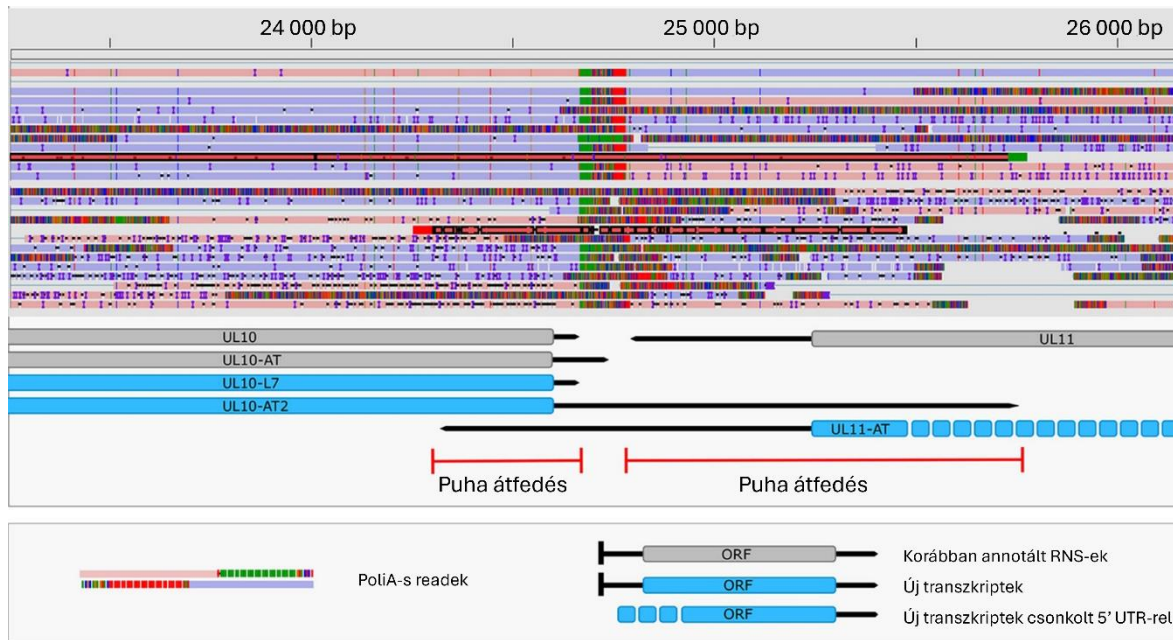
Az asRNS-ek vagy teljes mértékben, vagy részben átfednek egy mRNS-t, azzal ellentétes (antiparallel) állásban. Ezek az RNS-ek lehetnek akár saját promóter által szabályozottak, vagy pedig lehetnek a szomszédos, vagy akár távoli génekkel való konvergens, vagy divergens transzkripciók átfedés eredményei (**28. ábra**). Utóbbi esetben a transzkriptnek csak az átfedésben álló szakasza az antiszensz RNS.



28. ábra **Antiszensz RNS-ek a PRV-ben és a VACV-ban**²⁶². (A) A PRV ul30-35-ös genomi régiójáról számos RNS és TI íródik le, melyek átfednek egymással konvergens (pl. ul30 és ul31) vagy divergens (pl. ul32-31 és ul33) módon, bizonyos szakaszaik így antiszenszei egymásnak. Adott transzkript qPCR-rel detektált hosszabb 3' UTR variánsai még hosszabb átfedésre utalnak (szürke szaggatott nyilak). (B) Konvergensen átfedő asRNS-ek a VACVWR-00090-00100 genomi régióban.

6.3.3.3 Read-through RNS-ek

A read-through (rt)RNS-ek kódoló, vagy nem kódoló génekről is leíródhatnak, amikor a transzkripció terminációs szekvenciákat az RNP komplexek nem ismerik fel, s így a transzkripció nem áll meg, hanem transzkripció read-through (TRT, azaz továbbírás) történik (29. ábra). Amennyiben az rtRNS-nek van pontos terminációs helye és ugyanazt az ORF-et tartalmazza, mint a rövidebb, kanonikus TI, akkor az rtRNS egy TES izoforma. A cxRNS-ek, melyek ellentétes orientációjú ORF-eket is leírnak, szintén a TRT-k eredményei, míg a poligén RNS-ek csak akkor tekinthetők rtRNS-eknek, ha az upstream géneknek is saját transzkripció terminációs szignáljaik vannak.



29. ábra A HSV-1 transzkriptom részlete²⁸⁸. A HSV-1 transzkriptom finomhangolását a PacBio Sequel és az ONT MinION platformjaival végeztük el. Az ul10-ul11 konvergens génpár között 'puha' átfedést detektáltunk, amit a mindkét irányból leíró rtRNS-ek biztosítanak.

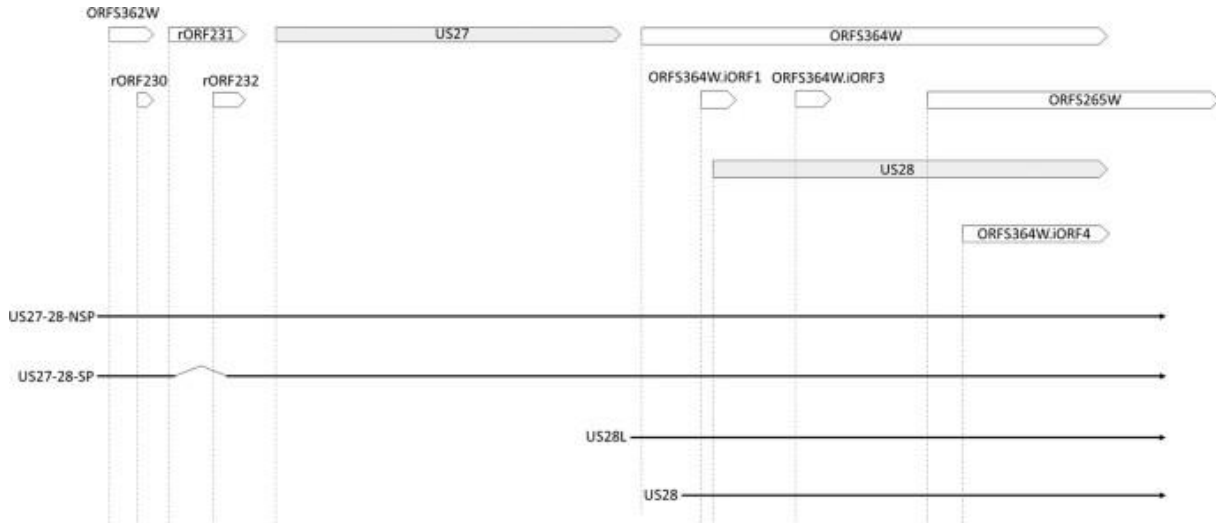
6.3.3.4 RNS izoformák

a. Transzkripció start hely (TSS) variánsok

A gének alternatív promóterek által szabályozhatóak, melyek eltérő TSS izoformákat fejtenek ki³⁷⁶. A több promóterrel szabályozott gének a vírus életciklusa során eltérő mértékben expresszálódhatnak³⁷⁷. Ezenkívül a TSS izoformáknak különböző funkcióik lehetnek, mivel a különböző 5'-UTR (nem kódoló régió) szerkezetek különböző módon szabályozhatják a translációt³⁷⁸. A TSS izoformák eredményezhetnek rövidebb transzkriptet, mint a kanonikus, ez esetenként csak rövidebb 5'-UTR-t eredményez, más esetben eRNS-t. A TSS izoformák egész hosszú 5'-UTR-eket is biztosíthatnak a géneknek. Minél hosszabb az UTR, annál nagyobb a valószínűsége, hogy upstream ORF-eket (uORF) tartalmaz.

Az mRNS-ek 5'-UTR-jében található uORF-ek a cisz-szabályozó elemek egy osztályát képezik, és egy alternatív translációs szabályozási módot képviselnek³⁷⁹. Az uORF-ek rövidek (<30 kodon); általában gátolják a downstream található fő ORF-ek translációját, azonban képesek lehetnek azok leíródásának serkentésére is³⁸⁰. Az ugyanazon gén által kódolt transzkript variánsok uORF összetétele (pl. rövid vagy hosszú TSS variánsokban) modulálja azok translációs szabályozását³⁸¹.

A HCMV transzkriptom kódoló kapacitása analizéséhez az experimentálisan validált ORF listát³⁸² használtuk. Az általunk azonosított különféle TSS TI-k többsége esetén azt kaptuk, hogy egyedi kódoló potenciáljuk van, azaz egyedi ORF kombinációkat tartalmaznak. A TSS-izoformák eltérnek az uORF-ek számában, amelyek a fő ORF-től upstream állnak, az alternatívan spliceolt izoformák és a TSS-izoformák eltérően rövidített változatait tartalmazták ugyanannak a fő ORF-nek, vagy több fő ORF-et is tartalmaztak (**30. ábra**).



30. ábra A transzkript izoformák eltérő ORF-eket tartalmaznak⁹. Az ábrán felül az ORF-eket, alul a TI-eket ábráztuk. A szürke nyilak a kanonikus ORF-ek, a fehér színűek a translációban aktív ORF-ek, melyeket Stern-Ginossar és munkatársai publikáltak³⁸². A függőleges pontozott vonalakkal a TSS-eket jelöltük. Az us27 és us28 gének a poligén szerkezet (US27-28 vagy csak az US28), az alternatív splicing (US27), vagy az alternatív transzkripció miatt különféleképpen olvasódnak le.

Bár az α HV-oknál is jellemzőek a TSS izoformák, gyakoriságuk és hosszuk azonban elmarad a HCMV-től, illetve egyéb β - és γ HV-októl, míg a pox-, bakulo- és afrikai sertéspestis vírusokhoz képest kevésbé diverzek. Utóbbiakra jellemző inkább a nagyobb mértékű heterogenitás mind a TSS-ek, mind a TES-ek tekintetében. Míg azonban az AcMNPV-nél²³ és az ASFV-nél²⁰ is azonosítottunk fix, abundáns TSS-eket, melyek uORF-eket foglalnak magukba, addig ezt a VACV-nál nem figyeltük meg. Ezt a nagyon sok genomi pozícióból induló transzkripció és az annotált transzkriptek (a többi, általunk vizsgált víruscsaládhoz képest mért) rövidegsége is magyarázza.

b. Transzkripció stop (TES) variánsok

Bizonyos esetekben a ribonukleoprotein komplexek (RNP-k) progressziója nem áll meg a transzkripció terminációs szekvenciáinál, ami hosszabb TES variánsokhoz vezet^{383,384}. Sok esetben ezek a hosszabb molekulák átfednek a downstream génekről leíró transzkriptekkel. Korábban az SRS-t használták az alternatív poliadeniláció azonosítására³⁸⁵. Az LRS azonban további információkat ad a különböző transzkript izoformák által használt különféle 5'- és 3'-UTR kombinációról. Kimutattuk, hogy a poliadenilációs helyeket szűrni kell az 'internal priming' [oligo(dT) primerek bekötődése aspecifikusan] tekintetében, miután az adeninben gazdag régiók fals poliadenilációs helyekként jelenhetnek meg^{8,386}.

Talán a legnagyobb kihívást az AT-gazdag (66,7 %) genommal rendelkező VACV jelentette a transzkriptek annotációja szempontjából. Ennek fő oka az extrém TSS és TES gazdagság (**22. ábra**).

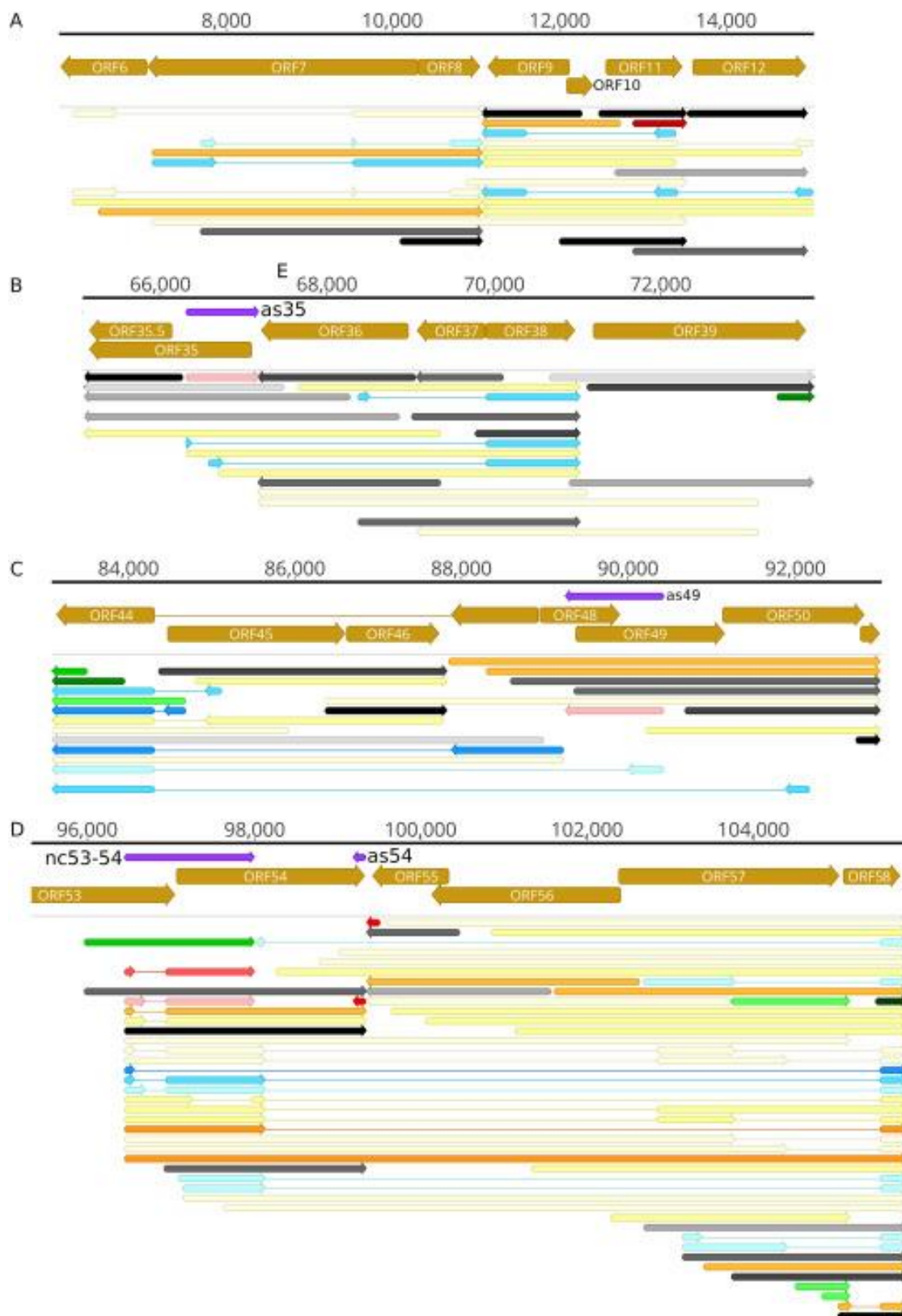
c. Splice variánsok

A PacBio Iso-Seq, illetve az ONT dRNS-Seq technika különösen alkalmas az új splice helyek azonosítására^{6,8,21,25,26,387-389}. A splicing események viszonylag ritkák az alfa-herpeszvírusokban, bakulovírusokban és orthomyxovírusokban^{17,274,390}, de gyakoriak a béta- és gammaherpeszvírusokban, retrovírusokban és hepadnavírusokban^{85,293,391,392}, míg a poxvírusokban egyáltalán nincs splicing¹⁴⁴. A LRS technikák alkalmazásának eredményeként radikálisan megnőtt az azonosított splice helyek és splice izoformák száma.

Az α HV-oknál is vannak ismert, spliceolt transzkriptek (pl. a PRV US1, EP0, UL15, vagy az utóbbival homológ ORF44 az EHV-1-ben), s ezeknek számos új variánsát detektáltuk a kanonikus változatok mellett^{18,25}. Az általunk a PRV-ben azonosított NOIR-1 ncRNS egy intronos, spliceolt változatát is azonosítottuk, s ennek kimutattuk a homológját az EHV-1-ben is, mely azonban két splice hellyel rendelkezik²⁵. Az AcMNPV-ben a már korábban leírt¹⁷¹ tizenkét, kis abundanciájú spliceolt transzkript mellett további ötöt azonosítottunk, míg poxvírusokban, a rokon ASFV-ben (ahogy az várható is volt) nem detektáltunk egyetlen spliceolt transzkriptet sem. Ez utóbbi tény az LRS módszer, az általunk alkalmazott szigorú kritériumok a splicing elemzése során, s a LoRTIA program együttesének megbízhatóságára is utal, hiszen az RT és a PCR template-switching révén hibákat okozhat a cDNS-ben, mely fals intronnak tűnhet, s hibás annotációhoz vezethet. A LoRTIA azonban kiküszöböli ezeket a műtermékeket azáltal, hogy észleli a splice junction konszenzusok hiányát, vagy az repeat régiók jelenlétét, amelyek elősegíthetik a TS-t.

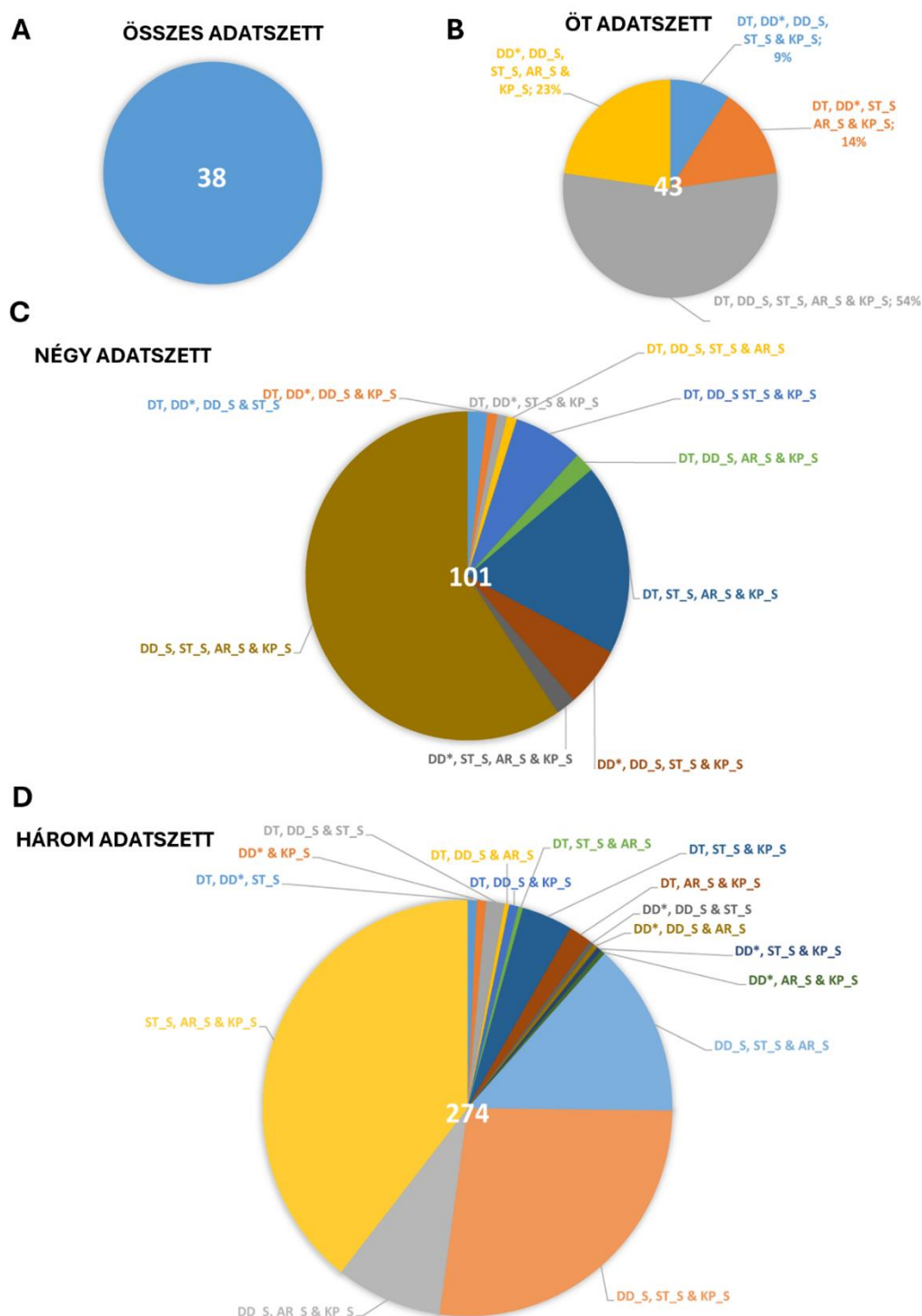
Az alternatív TSS-ek mellett az alternatív splicing is képes megváltozott kódoló potenciálú transzkripteket előállítani, ha a splicing az ORF-en belül történik. A splicing fúziós transzkripteket (fRNS) is eredményezhet, melyek legalább két szomszédos vagy közeli génből származó splice-olt szekvenciát képviselnek, UTR-eket és/vagy kódoló régiókat foglalnak magukba. Az ORF fúziókban lévő downstream partner lehet in-frame vagy kereten kívüli, azaz out of frame.

Az EHV-1 esetében összetett splicing mintázatokat észleltünk még azokban az RNS-ekben is, amelyek a rokon vírusokban nem spliceoltak (**31. ábra**). A legérdekesebbek ezek közül az fRNS-ek. Ezek egyik típusa ellentétes orientációjú gének szegmenseit használja fel (pl. ORF8, **31. ábra A**). Ezeknek a transzkripteknek a pre-mRNS-ei cxRNS-ek. Az ORF8 (*u151*) az ORF6 és ORF7 bizonyos részeit használja fel 5'-UTR-ként különböző kombinációkban. Érdekesség, hogy az ORF8 egy rövidebb, kódoló szekvenciája is leíródik. Hasonló összetett splicing mintázat figyelhető meg az ORF9 (*u150*) esetében is. Az ORF35-38 régióban (**31. ábra B**) kifejezett fúziós RNS-ek első három génje antiparallel orientációjú az ORF38-hoz képest. Ezek a transzkriptek tartalmazzák az ORF38 teljes kódoló régióját és különböző 5'-UTR szegmenseket az ORF35 és ORF36 génekből. Az ORF44 (*u15*) gén az α HV-ok egy különleges génje; az ORF folytonosságát két másik, ellentétes orientációjú gén (ORF45/*u17* és ORF46/*u16*) szakítja meg, amelyek kivágódnak az érett ORF44 RNS-ből. A downstream exon függetlenül is expresszálódik (**31. ábra C**). Az EHV-1 ORF44 még bonyolultabb: sok RNS egy sokkal hosszabb intront tartalmaz, amely magában foglalja az egész ORF48 és 49 géneket, valamint az ORF50 gén nagy részét. Az ORF53-58 (*u19-4*) régió legösszetettebb (**31. ábra D**), a gének különböző kombinációkban fúziós fehérjéket termelnek.



31. ábra **Spliceolt és fúziós transzkriptek**²⁵. Az ábrán látható transzkriptek mindegyike a drNS-Seq adatokban is jelen volt. A. ORF6-12. B. ORF35-38. C. ORF44-50. D. ORF53-58. A színek árnyalata a gyakoriság szerint van meghatározva: 1: 1–9 leolvasás (read), 2: 10–49 read, 3: 50–199 read, 4: 200–999 read, 5: >1000 read. Fekete: mRNS; piros: ncRNS; zöld: putatív emRNS; sárga: csak dcDNS adatokban volt jelen; kék: fúziós transzkript.

Saját, ONT és PacBio adataink alapján összesen 378 putatív intront azonosítottunk a HSV-1-ben. A splice helyek azonosításához egy meta-analízis-szerű elemzést folytattunk le, melybe a saját adataink mellett az összes, adatbankokban elérhető HSV-1 transzkriptomot felhasználtuk (32. ábra).



32. ábra A HSV-1 legalább három, független kísérletben azonosított intronjai²⁸⁹. (A) Harmincyolc intront mind a hat adatszettben megtaláltunk. (B) Öt kísérlettel validált intronok. A legnagyobb "öt adatszettes" kombináció (56%) azokat az intronokat tartalmazza, melyek a négy Illumina adatszettben és a sajátunkban került azonosításra. (C) Négy adatszett által alátámasztott intronok száma. Ezek 59%-át a négy Illumina adatszettben azonosítottuk. (D) 274 intron három független kísérletben volt jelen. DT^{7,288}; DD³⁴⁴; ST³⁶⁷; AW³⁶⁹; AR_S³⁶⁸ analysed by STAR; DD_S³⁴⁴ analysed by STAR; KP_S³⁷⁰; ST_S³⁶⁷ analysed by STAR. A betűk az adott publikáció első szerzői monogramjaira utalnak.

A BoHV-1 transzkriptom elemzése során olyan magas abundanciájú fúziós transzkripteket azonosítottunk, amelyek tartalmazzák a *bicp4* gén 5'-UTR régióját és a teljes *bicp0* gént, de a *bicp4* gén kódoló része ki van vágódva (**33. ábra**), ami egy korábbi publikáció eredményeit megerősíti³⁹³. Amikor a *bicp4* gén második kópiáját vizsgáltuk, hasonló fúziós transzkripteket találtunk, amelyek a *bicp4* 5'-UTR-jét és a *circ* gén teljes egészét tartalmazzák. Ezek a kiméra bICP4-CIRC transzkriptek átívelnek a cirkuláris virális DNS genomi határain, tartalmaznak egy nagy intront, amely magába foglalja a *bicp4* ORF-et.



33. ábra Az IE gének genomi régióiról leíró BoHV-1 transzkriptek²⁸⁴. A három IE gén nagy változatosságú splice-olt és nem splice-olt transzkripteket állít elő. A *bicp4* gén 3'-csonkolt nem kódoló RNS-eket és alternatív TES-eket is kódol. A *bicp0* transzkriptek jelentős részének kifejeződését az *icp4* promóter szabályozza. A *bicp22* gén mutatja a legösszetettebb expressziós mintázatot a BoHV-1 gének között, a több splice hely, TSS és TES használata, valamint nagyon hosszú transzkripciós átírások miatt. A TATA box2 a kanonikus transzkriptek alap promótere, míg a TATA box1 az OriS-en belül található. A TAATCGAT szekvenciáról feltételezik, hogy a promóter VP16 kötőhelye. A transzkript abundanciákat színárnyalatokkal jelöltük az ábrán: a két szélsőséggént a fekete nyilak magas abundanciát, míg a fehér nyilak alacsony abundanciájú transzkripteket jeleznek.

Bicisztronos ICP4-CIRC transzkripteket is azonosítottunk, amelyek a teljes *bicp4* és *circ* géneket tartalmazzák. Nem tudjuk, hogy ezek érés előtti vagy érett RNS-ek. Úgy gondoljuk, hogy a BoHV-1 *circ* gén IE kifejeződési jellemzőjét a *bicp4* promóter, az IE transzkripciós egység 1 (IEtu1)³⁹³ biztosítja. A *bicp4* promóter TAATGAGCT szekvenciáját a tegument VP16 transzaktivátor kötőhelyeként írták le³³⁹. Hasonló szekvenciát (TAATCGAGA) is azonosítottunk a *bicp22* gén távoli promóterében, vagyis IEtu2-ben³⁹⁴. Ezek a herpeszvírus IE gén motívumok, a TAATGARAT-szerű

szekvenciák nem voltak azonosíthatók a többi három IE gén promóterében. Így a négy IE gén közül a bICP0-át, bICP4-et és a circet az IEtu1, míg a bICP22-t az IEtu2 szabályozza.

Az EBV, a HCMV és a KSHV transzkriptomok analízise számos új splice izoforma azonosítását eredményezte. Az EBV-nél nagyszámú olyan, új intront írtunk le, mely ORF-en belül helyezkedik el, köztük több olyat, mely kereteltolódást, vagy nonszensz mutációt okoz. A KSHV-nál 10 genomi régiót azonosítottunk, amelyek fúziós transzkripteket kódolnak (**20. ábra**).

6.3.3.5 *Transzkripció átfedés hálózatok*

Eredményeink azt mutatják, hogy gyakorlatilag minden herpeszvírus gén produkál transzkripcionális átfedéseket (TO-kat) úgy mint a divergens (fej-fej melletti), konvergens (farok-farok melletti) és párhuzamos (farok-fej melletti) átfedéseket; az LRS technikák egy bonyolult transzkripció átfedést tártak fel a szomszédos és akár az egymástól távolabbi helyeken lokalizálódó gének között egyaránt. A tandem gének parallel átfedő, többgének, 3'-koterminális transzkripteket alkotnak, amely egységek a herpeszvírusok archetipikus genomikus szerveződését képviselik. Sok vírusgén esetén úgy találtuk, hogy 5' csonkított transzkriptek is leíródnak különböző TSS-ekkel, de ugyanazzal a TES-sel, amelyek ún. 'beágyazott' ORF-eket tartalmaznak, s N-terminálisan rövidített polipeptideket kódolhatnak. A legtöbb divergens gén 'kemény' TO-kat produkál, a kanonikus transzkriptek átfedik egymást. Azonban néhány esetben csak a hosszú TI-k hoznak létre fej-fej melletti TO-kat (ún. 'puha' TO-k), nem pedig a kanonikus RNS. A konvergensen orientált gének 'puha' TO-kat formálnak transzkripció átírás révén, s csak néhány esetben figyeltünk meg 'kemény' TO-kat (pl. az α HVs-ben az *ul7/ul8*, *ul30/ul31* és *ul50/ul51* génpárok esetében) (**34. ábra**).

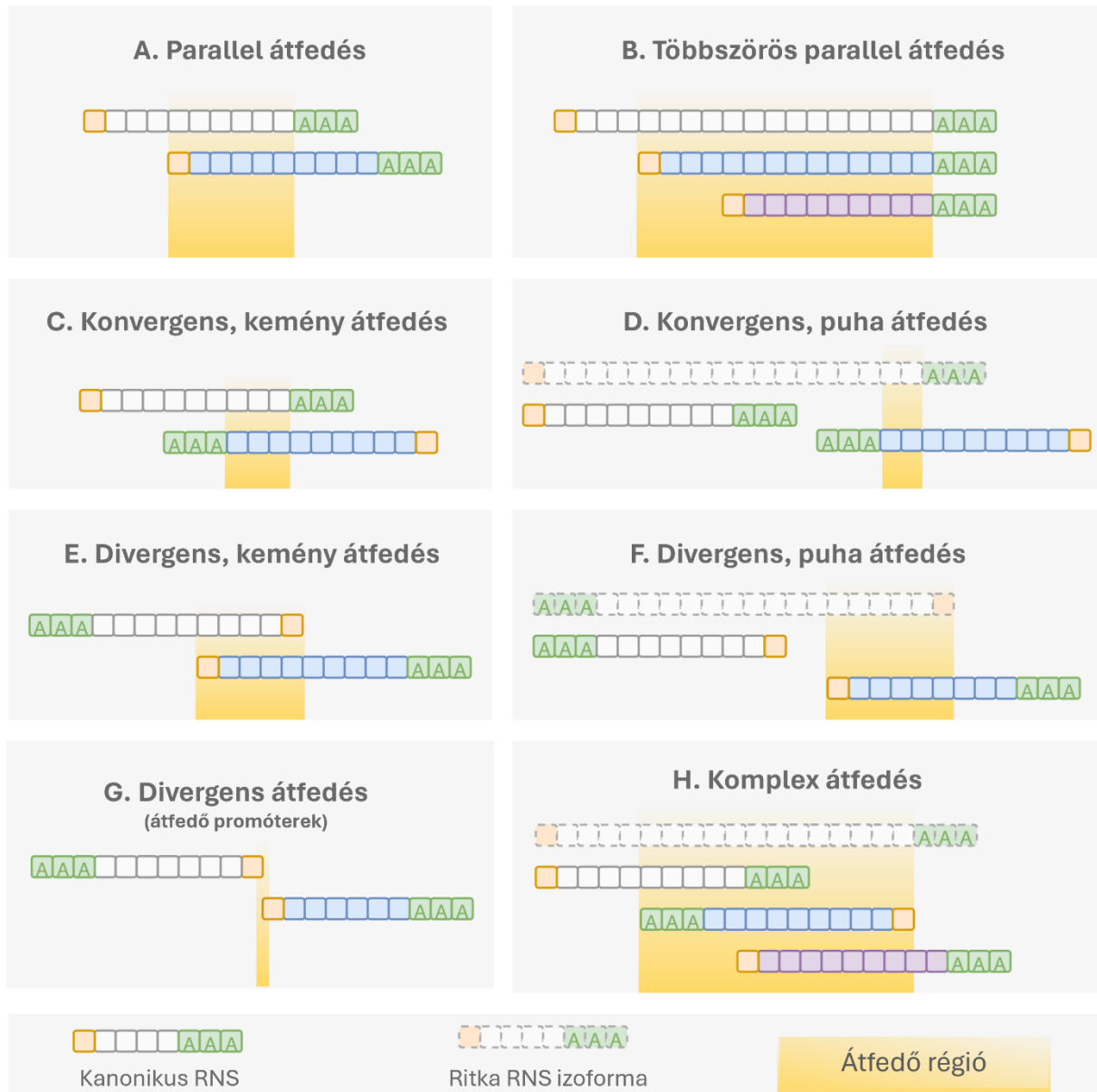
Projekjeink eredményeként jelentősen nőtt a már ismert transzkripció átfedések száma a szomszédos és a távoli gének között egyaránt. Kimutattuk, hogy gyakorlatilag a legtöbb vírusgenom hosszú, átfedő RNS-eket expresszál.

a. Parallel (tandem) átfedések

A tandem orientációjú gének között az átfedés lehet részleges, vagy teljes. Elemzésünk feltárta, hogy gyakran e két jelenség együttesen előfordul ugyanazon génpár között.

b. Konvergens átfedések

A konvergens átfedések az egymással szembenálló, ellentétes orientációjú RNS szintű interakcióit biztosítják. Megkülönböztetünk úgynevezett kemény és puha átfedéseket (hard és soft overlap), előbbiek esetén az érintett génekről leíródo összes RNS átfedésben áll a vele szembenálló génről leíródo transzkriptekkel, míg utóbbi esetben ez csak a transzkriptek hosszabb TES variánsai esetén fordul elő.



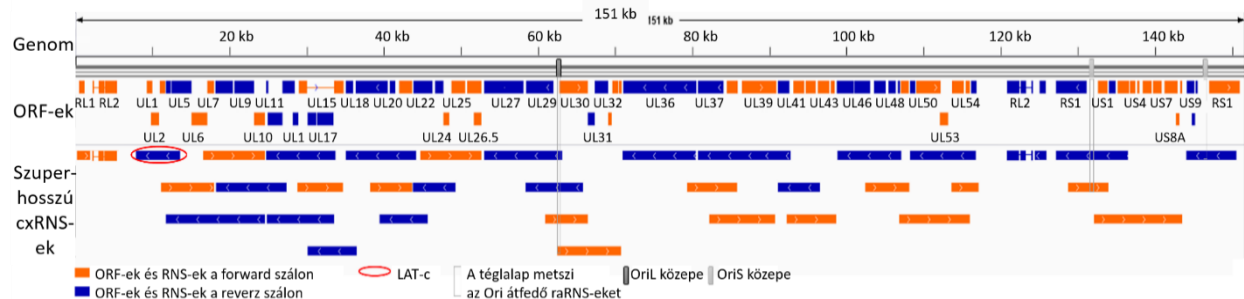
34. ábra **A vírustranszkriptomok analízise során megfigyelt transzkripciós átfedés típusok sematikus ábrája**²⁶². A virális RNS-molekulák különböző típusú átfedéseket képezhetnek az átírások irányát és hosszát illetően. A herpesz-, pox- és a bakulovírusok transzkriptomjának prototipikus szerveződése az, hogy a tandem génjeik átfedő transzkripteket fejzenek ki közös 3'-végekkel. A 'soft' átfedéssel rendelkező transzkripteknek rövidebb nem átfedő változatai vannak (ezek a kanonikus RNS-ek). A komplex átfedések legalább két teljes, ellentétes orientációjú gént foglalnak magukba.

c. Divergens átfedések

A divergens átfedések az egymással ellentétes orientációjú, egymásnak "háttal" álló gének között figyelhetők meg. Több szintű átfedést azonosítottunk: a 'legerősebb' esetben a transzkriptek a szomszédos gén kódoló régióit (KR) is érintik: KR/KR átfedés, ha ez mindkét partnergén esetén fennáll. Más esetben csak az egyik génről képződő RNS fedt át a másik gén kódoló régióját, míg a másik csak az 5'-UTR régióba ér bele: 5'-UTR/KR átfedés. Előfordul, hogy csak a két gén 5'-UTR-jeivel fednek át az érintett génekről expresszálódó RNS-ek (5'-UTR/5'-UTR átfedés), míg bizonyos esetekben nincs transzkript szintű átfedés, ám a transzkripció szabályozásában kulcsszerepet játszó promóter szekvenciák (TATA boxok) átfednek egymással.

d. Átfedések a távoli gének között

A transzkripció átfedések egy további formáját a nem szomszédos, tandem orientációjú gének alakítják, ezeket távoli párhuzamos átfedéseknek nevezünk. A különböző orientációjú géneket tartalmazó transzkriptek (cxRNS-ek) meglepte arra utal, hogy a lokális gének mellett a disztális gének is kölcsönhatásba léphetnek egymással különböző típusú transzkripció átfedések, úgynevezett komplex átfedések kialakításával. A HSV-1-ben ilyen szuperhosszú cxRNS-eket a genom egészén azonosítottunk, elsősorban a dRNS-Seq módszerrel (35. ábra).



35. ábra Az 1-es típusú Herpes simplex vírus szuper-hosszú RNS-ei²⁸⁹. Ezeket a nagy (≥ 4 kbps) RNS molekulákat az ONT MinION dRNS-Seq és PacBio Sequel technikák segítségével azonosítottuk. Sokuknak bizonytalan a TSS-e, különösen azoknak, amelyeket dRNS-Seq módszerrel mutattunk csak ki. Az adott genomi régióban csak a leghosszabb RNS-eket jelöltük, kivéve az egymással komplementer, átfedő transzkripteket.

6.3.4 Funkció a strukturális komplexitás mögött

Az LRS technológiák megjelenése jelentősen felgyorsította az új vírus transzkriptek és TI-ik, úgy, mint a splice, TSS és TES változatok felfedezését. Számos lncRNS-t azonosítottunk az α HV-ok OriS és OriL régiói közelében^{6,17,262,272}. Ezek funkciója még nem ismert, csak feltételezzük a replikációban betöltött szerepüket (lokációjuk, 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^{7,288,395}. Továbbá olyan RNS-ek, melyek az Ori-k körüli, a vírus transzkripció, illetve replikáció szabályozásában résztvevő génekkel fednek át²⁶².

A herpeszvírusok mindhárom alcsaládjában azonosítottunk raRNS-eket, melyek a vírus genom Ori régiói köré térképeződnek^{5,7,8,13,272,299}, továbbá a bakulo²⁷³- és cirkovírusokban²⁹² 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 (RAMPAGE)³⁹⁶] adatait is bevontuk, illetve konkrétan e régiók vizsgálata céljából végeztünk további LRS és CAGE-Seq kísérleteket, qPCR validálást⁸².

Habár nem kódoló RNS-ekről beszélünk az raRNS-ek esetében is, s így helyesebb lett volna a 6.3.3.-as fejezet alfejezeteként tárgyalni őket, azonban speciális lokációjuk és az ebből (is) következő feltételezett szerepük miatt nagyobb hangsúlyt helyezek a disszertációban rájuk, s egy főbb alfejezetként tárgyalom őket.

6.3.4.1 *raRNS-ek: transzkriptek a replikációs origók közelében*

Azok az RNS-ek tartoznak ide, melyek átfedik az OriK valamelyikét, vagy azok közelébe térképeződnek. Vannak köztük lncRNS-ek és mRNS-ek hosszú, 5'-, illetve 3'-UTR izoformái egyaránt.

a. Alfaherpeszvírusok—OriS

Az OriS lokációja az α -herpeszvírusok 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 *Simian Varicellovirus* (SVV) mindegyike esetén azonosítottunk új, az OriS körül leíró, illetve azzal átfedő transzkripteket és új RNS izoformákat.

A PRV-nél leírtuk egyrészt a NOIR-1 transzkript családot (**n**on-coding RNA in the **i**nverted **r**ep**e**at), melynek különböző hosszvariánsai mind az IR régióban íródnak le és közös 3' koterminálisban végződnek. Közös a 3' végük az antiszensz transzkripttel (AST) és az LLT-vel is⁶. A NOIR-1 spliceolt variánsát is leírtuk. Az általunk leírt NOIR-2 szintén az IR régióból íródik le, a *noir-2* gén az *ie180* géntől downstream helyezkedik el, azzal parallel orientációban, míg konvergens állásban a *noir-1*-hez képest. A NOIR-2 nagyon alacsony abundanciájú transzkript. 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. Azonosítottunk egy AZURE (**a**ntisense transcripts in the IR-**u**S overlapping **r**egion) névre keresztelt transzkriptet is, 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-t²⁵. 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 (**36. ábra**).

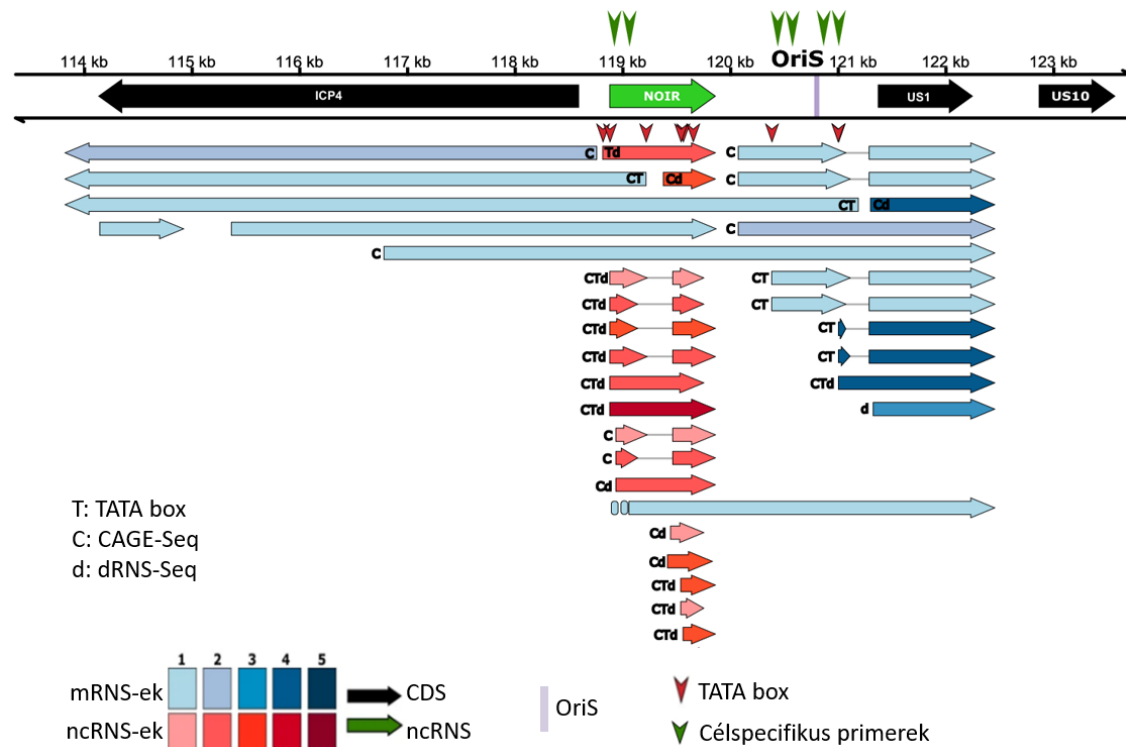
A BoHV-1 OriS-eit két egymással ellentétes orientációjú IE transzkript fedi át. Egyike ezen raRNS-eknek a bICP4 nagyon hosszú TSS transzkript izoformája, a másik pedig a bICP22 nagyon hosszú TSS variánsa. Ez a két TSS variáns az OriS mellett egymással is átfed²⁸⁴. Abundanciájukat (lévén nagyon hosszúak) az LRS alábecsüli. A bICP22-nek több izoformáját is azonosítottuk. A TAATGARAT és a TAATGARAT-szerű szekvenciák a herpeszvírusok IE génjei aktivitásához szükségesek³⁹⁷. 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 a fertőzés későbbi időpontjaiban jelennek csak meg.

A HSV-1-ben ugyanezt az organizációt figyeltük meg, míg a VZV-ben és a PRV-ben csak a bICP22 homológjának hosszú 5' UTR variánsát detektáltuk³⁹⁸.

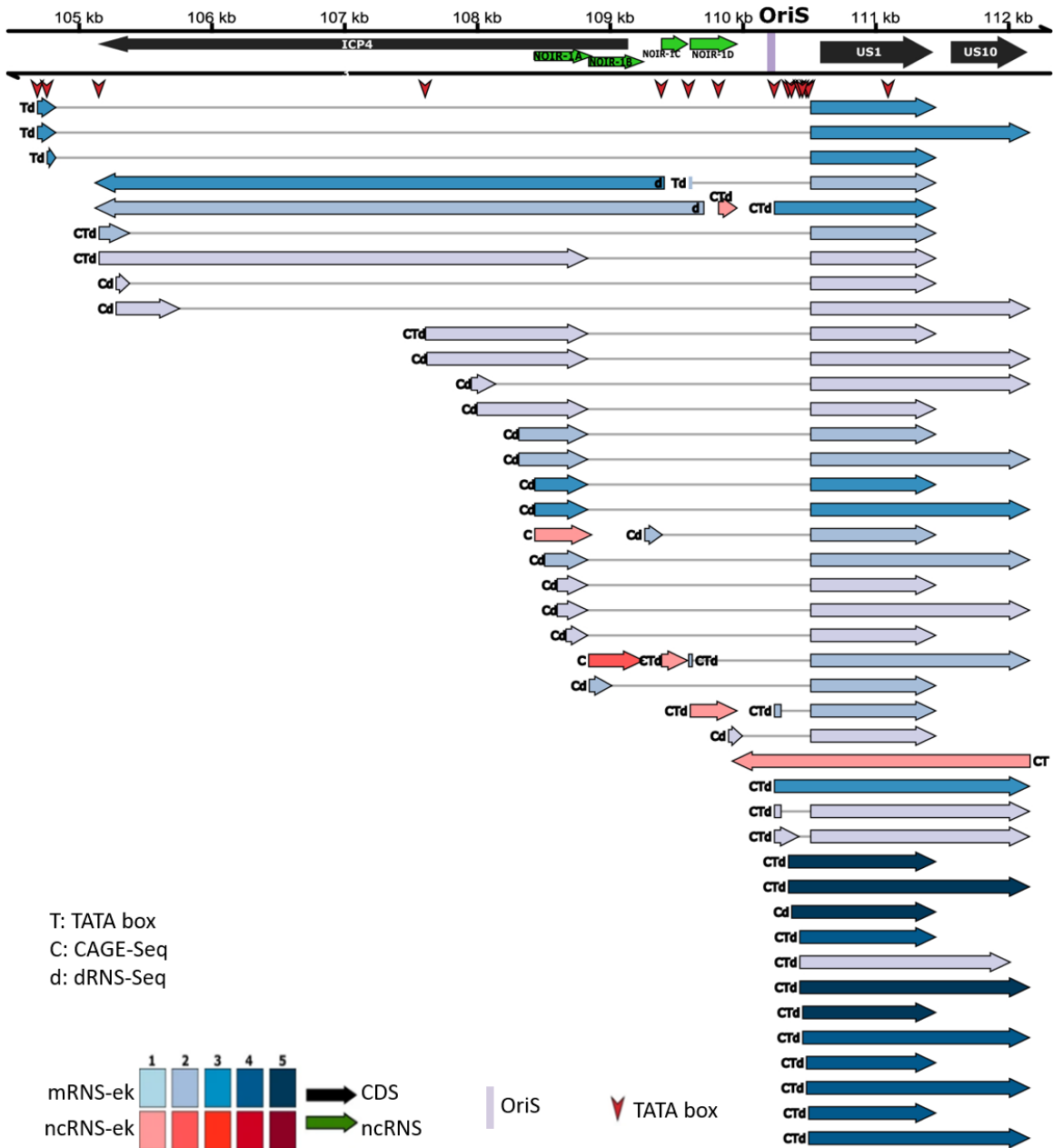
Számos raRNS-t azonosítottunk a VZV OriS-sel átfedésben is. A VZV-ben a csoportunk által kimutatott¹⁴ raRNS-eknek egy csoportja a kanonikus ORF62-ben (*icp4* gén homológja) helyezkedik el, részben azzal ellentétes orientációban (asRNS-ek). Továbbá, az *us1* homológ ORF63 hosszú TSS variánsai is átfedik az OriS-t (**37. ábra**).

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

Összefoglalva, az α HV-ok OriS közeli régiójában 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 transzkripció szabályozó gének (mint az *us1* és az *icp4*) nagyon hosszú 5' TI-jei átfednek az OriS-sel²⁵. Az α 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 α HV esetében olyan TSS-ekkel rendelkező TI-ket találtunk, amelyek nagyon közel helyezkednek el az OriS-en belüli TATA-boxokhoz, ami arra utal, hogy ezek a promóter elemek funkcionálisak lehetnek.



36. ábra Az EHV-1 OriS-közeli transzkriptjei⁸². Az ábrán az EHV-1 OriS régiója és azok környékéről leíró transzkriptek láthatóak. Fehérje kódoló gének: fekete nyíl; nem-kódoló gének: zöld; mRNS-ek: kék; ncRNS-ek: piros nyíl. A többi HV-vel való összehasonlíthatóság érdekében a HSV-1 terminológia szerint neveztük a géneket. 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 reprezentálják. A PCR során alkalmazott primerek pozícióit zöld nyilakkal jelöltük.



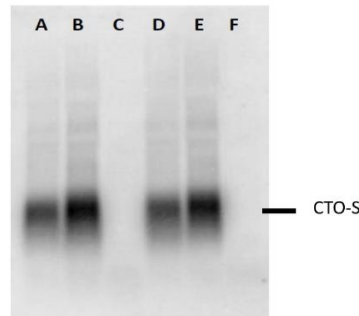
37. ábra A VZV OriS környéki transzkriptjei⁸². Az ábra a VZV által kódolt, az OriS régió környékén leíró transzkripteket mutatja. Fehérje kódoló gének: fekete nyíl; nem-kódoló gének: zöld; mRNS-ek: kék; ncRNS-ek: piros nyíl. A többi HV-vel való összehasonlíthatóság érdekében a HSV-1 terminológia szerint neveztük a géneket. Színek intenzitása az abundanciára utal: 1: 1–9 leolvasás, 2: 10–49 leolvasás, 3: 50–199 leolvasás, 4: 200–999 leolvasás, 5: > 1000 leolvasás.

b. Alfaherpeszvírusok—OriL

Az OriS régióval ellentétben az OriL változatos lokációjú az α HV-oknál: hiányzik a VZV és BoHV-1 genomokból, míg a többi vírus esetén más-más régiókban található. A HSV-1-nél az *ul29* és *ul30* gének között, míg a PRV-nél és az EHV-1-nél az *ul21* és *ul22* gének közt³⁹⁵ van.

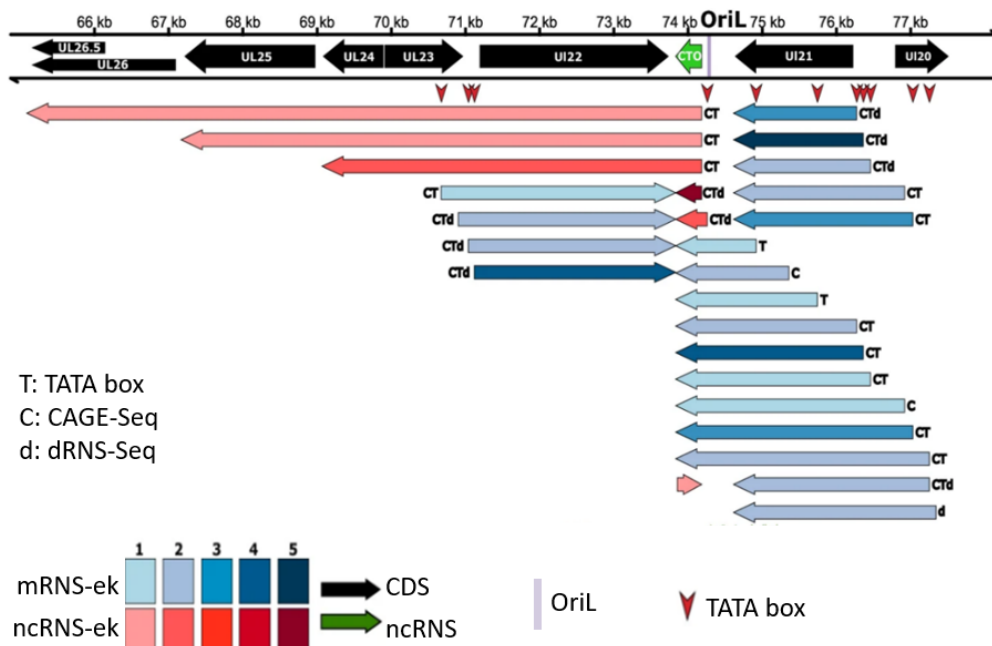
A 'close to the OriL' (OriL-hez közeli) CTO családot a PRV-ben azonosítottuk először^{3,6,272}, három, 3' koterminálisban végződő transzkriptet írtunk le. A rövid változat a CTO-S (short) (38. ábra), a közepesen hosszú a CTO-M (medium), mely az *ul21* gén poliA szignálja körül inicializálódik, illetve a hosszú, azaz CTO-L (long) mely az *ul21* gén transzkripció átírásából ered. A CTO-L az *ul21* gén

hosszú 3' UTR variánsaként értelmezhető. Az ilyen hosszú 3' UTR izoformák, egyéni TES-ek rendkívül ritkák az α HV-oknál. A CTO-S hosszú 3' UTR variánsa és a vele szemben, konvergens állású *ul22* gén hosszú TES transzkript variánsa között konvergens átfedést azonosítottunk. További technikák, nagyobb lefedettséget biztosító szekvenálást követően további RNS-eket írtunk le ebben a régióban¹⁹, többek között a CTO-S új TES izoformáját, a CTO-M és -L TSS verzióit, s egy nagyon hosszú, komplex RNS molekulát (CTO-S-cx). Azonosítottunk antiszensz expressziót is a CTO-S genomi lokuszában. A CTO-S messze a legabundánsabb PRV transzkript, a többi verzió relatíve alacsony mértékben fejeződik ki. A teljes PRV transzkriptomot bemutató ábrán is feltüntettük a CTO-S, -M és -L transzkripteket (**17. ábra**). Mind LRS-sel, mint qPCR-ral meghatároztuk a CTO expressziós dinamikáját, s az L kinetikai osztályba soroltuk²⁷².



38. ábra A CTO-S transzkript kimutatása Northern blot analízissel²⁷². Az A és a D sávokba a fertőzést követő 4, a B és E sávokba 6 h-s RNS-t vittünk fel, míg a C és F sávokba nem fertőzött sejtekből izolált RNS-t.

Az EHV-1-ben is azonosítottunk CTO transzkripteket (**37. ábra**).



39. ábra Az EHV-1 OriL-közel transzkriptjei⁸². Az ábrán az EHV-1 OriL régiója és azok környékéről leíró transzkriptek láthatóak. Fehérje kódoló gének: fekete nyíl; nem-kódoló gének: zöld; mRNS-ek: kék; ncRNS-ek: piros nyíl. A többi HV-vel való összehasonlíthatóság érdekében a HSV-1 terminológia szerint neveztük a géneket. Színek intenzitása az abundanciára utal: 1: 1–9 leolvasás, azaz read, 2: 10–49 read, 3: 50–199 read, 4: 200–999 read, 5: > 1000 read.

Az EHV-1 CTO-S transzkript nagyon hosszú TES izoformája koterminálisban végződik az ORF35, 36 és 37 génekkel. Ez a transzkript egy cxRNS, hiszen antiparallel orientációban is tartalmaz géneket (ORF38 és 39), illetve feltételezéseink szerint ncRNS, mert az első ATG (az ORF37-é) túl távol esik

a TSS-étől. Egy, a CTO-S-sel antiparallel transzkriptet is detektáltunk. Az ORF40 gén (az *ul21* homológja) kódol egy TES variánst (CTO-L), mely koterminálisban végződik a kanonikus CTO-S-sel. A PRV CTO-M-mel homológ transzkriptet nem azonosítottunk (feltételezzük, hogy csak a régióban kapott relatíve alacsony lefedettség miatt). A hosszabb CTO-S izoforma TATA boxa az OriL-lel egybeesik. A PRV-hez hasonlóan az EHV-1-ben is extrém abundáns a CTO-S RNS.

c. Bétaherpeszvírusok

Kimutattuk, hogy a korábban azonosított HCMV RNA4.9⁸⁵ transzkripciója az OriLyt-ről indul. Kimutattuk az UL59³⁹⁹, SRT⁸⁸ és vRNA-2⁸⁹ transzkripteket is^{21,82}. Továbbá, az UL58 lncRNS-nek két hosszabb, míg az UL59 lncRNS-nek egy rövidebb izoformáját is azonosítottuk⁸².

d. Gammaherpeszvírusok

Az EBV BCRF1 génjének hosszú, TSS variánsa átfedi az OriP-t⁴⁰⁰. Hasonlóképpen, a BHRF1 gén hosszú TSS izoformája az OriLyt-tel fed át. A BHLF1 gén promotere az OriLyt-en belül van⁹⁰. Mi leírtunk számos új lncRNS izoformát, melyek vagy átfedik az OriLyt-R-t az intronjaik révén, vagy az Ori-ből inicializálódnak⁸².

Kimutattuk, hogy a KSHV litikus reaktiválódása során az 1,4 kbp hosszúságú lncRNS mellett több különböző hosszúságú OriLyt-L-asszociált ncRNS is leíródhat ugyanabból a TSS-ből. Az OriLyt-L rövid, fehérje kódoló génekkel van körülvéve. Ezek balról a K4.2, K4.1 és a K4, míg jobb oldalról a K5, K6 és K7 gének^{401,402}. Korábban leírták, hogy a K4, K4.1 és a K4.2 mind mono-, bi- és tricisztronos formában is leíródnak⁴⁰³, a mi elemzésünk azonban egy sokkal összetettebb RNS profilt tárt fel, beleértve spliceolt, nem spliceolt, s különféle hosszúságú RNS variánsokat egyaránt. Feltártuk azt is, hogy a K5 és K6 gének nem csak egyedileg fejeződnek ki, hanem splicingon keresztül is, ami különböző hosszúságú első exonnal rendelkező mRNS-eket eredményez. Fontos megjegyezni, hogy eredményeink összhangban vannak a korábbi transzkriptomikai tanulmányokkal⁴⁰⁴⁻⁴⁰⁶, ugyanakkor az OriLyt-L lokuszról leíródo különböző virális RNS-ek számát is bővítik. A KSHV látencia lokusza négy fehérjekódoló gént [K12, K13, ORF72, ORF73 (LANA)] és tizenkét pre-mikro RNS-t (pre-miRNS-t) kódol⁴⁰⁶⁻⁴⁰⁸. Itt több lncRNS-t is azonosítottunk, amelyek antiszenszerek a miRNS-kódoló genomikus régiókkal szemben.

6.3.4.2 *raRNS-ek: a transzkripciót szabályozó gének RNS izoformái*

Az *us1* génről leíródo RNS-ek közül számos nagyon hosszú 5'-UTR izoforma, amelyek divergens TO-kat alakítanak ki az ICP4 transzkriptekkel az EHV-1-ben (**36. ábra**), VZV-ben (**37. ábra**) és HSV-1-ben. Az HSV-1-ben az US10-12 poligén transzkriptek 5'-UTR izoformái divergens TO-t hoznak létre az *icp4* génekkel. Továbbá, az HSV-1-ben észleltünk egy ICP4 TI-t egy kiterjedt 5' UTR-rel, ami átfedi az *us1* gént. A BoHV-1-ben az *icp4* 3' UTR izoformája parallelt TO-t alkot a downstream *icp0* génnel, de az *icp4* ORF ki van vágódva ebből a transzkriptből, s így egy fúziós RNS-t eredményez, amely tartalmazza a teljes hosszúságú *icp0* gént és az ICP4 RNS 5' UTR-jének egy részét. Továbbá, ICP4 TI-k hasonló fúziós és bicisztronos RNS-eket hoznak létre a BoHV-1 CIRC RNS-sel.

6.3.4.3 A transzkripciót szabályozó gének közelébe térképeződő ncRNS-ek

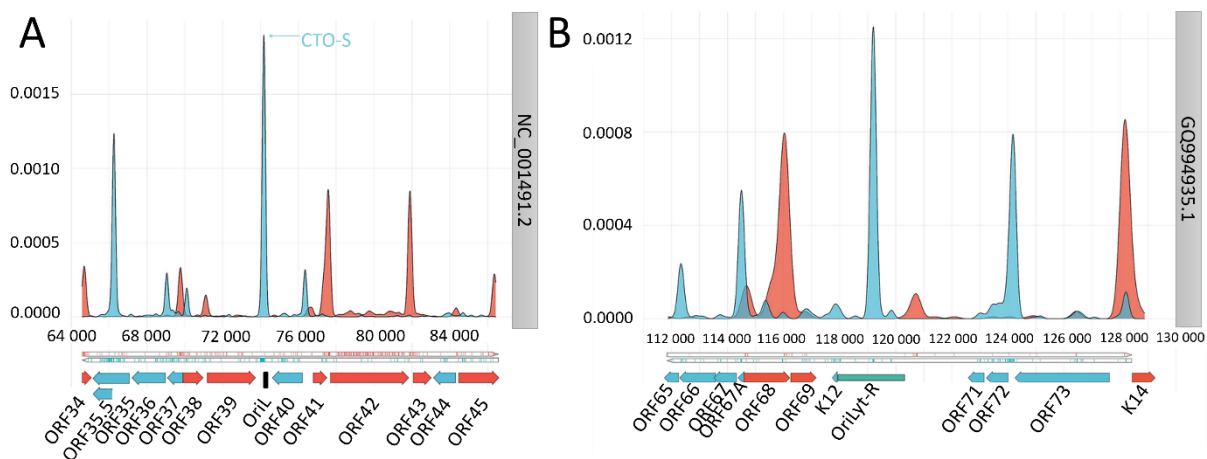
A BoHV-1-ben és PRV-ben olyan asRNS-eket azonosítottunk, amelyek átfedik az *us1* gént. Az ELIE az *icp4* és *icp0* gének között helyezkedik el, az egyik TI-je 5'-koterminális a NOIR-1 transzkriptokkal. Az ELIE-t a PRV mellett az EHV-1-ben is azonosítottuk. Továbbá leírtuk az AZURE IncRNS-t a PRV-ben, az *as64*-et az EHV-1-ben, s számos egyéb, RNS-t és TI-t melyek a transzkripció szabályozásában részt vevő gének környékén íródnak le, az azokról leíródo transzkriptokkal, vagy azok TI-ival átfednek^{6,25}.

6.3.4.4 A DNS replikációban részt vevő gének átfedései

A Simplexvirusokban a divergens *ul29-ul30* génpár hosszú 5' UTR izoformái nem csak az OriL-t, hanem egymást is átfedik. Érdekeség, hogy mindkét gén a DNS replikáció fő szabályozóit kódolja. Mind a HCMV (*ul57*), mind a HHV-6 (*ul42*) esetében az *ul29* ortológok az OriLyt mellett található. Kiemelendő, hogy az α HV-okban összesen három "kemény" TO létezik a génpárok között, s ezek közül az egyik partner mindig egy olyan gén, amely a vírusreplikációban vesz részt. Ezek a TO-k magukban foglalják az *ul30/ul31*, *ul6-7/ul8-9*, *ul50/ul51* párokat (*ul30*: DNS-polimeráz; *ul8*: DNS helikáz; *ul9*: origin binding protein - OBP; *ul50*: dezoxiuridin-trifoszfátáz).

6.3.4.5 A TSS-ek meghatározása

A transzkriptom kutatások egyik nagy kihívása az RNS-ek TSS-einek pontos meghatározása. Amellett, hogy számos LRS és SRS módszert alkalmaztunk, validálásként a CAGE-Seq technikát is bevezettük. Bizonyos vírusok esetén végeztek korábban CAGE-Seq-et, ezen esetekben csak az adatokat használtuk fel, azonban számos vírus, így pl. KSHV, EHV-1 (**40. ábra**), vagy az MPXV esetén saját kísérleteket végeztünk. A KSHV eredményeinket mások RAMPAGE adataival³⁹⁶ is összevetettük.



40. ábra TSS-eloszlás a vizsgált genomi régiókban a CAGE-Seq alapján⁸². Az 5' végeket simított sűrűségi diagramokon ábrázoltuk a CAGE adatok alapján. A TSS-ek eloszlását az EHV-1 és a KSHV egy-egy genomiális régiójával szemléltettem: (A) EHV-1 OriL; (B) KSHV OriLyt-R; Nagyobb felbontást használtunk az alacsony gyakoriságú TSS-ek jobb láthatósága érdekében. A CTO-S transzkript expresszáója nagyon nagy mértékű (A). Az y tengely az 5' végek valószínűségi becslését mutatja valószínűségi sűrűségfüggvény segítségével⁸². Az ábrák alsó részében kerültek feltüntetésre a megfelelő genomokhoz tartozó kódoló szekvenciákkal kapcsolatos információk (ORF-ek, Ori-k; míg a jobb oldalon a genomok GenBank referencia azonosítói. A pozitív szál lefedettség adatait és a kódoló szekvenciák annotációját piros, míg a negatív szálat kék színnel ábrázoltuk. Az EHV-1 Ori régiója feketével, míg a KSHV-ben zölddel van ábrázolva.

6.3.5 Dinamikus transzkriptom: a génexpresszió időbeli változásainak elemzése

Számos vírus esetén meghatároztuk a géntermékek relatív expressziós szintjét. Az IE gének azonosításához 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 PRV⁷ és a Sequel-lel a VACV^{17,297} esetén, míg ONT MinION-on szekvenáltuk a HSV-1²⁸⁸, a VACV^{17,297} és a BoHV-1²⁸² dcDNS könyvtárakat a dinamikus RNS profiljuk elemzéséhez.

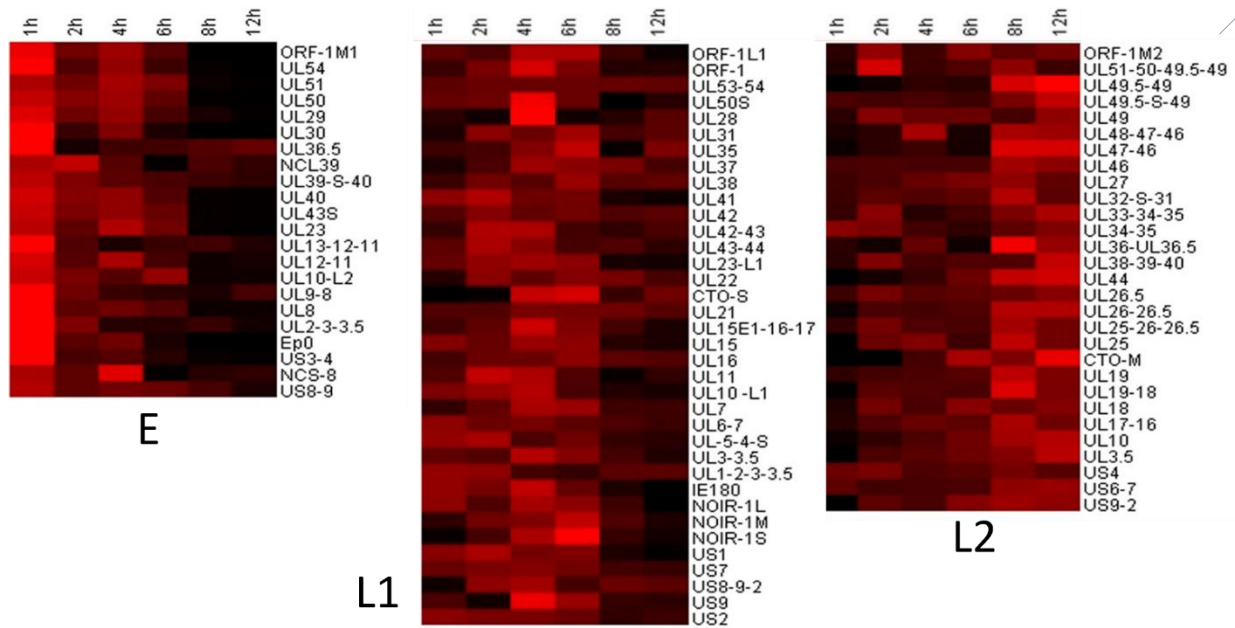
6.3.5.1 A dinamikus PRV transzkriptom

A herpeszvírusok transzkriptomának (főként HSV-1 és PRV) dinamikus tulajdonságait korábban számos módszerrel tanulmányozták, beleértve a microarrayt^{409,410} Illumina szekvenálást^{299,411}, valamint az általunk végzett qRT-PCR elemzéseket^{2,412}. Saját PhD témám is a PRV génexpresszió qRT-PCR-rel való elemzését emelte a fókuszba², s azt követően, az LRS módszer bevezetését megelőzően is ezzel a technikával foglalkoztam, vizsgáltuk különféle mutáns PRV törzsek fehérje kódoló génjeinek kinetikáját, a génmutációk össz-génexpresszióra gyakorolt hatását (ezt a 6.3.7.-es fejezetben tárgyalom²⁷⁹⁻²⁸¹).

Az LRS módszerek megjelenésével azokat elsőként vezettük be a vírusok dinamikus transzkriptom profilalkotásához is⁷. 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, másodsor pedig annak bemutatása, hogy az LRS módszer hogyan használható és mennyire hasznos a globális transzkripció kvantitatív elemzésében, az időben változó génexpresszió össz-genomszintű profilozásában.

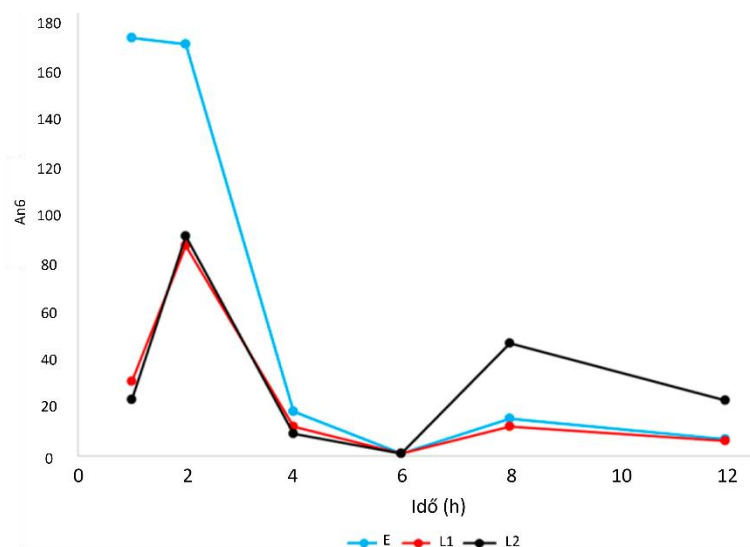
A könyvtárkészítést a PacBio „Very Low Input” protokollja szerint végeztük, mely a relatíve rövid (1-2 kb) hosszúságú readeknek kedvez, ezért az ennél hosszabb és rövidebb leolvasások alulreprezentáltak a mintánkban. Az abszolút kvantifikálást nem tette lehetővé a módszerünk, azonban a relatív expressziós (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 poligénus természete egy további nehézség volt a transzkriptek azonosítása során. Azokat a poligénus egységekről származó, inkomplett readeket, melyek legalább 8 bp-t tartalmaztak az upstream génből, a hosszabb RNS molekulához kalkuláltuk. Ezt a küszöböt a PRV statikus elemzésekor⁶ tapasztalt, a transzkriptek TSS-ére vonatkozó (~ 4 bp-os) hossz polimorfizmus miatt állítottuk be. A vírus transzkriptek 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 (**41. ábra**).



41. ábra A PRV gének transzkripciós kinetikája hőterképés ábrázolása⁷. 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

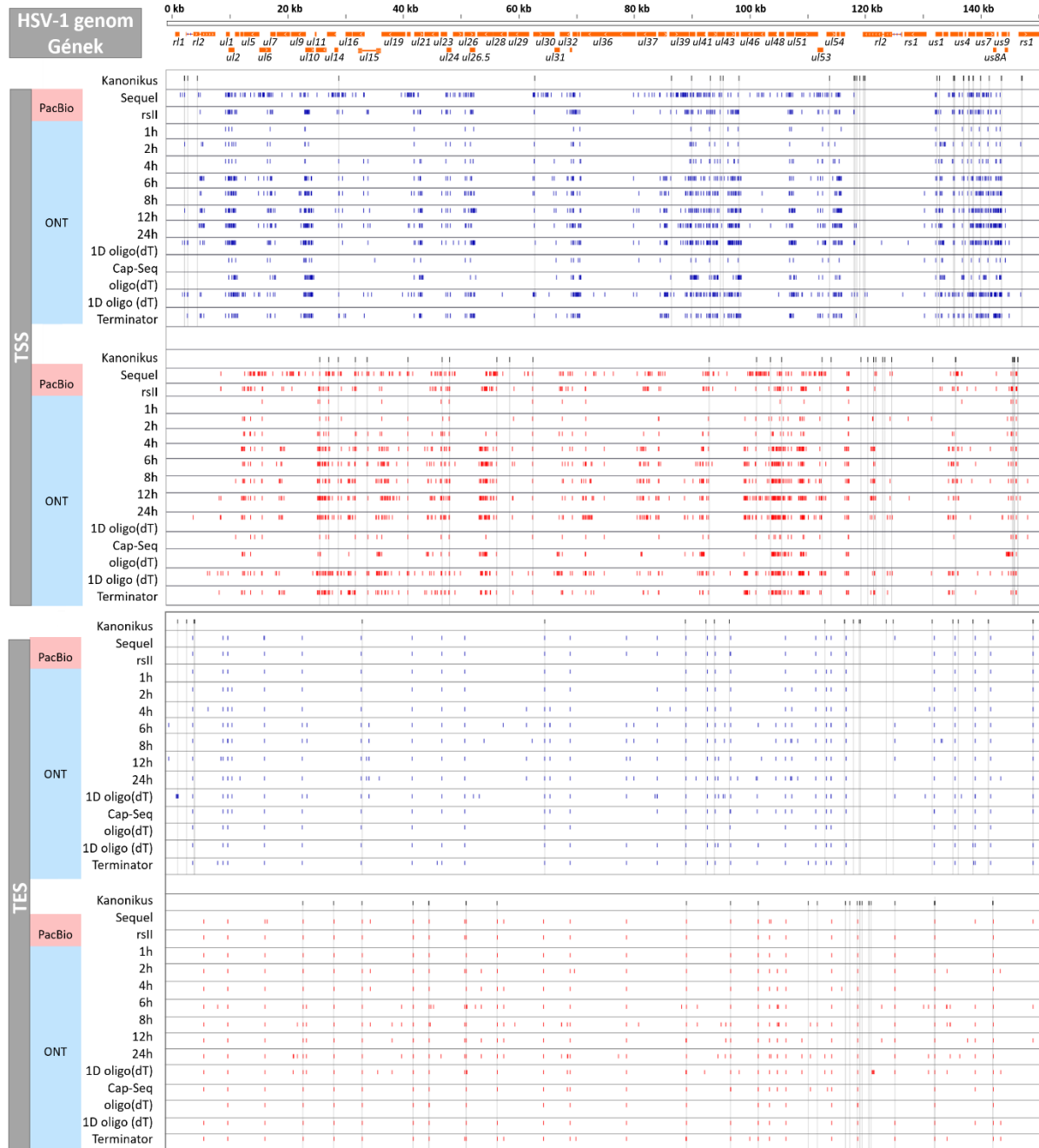
A 42. ábra azt mutatja, hogy az általános E, L1 és L2 génkifejeződések jelentősen különböznek egymástól. Az E gének magas relatív kifejeződést mutatnak a DNS replikáció kezdete előtt, amely 12 h-val a fertőzés után jelentősen csökken. Az L2 gének 'inverz' kinetikát mutatnak az E-khez képest. Az L1 gének kezdetben eltérően viselkednek az E génektől, míg később expressziós görbájük 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.



42. ábra A három kinetikai osztályba tartozó transzkriptek Fx értékei átlaga⁷. Az E gének relatív expressziós értéke magas a fertőzést követő első órákban, míg az L2 géneké a késői időpontokban. Az L1 transzkriptek expressziós profija intermedier.

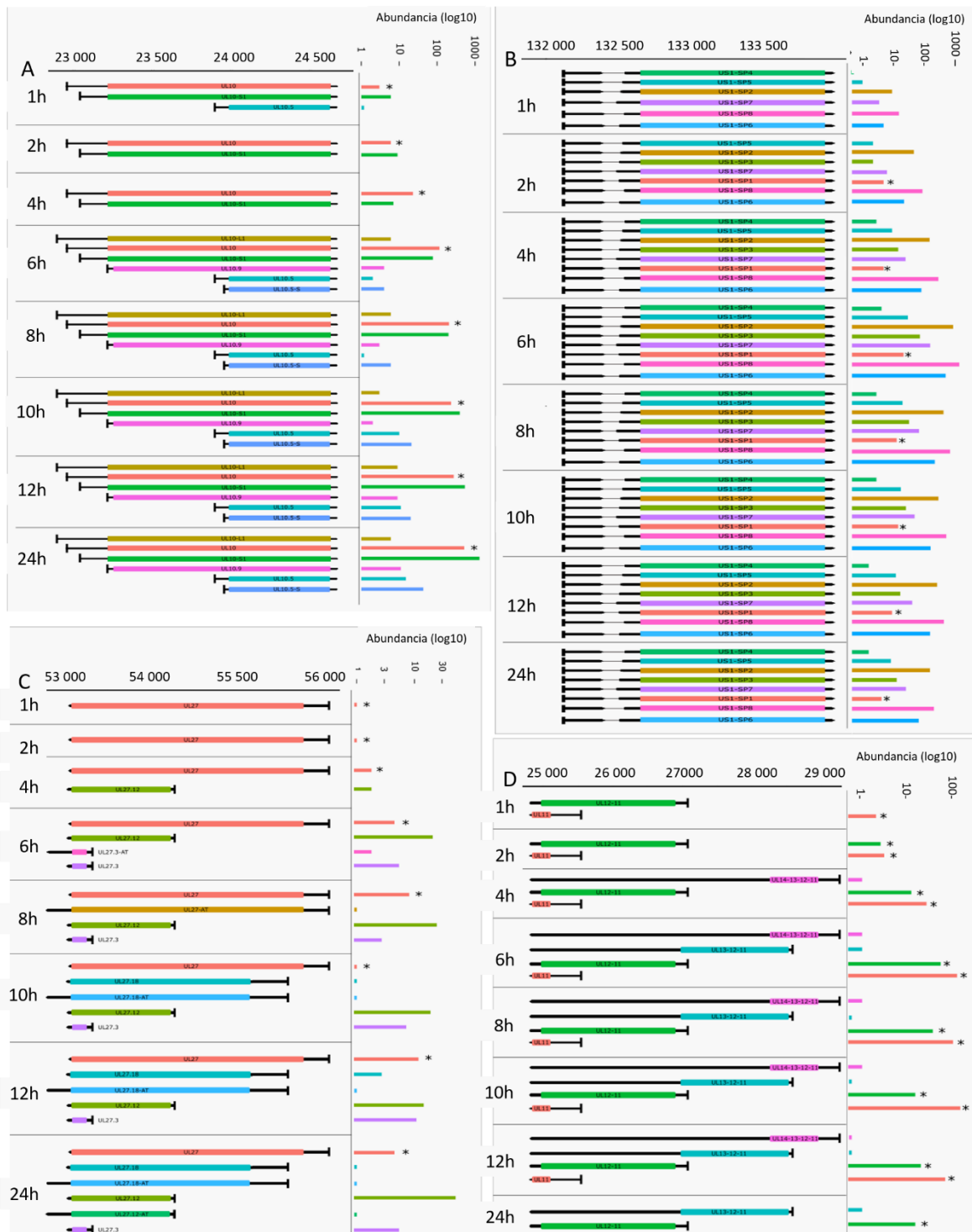
6.3.5.2 A HSV-1 gének kinetikai analízise

A HSV-1 gének a fertőzést követő expressziós dinamikáját is elemeztük. Az egyes transzkripteket a LoRTIA programcsomaggal határoztuk meg. A program azonosítja a TSS-eket, a TES-eket, az intronokat, majd ezekből „építi fel” a transzkripteket, TI-akat. A **43. ábra** a HSV-1 gének TSS-ei és TES-ei eloszlását, s ezek időbeli változását mutatja összgenom szinten. Az ábrákon az ONT szekvenálás nagy időbeli felbontású adatai mellett a többi könyvtárkészítési és szekvenálási módszerrel (mix RNS-ekből származó) nyert adatainkat is feltüntettük.



43. ábra A HSV-1 TSS és TES pozícióinak összgenom szintű ábrázolása²⁸⁸. A TSS-eket (A) és a TES-eket (B) a LoRTIA szoftverrel állapítottuk meg. A kék szín a forward, míg a piros a reverz szálon aktív TSS és a TES pozíciókat mutatja. A fekete vonalak a kanonikus, korábban már azonosított TSS-eket, illetve TES-eket jelzik, ezek szürke vonallal az egész ábrán keresztül mutatják az ismert pozíciókat. A narancssárga téglalap alakú blokkokkal az ORF-eket ábrázoltuk.

A különböző transzkript kategóriák dinamikáját a **44. ábrán** szemléltetjük, beleértve a TSS izoformákat (A panel), splicing variánsokat (B panel), TES izoformákat (C panel), és poligénes RNS-eket (D panel). Számos mono- és policisztronos RNS és TI eltérő mértékben expresszálódik a vírus fertőzési ciklusa során, az egyes (általunk vizsgált) időpontokban.



44. ábra **Példák a dinamikus HSV-1 transzkriptomból**²⁸⁸. Az ábrák bal oldalán a TI-k strukturáját ábrázoljuk sematikusan, míg jobb oldalt az egyes izoformák időpontoként változó abundanciáját bar ploton mutatjuk log₁₀-es skálán. (A) Az ul10 gén 5' UTR variánsai és 5' végén csonkolt izoformái. (B) Az us1 splice variánsai. (C) Az ul27 és izoformái, pl. alternatív termináció. (D) Az ul11-14-es régió koterminálásban végződő transzkriptjei – mono- és poligénes RNS-ek egyaránt. * mások által leírt variánsok

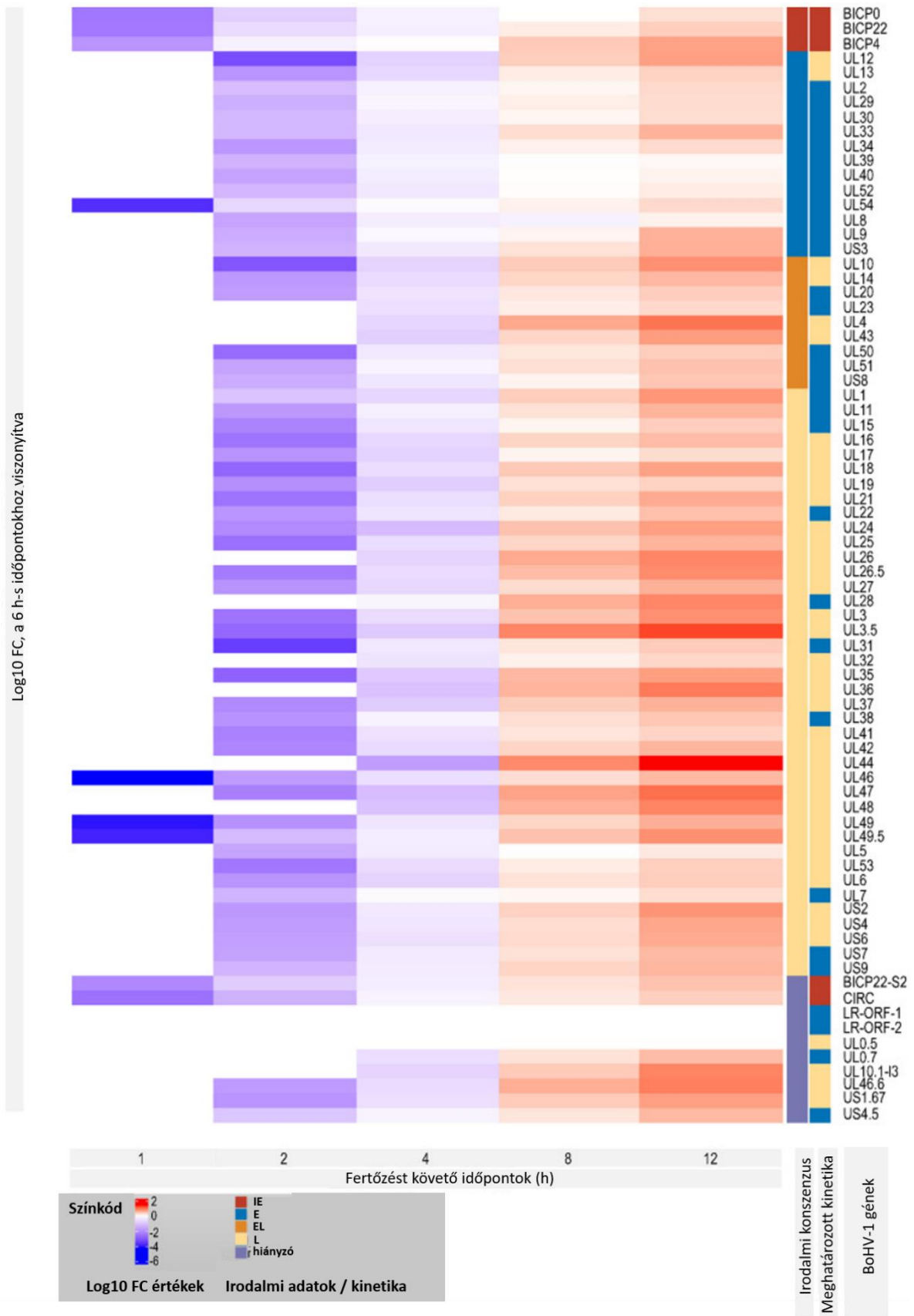
6.3.5.3 A BoHV-1 összgenom-szintű génexpressziós változásainak vizsgálata

A CHX-szel kezelt mintákban az alábbi gének esetén mértünk expressziót: *bicp4*, *bicp22*, *bicp0*, *circ*, és *ul54*. A *bicp22* és a *bicp4* nagyon magas, míg a többi három mérsékelt expressziót mutatott. A *bicp22* gén expressziós szintje magas a teljes vizsgálati periódus alatt a nem kezelt sejteken is, *bicp4* viszont kevés RNS-t termelt. Azt feltételezzük, hogy a CHX-kezelt mintákban való expressziója ellenére az *ul54* nem valódi IE gén, mert eltérően a *bicp4*, *bicp22*, *bicp0*, és a *circ* génektől, melyek esetén az expresszió mértékét a CHX kezelés (és ennek koncentrációja) gyakorlatilag nem befolyásolja, az *ul54*-é drasztikusan csökken a növekedett CHX dózis hatására.

Az IE gének aktivitásához szükséges TAATGARAT szekvenciához a tegument VP16 fehérje köt, amely az Oct-1 sejtes faktort segíti⁴¹³ abban, hogy az IE transzkripcióban részt vegyen.

Mi nem azonosítottunk TAATGARAT-szerű szekvenciát az *ul54* promóterében. Nem zárjuk ki azonban, hogy az *ul54* valóban egy IE én a BoHV-1-ben, de – másokhoz hasonlóan inkább ennek ellenkezőjét valószínűsítjük. Az *ul54* gén CHX-kezelt sejteken mutatott expressziója azzal a ténnyel magyarázható, hogy ez a gén a korai fertőzési időpontokban nagyon aktív, melyet a szuboptimális (20 mg/mL) CHX közegben is láthatunk. Csak feltételezni tudjuk, hogy a 100 mg/mL CHX koncentráció esetleg nem volt elegendő az *ul54* gén transzkripciójának teljes blokkolásához, vagy elképzelhető az is, hogy a tegument bICP4 transzaktivátor képes volt az említett génről az észlelt alacsony szintű expressziót kiváltani (a BoHV-1 bICP4 transzaktivátort szintén kimutatták a vírus tegumentjében⁴¹⁴). Az *ul54* IE gén a *Simplexvirus* genuszban, de nem az a *Varicellovirus* nemzetségben, ahová a BoHV-1 is tartozik. Azt találtuk, hogy az *icp4* promótere és a gén hosszú izoformáját szabályozó promóter TAATGARAT-szerű szekvenciát tartalmaz. A *circ* gént egy korai tanulmánytól eltekintve⁴¹⁵ nem tekintik IE génnek. Eredményeink szerint a *circ* IE expressziója legalább részben az *icp4* promóter által szabályozott. Erre a bizonyíték a kimerikus ICP4/CIRC transzkript. Ugyanez a helyzet a *bicp0* és *bicp4* génekkal, melyek az IETu1-ről olvasódnak le, mint egy egységnyi transzkript, mely aztán különféle splicing eseményeken megy keresztül, s így generálja a bICP4 RNS-t (IER4.2) és a bICP0 RNS-t (IER2.9)³⁹³.

Ezt követően a kanonikus, illetve a nagy abundanciájú, alternatív TSS-ek transzkripciók kinetikáját vizsgáltuk a vírusfertőzés során (**45. ábra**). A TSS-ek expressziós értékét úgy határoztuk meg, hogy a readszámukat elosztottuk a minta össz vírus readszámával, majd a géneket az expressziós görbéjük alapján kategorizáltuk. Továbbá, összehasonlítottuk az eredményeinket a korábbi, klasszikus módszerekkel meghatározott kinetikai besorolásokkal^{122,416-419}. Egy transzkriptet IE-nek határoztunk meg, ha a maximum expressziós szintje 1, vagy 2 h-kor van (utóbbi esetben csak akkor, ha 1 h-kor is nagymértékű a kifejeződése). Korai kategóriába azokat a géneket soroltuk, melyek 2, vagy 4 h-kor érik el az expressziós csúcst, előbbi esetben csak, akkor, ha 1 h-kor alacsony a kifejeződésük mértéke. Az *ul54* gén ez a megközelítésünk alapján is az E kategóriába került. Az E/L gének a maximumukat 6 h-kor, míg az L gének 8, vagy 12 h-kor érik el. Ez a kategorizálás felfedte a *circ* gén kanonikus promóterét, és azt, hogy mind a kanonikus, mind a legerősebb alternatív promóter (amely a bICP22-S2 transzkript kifejeződését szabályozza) IE kinetikát mutat, annak ellenére, hogy nem tartalmaznak TAATGARAT-szerű szekvenciákat. Így, ahogy fentebb tárgyaltuk, ennek a két génnek az IE jellemzői nem kizárólag az *icp4* promóter által vannak meghatározva.



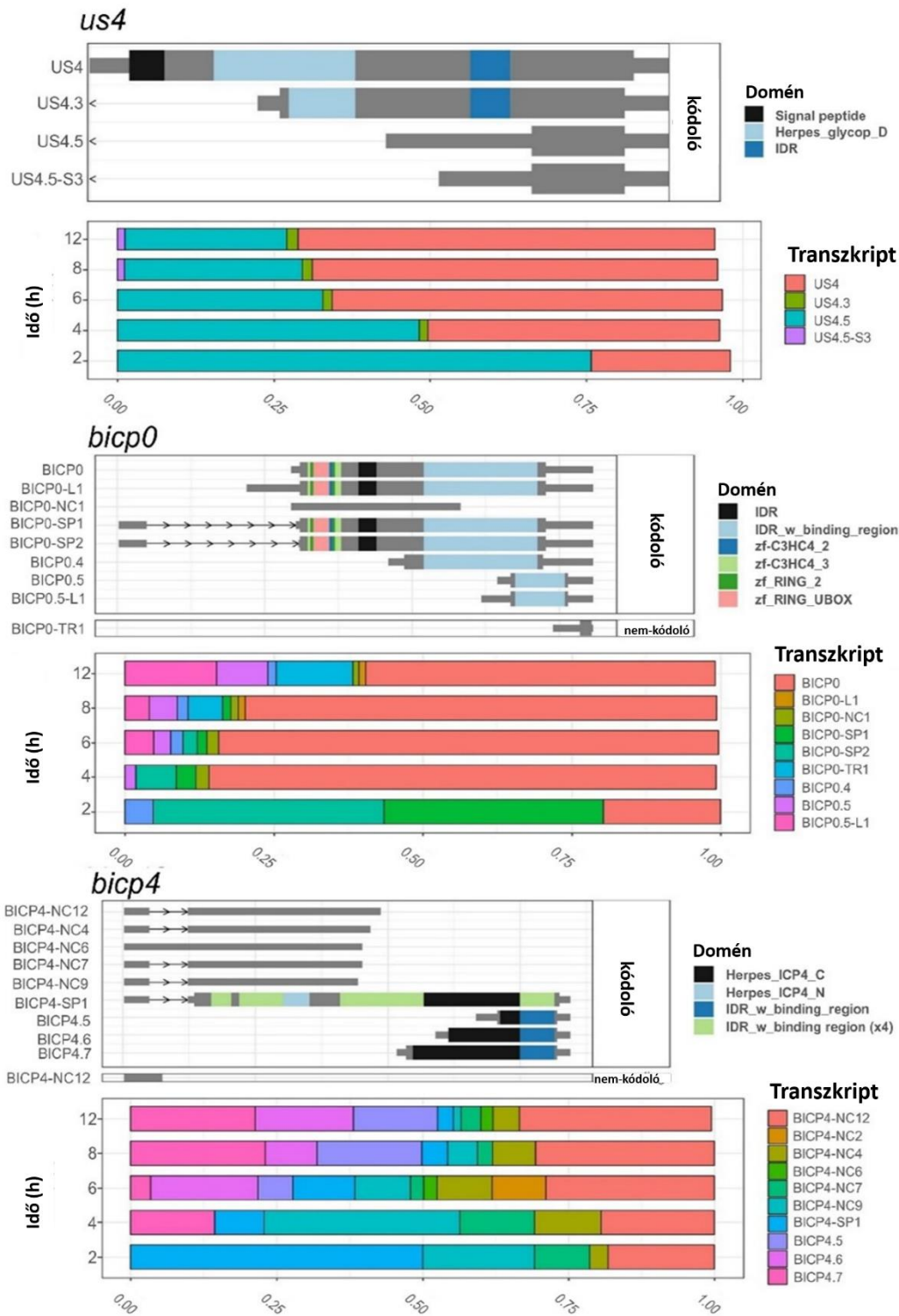
45. ábra A BoHV-1 normalizált génexpressziót ábrázoló hőterkép²⁸⁴. Az adott gén adott időpontbeli TSS abundanciáját osztottuk a 6 h-s értékkel, s ennek log10-es értékét ábrázoltuk a heatmapen. A 6 h-s adatokat nem ábrázoltuk. A két jobbszélső oszlop a kinetikai annotációkra utal, a belső az irodalmi, még a külső a saját adataink alapján.

6.3.5.4 A vírusgén domének időfüggő kifejeződése

Az izoformák (különösen a splice variánsok) amelyek különböző fehérjedoméneket hordoznak, gyakoriak az emberi sejtekben és bizonyították, hogy fontos szerepet játszanak a rák kialakulásában⁴²⁰. 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 funkcionalitást indukálhat, azaz biológiai következményeket változtathat meg például a kódoló szekvenciák fehérjedoméneinek jelenlétének/hiányának megváltoztatásával. A BoHV-1 transzkriptom vizsgálata során a vírusfertőzés alatt e lehetséges jelenségek megfigyelése érdekében az isoformswitchanalýser eszközt³³⁸ használtuk a LoRTIA által kapott transzkript mennyiségeken.

A **46. ábra** az annotált fehérjedoméneket, jelpeptideket és IDR-eket mutatja be a néhány példagén transzkriptjeiben. A transzkripteket 0,001-es küszöbértékre szűrtük (csak azokat a transzkripteket tartottuk meg, amelyek elérték 0,1%-ot a szülő génjük abundanciájából bármelyik mintában). Az elemzés eredményei szerint az *us4.5* gén 5' végén csonkolt transzkriptjei nem rendelkeznek az összes azonosított fehérjedoménnel 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 (**46. 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: a 4 órás mintákban sok rövid izoforma íródik le a kanonikus UL44 mellett, de 6 órára ezek mennyisége jelentősen csökkent, és a kifejeződési profil a kanonikus izoformára állt be. Az azonnali korai *bicp4* gén esetében ez éppen fordítva volt: a kanonikus bICP4-SP1 transzkript csak a 2 órás mintában volt domináns. A 4 órás mintákban a 3' végén csonkolt RNS-ek magas expresszióját mértük; majd a 6 órás mintáktól csökkent az arányuk az 5' végén rövidült transzkriptekhez (bICP4.5, 4.6 és 4.7) képest. A 8 és 12 órás mintákban azok a transzkriptek, amelyek nem hordozzák az ICP4 fehérjedoménjének C-terminális részét (Herpes_ICP4_C), váltak a legabundánsabb bicp4 izoformává.

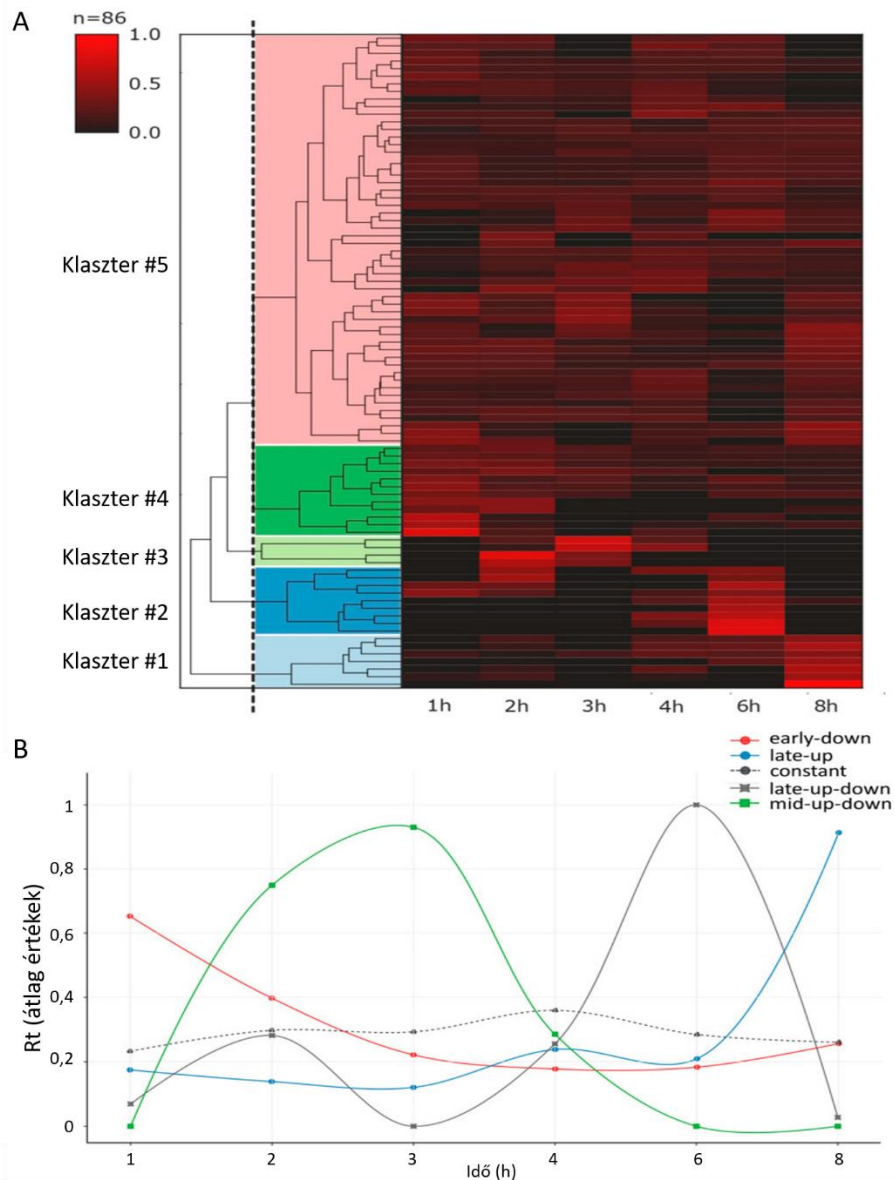
A *bicp0* gén két splice variánst fejez ki (azonos doménszerkezettel) a kanonikus izoformával nagyjából egyenlő mennyiségben, de 2 óra után jelentős csökkenést figyeltünk meg. Több 5'-csonkolt izoforma (néhány annotált domén nélkül) kisebb abundanciában fejeződött ki a vírusfertőzés későbbi szakaszaiban. Több másik gént is azonosítottunk az isoformswitchanalýserrel, 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.



46. ábra A színes boxok a Pfam adatbázis használatával azonosított fehérje doméneket mutatják²⁸⁴. Az ábrán a *bicp4*, *bicp0* és az *us4* gének a fertőzés során mutatott izoforma struktúráit és az expressziós profiljukat tüntették 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 prediktált ORF-eket tartalmaznak.

6.3.5.5 A VACV dinamikus transzkriptoma

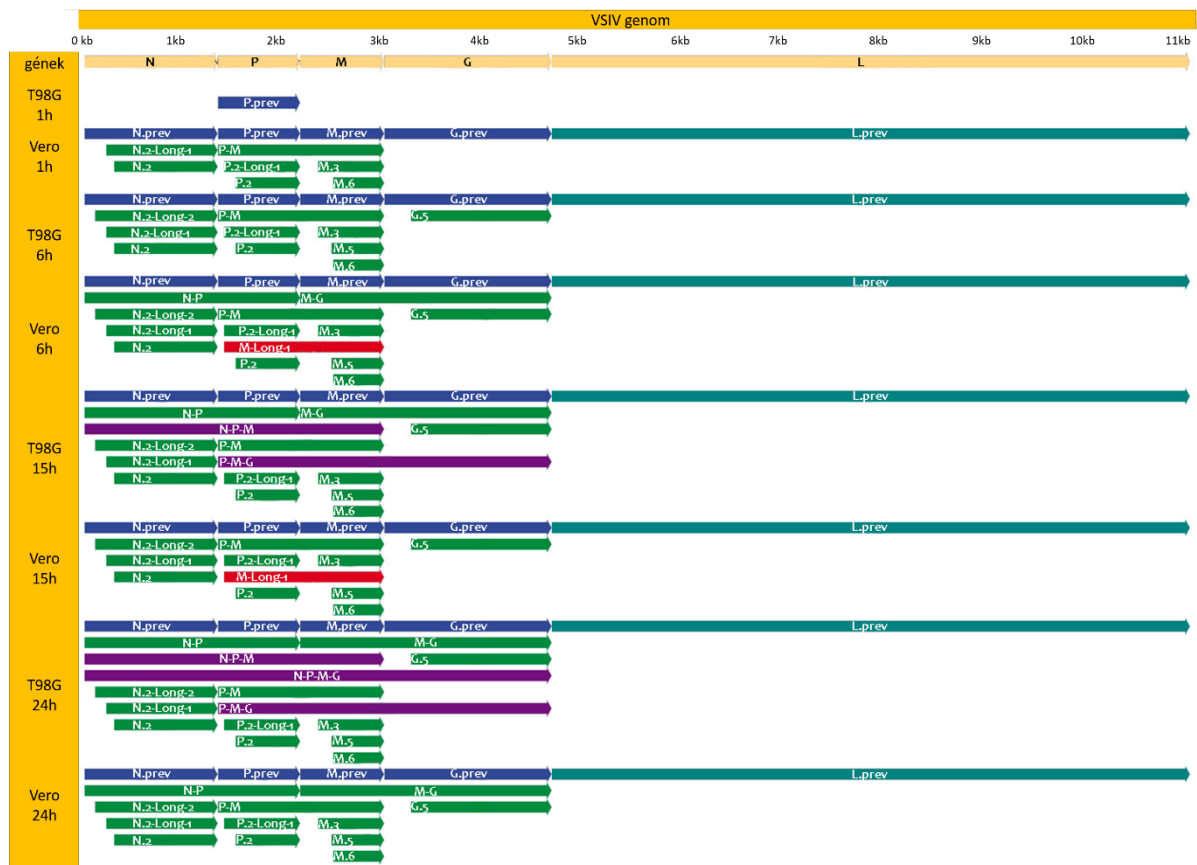
A LoRTIA-val azonosított transzkriptok kinetikai analizéséhez k-means és hierarchikus klaszterezést végeztünk (47. ábra, A). A k-means klaszterezés arra szolgált, hogy leírja ezeknek a klasztereknek a temporális kifejeződési szintjeit (47. ábra, B). Az alábbi expressziós profilokat kaptuk: az „early-down” csoport folyamatosan csökkenő Rt értékeket mutatott a fertőzés kezdetétől, míg a „mid-up-down” klaszter helyi maximuma 2 és 3 h között figyelhető meg. A „late-up-down” RNS-ek klaszterének Rt értékei 6 h-nál érték el a maximumot; a „late-up” klaszterbe tartozó transzkriptoknak egyértelmű késői aktivitása van, 8 h-nál érve el az Rt értékek maximumát, valamint a konstans, „constant” klaszterbe tartozó transzkriptok a teljes vírusciklus során kismértékű expressziós változást mutattak.



47. ábra A legabundánsabb VACV RNS-ek és TI-k expressziós profiljai²⁹⁷. A vírus transzkriptok Rt értékei öt kinetikai csoportban különböző expressziós profilokat mutatnak. (A) A vírus transzkript klaszterek expressziós mintázata. Öt különböző VACV transzkript klasztert ábrázolja a hőtérkép mátrix. Minden sor az RNS relatív expressziós szintjeinek változásait reprezentálja. A piros téglalapok magas relatív expressziós értékeket, a fekete téglalapok alacsony relatív expressziós értékeket jeleznek. (B) K-means klaszterelemzést használtunk az öt, hierarchikus klaszterezéssel nyert csoport időbeli expressziójának jellemzésére. A klaszterek relatív Rt átlagértékeit ábrázoltuk a fertőzés vizsgált időpontjaiban.

6.3.5.6 A VSIV transzkriptom időbeli elemzése

Az LRS módszerrel kimutattuk, hogy a látszólag egyszerű, kisméretű VSIV genom egy valójában bonyolult transzkriptomikai struktúrát kódol^{285,421}, s hogy ez a vizsgált két sejtvonal esetében jelentős különbségeket mutat, mind a transzkriptek felépítését, mind pedig azok időbeli kifejeződési dinamikáját illetően. A transzkripteket, beleértve az egyes mintákban lévő RNS-ek readszámait is, a LoRTIA programmal határoztuk meg, s vizsgáltuk a transzkriptek kódoló kapacitását is. Ezzel a megközelítéssel összesen 16 új transzkriptet azonosítottunk (mindkét sejtvonalban), amelyek kifejeződnek a VSIV genomról (**48. ábra**). Ezenkívül találtunk az egyes sejtvonalakra jellemző transzkript variánsokat is.



48. ábra A VSIV transzkriptom kinetikai profilja⁴²¹. Az ábrán az annotált transzkripteket ábrázoltuk időpontként és sejtvonalanként. A sárga nyilak a VSIV géneket jelölik; a kék nyilak a korábban már leírt, kanonikus RNS-eket illusztrálják; a zöld nyilak azon transzkripteket illusztrálják, amelyek mindkét sejttípusban azonosak; az indigókék nyilak csak a gliasejtekben található transzkripteket jelölik; a sötétvörösek pedig csak a fibroblaszt sejtekben találhatóakat. Nem kaptunk olyan readet, amely a teljes L gént leírta. Ezt teal (kékeszöld) színnel jelöltük. N: nukleokapszid, P: foszfoprotein, M: mátrix fehérje, G: glükoprotein, L: polimeráz

A VSIV transzkriptek kinetikai elemzése különböző szerkezeti és időbeli expressziós mintázatot tárt fel a VSIV gének esetében a két sejtvonalban. Eltérés figyelhető meg az 5'-csonkított és a kanonikus RNS-ek arányában a gének esetében minden időpontban és mindkét sejttípusban. Általánosságban a fibroblaszt sejtek nagyobb százalékban termeltek beágyazott transzkripteket. Kivételt képeznek az L és a G gének, ahol nagymértékű a beágyazott gének expressziója. Ezeknek a rövid RNS-eknek az aránya hasonló mintát követett: nagyon alacsony százalék az infekció kezdetén, csúcs 6, vagy 15 h-nál, majd csökkenés 24 h-nál. Ezeket izoforma-váltási eseményeknek tekinthetjük. Az 5'-csonkított transzkriptek legmagasabb arányát az M és N gének esetében

figyeltük meg: itt a gén kifejeződésének 60–75%-a állt ezekből transzkriptekből. Ez a minta még inkább megfigyelhető volt az M gén esetében a gliasejtekben is, csak itt az arányok alacsonyabbak voltak, de a görbe alakja egyértelműen hasonló. Ez arra utal, hogy ezeknek a rövidített transzkripteknek a kifejeződése eltérően szabályozott a gazda (kanonikus) RNS-eihez képest, valamint a két sejtvonalban is különböző. A P gén esetében a poligénus transzkriptek hasonló mintát mutattak, de arányuk alacsonyabb volt. Mindkét sejtvonalban a legmagasabb értékeket a P gén 6 h-s mintáiban észleltük: gliasejtekben 1,8%, fibroblaszt sejtekben 5,5%. Az arányok ezek alatt az értékek alatt ingadoztak egyértelmű trend nélkül.

6.3.6 Poszttranszkripciók modifikációk elemzése

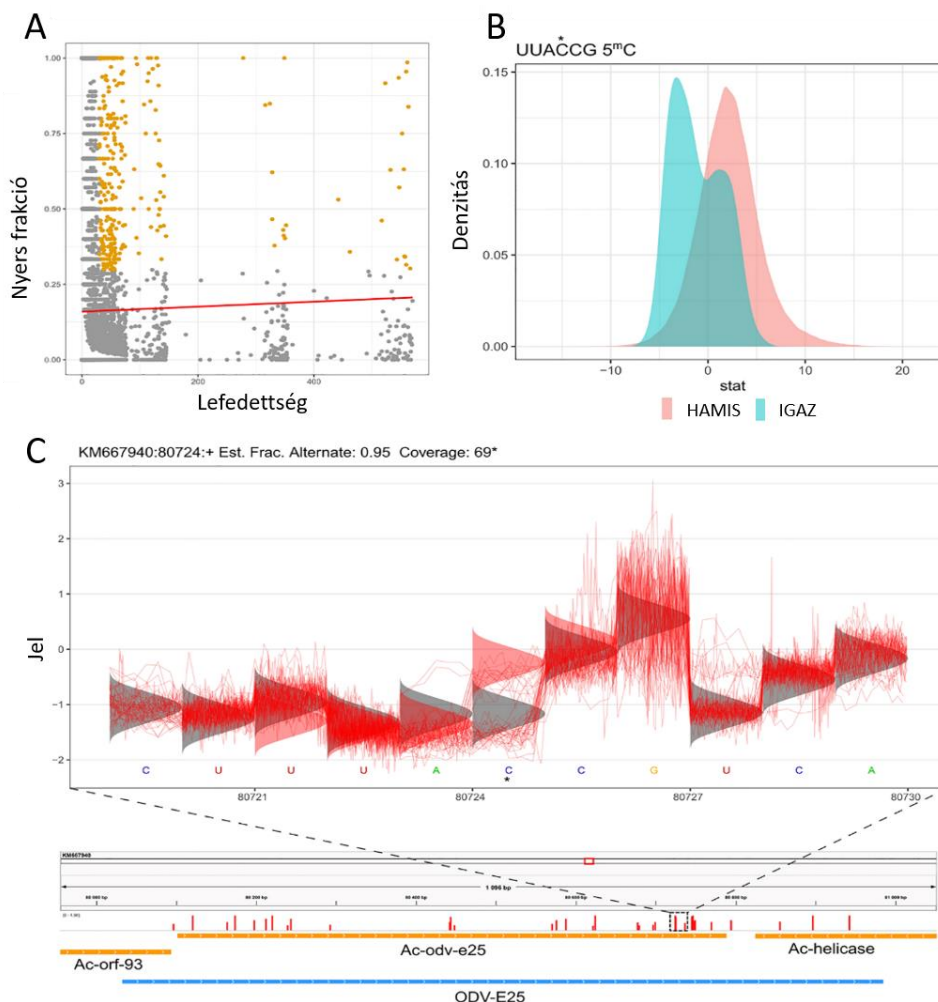
Az RNS-ben lévő 5-metilcitozin (5mC) egy poszttranszkripciók módosítás, amely hasonlóan működik, mint a DNS-ben található 5mC, de az RNS molekulák funkcióját és stabilitását befolyásolja. Az RNS-beli 5mC módosítás hozzájárul az RNS molekulák strukturális integritásához, befolyásolja az RNS splicingot, az RNS stabilitását, a fehérjeszintézist, és szerepet játszik az RNS közvetítette epigenetikai szabályozásban is. Az RNS-metiltranszferáz enzimek specifikusan metilálják a citozin nukleotidokat az RNS-ekben. Az RNS-metiláció széles körben elterjedt az élővilágban, és megtalálható mind az eukariótákban, mind a prokariótákban. Az RNS 5mC szerepét aktívan tanulmányozzák számos biológiai folyamatban, beleértve a sejt differenciációt, a stresszválaszt és a betegségek kialakulását. A kutatások arra is rávilágítottak, hogy az RNS 5mC szintje változhat különböző körülmények között, például fejlődés során vagy válaszul környezeti stresszre, ami szabályozza a gének kifejeződését és befolyásolja a sejt viselkedését.

Az A-tól I-ig (adenin-tól inozin) hipereditálás szintén egy poszt-transzkripciók módosítási folyamat⁴²², amely során az adenin nukleotidok inozinná alakulnak át az RNS molekulákban. Ezt az átalakítást az ADAR (adenosine deaminases acting on RNA) enzimcsalád végzi. Az inozin a szekvensspecifikus fehérjék által guanozinra (G) hasonlít, ezért az RNS-transzláció során gyakran G-ként értelmezik, ami változásokat eredményezhet az RNS molekulák szerkezetében és funkciójában. Az A-tól I-ig hipereditálás szerepet játszik számos biológiai folyamatban, beleértve a neurotransmissziót, az RNS splicingot és az immunrendszer működését. Ez a módosítás hozzájárul az RNS molekulák diverzitásához és komplexitásához, valamint fontos szerepet játszhat bizonyos betegségek kialakulásában is.

Az ONT dRNS-Seq módszere lehetővé teszi a nukleotid metiláció és egyéb, az RNS szerkesztés során megváltozott nukleotidok elemzését közvetlenül, a natív RNS-ek szekvenálása révén. A biszulfid konverziót alkalmazva, s az így átalakított RNS populációt Illumina könyvtárkészítéshez felhasználva, a metiláció azonban SRS módszerekkel is detektálható. Az AcMNPV példáján keresztül szemléltetem az RNS-ek módosulásaival kapcsolatos eredményeinket.

6.3.6.1 5mC metiláció – az epitranszkriptom vizsgálata

A dRNS-Seq nyers, illetve a biszulfid konverziós adatokat használtuk a transzkriptek metilációs helyeinek azonosításához. Az ONT leolvasásokat a Tombo szoftverrel elemeztük. A fals pozitív eredmények csökkentése érdekében kiszűrtük a 30-szoros lefedettséget el nem érő transzkripteket az analízisből, illetve azokat is, ahol a módosításokkal rendelkező readok frakciója nem érte el a 30%-ot. Valójában nem találtunk szignifikáns korrelációt a nyers frakcióban a lefedettség és a metilált nukleotidok között (**49. ábra, A**). Azonosítottunk egy lehetséges metilációs konszenzus szekvenciát (UUAC*CG, ahol a * a metilált C-re utal), ami a log-likelihood arányok helyes eloszlását mutatta (**49. ábra B**). A biszulfid-konverziós Illumina szekvenálás eredménye alátámasztotta a fenti konszenzus szekvencia metilációját. A kanonikus C helyektől való eltérés is egyértelműen észlelhető volt (**49. ábra C**). A potenciális fals pozitív helyek azonosítása után 325 lehetséges 5mC metilációs pozíciót kaptunk 12 vírusgénben (*ac-39k*, *ac-bro*, *ac-ctl*, *ac-odv-e25*, *ac-orf-58*, *ac-orf-73*, *ac-orf-74*, *ac-orf-75*, *ac-p40*, *ac-p6.9*, *ac-polyhedryn* és *ac-vp39*). A kanonikus C helyektől való eltérés itt is egyértelműen észlelhető volt (**49. ábra C**).

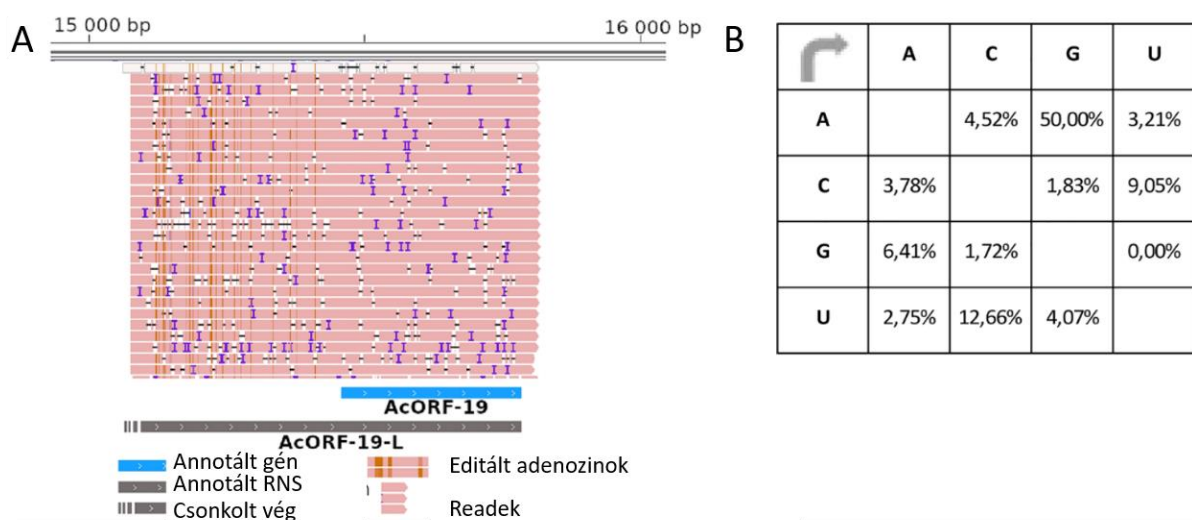


49. ábra Az AcMNPV RNS-ek 5-mC metilációja²³. (A) Nem figyelhető meg jelentős korreláció a lefedettség és a nyers frakcióban lévő metilált nukleotidok száma között. A sárga pontok a további elemzést igénylő pozíciókat jelölik. (B) A UUAC*CG szekvencia (metilált pozíciók csillaggal jelölve) tesztstatisztikái. Ez a panel a motívumot követő és nem követő helyek tesztstatisztikáinak eloszlását ábrázolja az összes megadott motívumra. (C) Egy potenciális metilált citozin az ODV-E25 transzkriptben, az ORF-en belül (5mC csillaggal jelölve). A piros görbék az elektromos jeleket jelzik, míg a sűrűségek az alternatív jel szinteket.

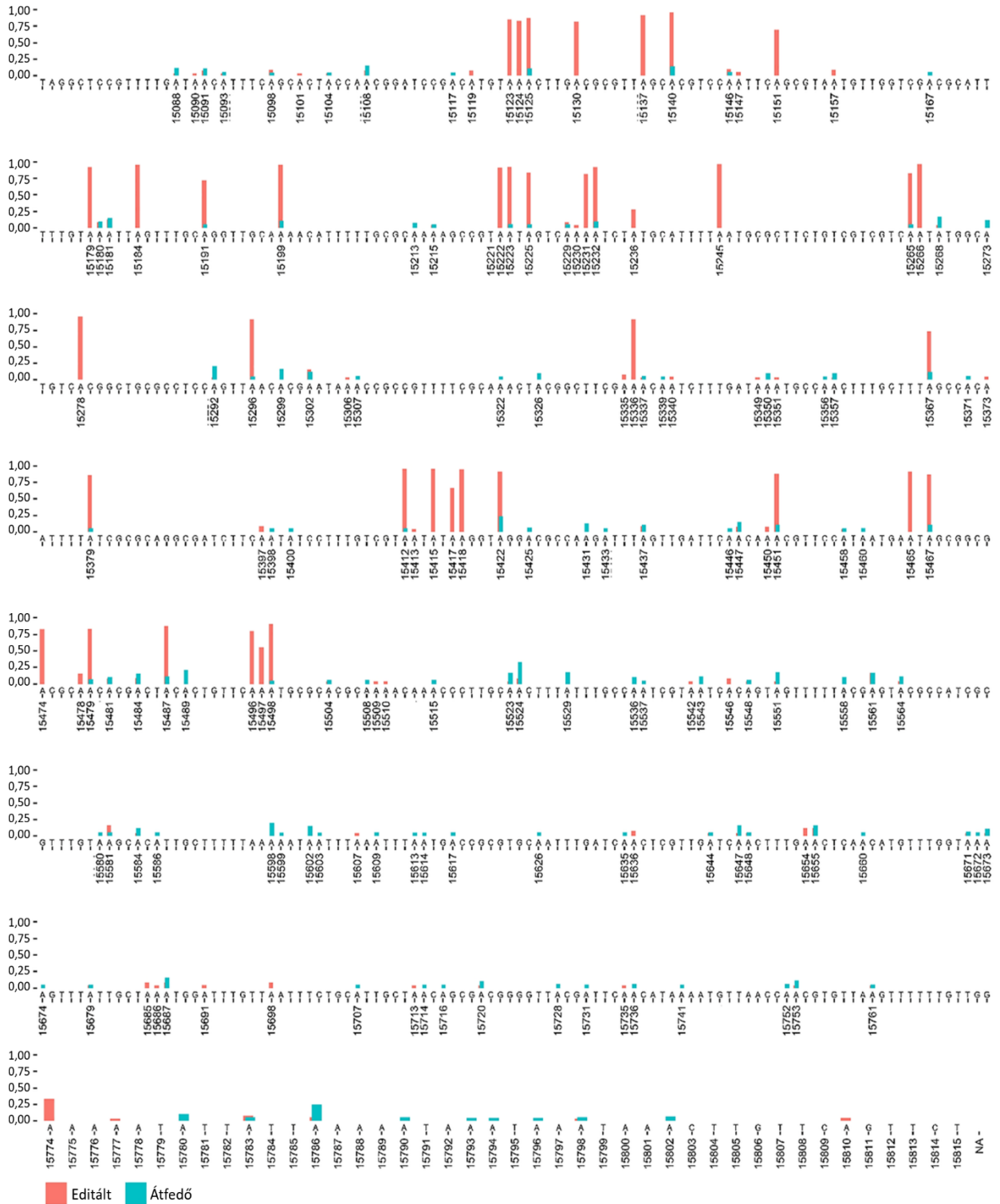
A dRNS-Seq-kell ellentétben, az Illumina biszulfid szekvenálás nagy lefedettséget eredményezett. A biszulfid konverzió értékeléséhez kontroll (konvertálatlan) minták szekvenálását is el kellett végezni. A Tombo-val azonosított 325 metilált pozícióból 234-et kimutattunk a biszulfid konvertált mintákban is. Itt a fals pozitívok elkerülése érdekében 25-szörös lefedettséget határoztunk meg minimumként. Összességében 7897 potenciális metilált pozíciót azonosítottunk, 99 génben. Az ac-Orf-12 transzkript 3'-UTR-ében 31 potenciális citozin pozíciót detektáltunk, mely minden esetben konvertált, azaz metilált volt. Az azonosított, potenciális metilált helyek (minimum 25-szörös lefedettséggel) 88%-a a gének ORF-jeiben van, míg 20%-uk UTR-ekben.

6.3.6.2 A-I hiperszerkesztés /hipereditálás

Az ORF19-L readek nagy százalékában detektáltunk A- I szubsztitúciót (valójában a szekvenáláskor A helyett G-t), mely az ezzel a génnel parallel átfedő, rövidebb readekben nem volt megfigyelhető. Megállapítottuk, hogy az összes szubsztitúció 50%-a A-ból G-be történő átalakulás volt (**50. ábra**), míg az átfedő readek esetén 16,9% ($p < 0,0001$, egyoldali Fisher-féle egzakt próba) (**51. ábra**). Beállítottunk egy 16,9%-os szubsztitúciós küszöbértéket a szekvenálás pontatlanságából eredő esetleges zaj kiszűrésére. Eredményeink szerint az ORF19-L minden adeninjének 18%-a magas szintet ($\bar{x} = 0.839$, $\sigma = 0.153$), míg az átfedő readek adeninjének 4%-a alacsony szintű A-tól G-ig történő átalakulást mutatott ($\bar{x} = 0.224$, $\sigma = 0.051$). Az ADAR által felismert lehetséges szerkesztési motívum jelenlétének azonosítása érdekében kiszámítottuk a szerkesztett A körüli ± 5 nt bázisfrekvenciáját. Korábban kimutatták, hogy egy G-ben gazdag környezet és egy 'felfelé irányuló' U stabilizálja az RNS-ADAR komplexet az emlős sejtekben⁴²³. Jelentősen magasabb U frekvenciát észleltünk közvetlenül a szerkesztett bázis előtt ($\chi^2(1, N = 79,455) = 79,338.023$, $p < 0.01$), míg a G frekvenciája csak enyhén volt magasabb ($\chi^2(1, N = 79,454) = 79,340.021$, $p < 0.05$) a szerkesztett bázistól downstream a +5 pozíciónál.



50. ábra Az AcORF-19 transzkript hiperszerkesztése²³ (A) A-tól I-ig hiperszerkesztett AcORF-19 readek. Az ábra az AcORF19-L transzkript leolvasásait mutatja (rózsaszín sávok). A barna vonalak a readekben bázismódosításokat jelölik. Ezek a módosítások a kanonikus RNS 5'-UTR-jében található. (B) Az AcORF-19 readek szubsztitúciós mátrixa. A referencianukleotidok a szubsztitúciós mátrix bal oldalán található, míg a mátrix jobb oldala a nukleotidok százalékos arányát tartalmazza. Látható, hogy több mint 50%-uk G betű.



51. ábra Az AcORF-19 transzkript hiper-szerkesztése²³. Az AcORF-19 RNS-en az A-tól G-ig történő szubsztitúció helyei és gyakorisága (a transzkriptet kódoló genomi régió jelölve). A magas gyakoriságú szubsztitúciók az A-tól G-ig történő szerkesztési eseményeket jelzik, míg az alacsony gyakoriságú szubsztitúciók a szekvenálási hibákra utalnak. A piros szín a magas, míg a kék szín az alacsony szerkesztési gyakoriságot jelöli.

6.3.7 A géndelációk hatása a virális transzkriptomra

A homológ rekombinációval készített mutáns PRV-kben a kiütött gén teljes virális transzkripcióra gyakorolt hatását qRT-PCR technikával vizsgáltuk. Az elemzéshez minden esetben ugyanazon körülmények között a vad típusú vírus transzkriptomot is vizsgáltuk, kontrollként. Az alábbiakban három kulcsfontosságú gén deléciója okozta fontosabb hatások egy szeptetét foglalom össze.

6.3.7.1 *u154*Δ

A PRV *u154* génje az E gének közé tartozik, míg homológja, az *icp27* a HSV-ben IE gén. Az ICP27 és homológjainak számos funkciója ismert, többek között a transzkripció⁴²⁴, és a DNS replikáció⁴²⁵ szabályozása, a gazdasejt fehérjeszintézisének leállítása, a poliadenilációs helyek használata⁴²⁶. Kimutatták, hogy az *u154* gén deléciója súlyos replikációs hibákhoz vezet⁴²⁷. Egy korábbi tanulmány⁴²⁸ arról számolt be, hogy az *u154* gén és fehérjeterméke nem alapvető a PRV növekedéséhez és replikációjához szövettenyészetben; azonban a mutáns vírus csökkent növekedési képességet mutatott. Azt is kimutatták, hogy ez a multifunkcionális fehérje nem elengedhetetlen a gazdasejt leállításához, és hogy hiánya kései fehérjék felhalmozódását okozza a fertőzés korai szakaszában, sejttípustól függő módon⁴²⁸. Az *u154* gén a vírus életciklusában betöltött szerepe alig ismert.

Eredményeink szerint a fertőzést követő 1 h-ban ez E gének expressziója jóval alacsonyabb a mutáció hatására, mint a *wt*-ban, míg 2 h-val a fertőzést követően gyakorlatilag egyforma a kifejeződés mértéke a két genom esetén (**52. ábra**). Ellenben az E/L és az L génekkel, amelyek ugyanebben az időszakban túl-expresszálódnak a mutánsban. Ez utóbbi eredmény összhangban van Schwartz és munkatársai megfigyelésével, akik a késői vírustermékek felhalmozódását mutatták ki a fertőzés E szakaszában⁴²⁸. A transzkriptek szintjének általános csökkenését figyeltük meg 4 h-kor, ami összefüggésben lehet a DNS-szintézis kezdetének a két genotípusra gyakorolt különböző hatásával. A különböző kinetikai osztályokba tartozó gének 6 h-kor eltérően viselkednek; jelentős esés figyelhető meg az L gének expressziója mértékében, míg adataink alapján az E génekre ekkor nincs hatással a mutáció. Végül, a PRV gének minden kinetikai osztályába tartozó gén jelentősen represszálttá válik 8 h-ra. Kiemelném, hogy a fertőzés után a 4. h-ban, amikor minden gén expressziója jelentősen csökkent a mutánsban a *wt*-hoz viszonyítva, az egyetlen kivétel a LAT, mely közel 8-szoros növekedést mutat. A vizsgálati periódus egésze alatt a legnagyobb génexpressziós növekedést a mutáns genom *u153* génjénél detektáltuk, (8,25-szeres) 6 h-kor, s ez a gén az egyetlen, amely 8 h-kor abundánsabb az *u154*Δ vírusban, mint a *wt* háttérében. Ezt a jelenséget azzal magyarázhatjuk, hogy az *u153* és a deletált *u154* gének egymás mellett helyezkednek el a genomon, és az *u154* cisz- vagy transz-ható, elnyomó hatást gyakorolhat az *u153*-ra, ami a mutáns vírustól hiányzik. Azonban az *u152* gén, amely szintén koterminális RNS-eket termel az *u154*-gyel, kevésbé abundáns a mutánsban, ebben a fertőzési szakaszban.

R _r	1h	2h	4h	6h	8h
\bar{E}	0,51	1,05	0,59	1,03	0,41
\bar{E}/\bar{L}	1,76	1,87	0,47	2,14	0,57
\bar{L}	1,61	1,56	0,34	0,50	0,30

52. ábra A kinetikai osztályokba tartozó gének átlagos R értékei arányának (az R_r értékek) hőtéskép-szerű ábrája²⁸⁰. A piros szín azokat az értékeket mutatja, ahol a mutáns vírusban az R érték magasabb, mint a *wt*-ban, míg a fekete fordítva.

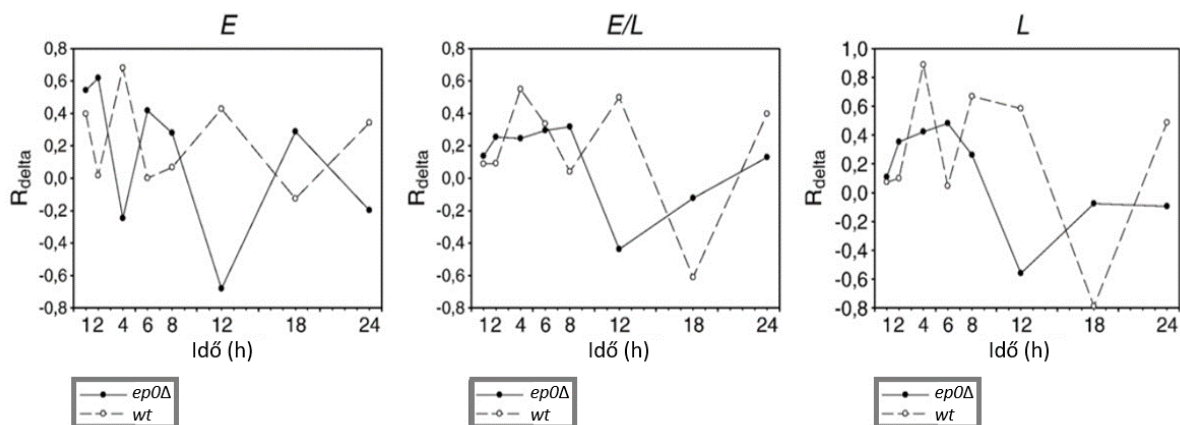
6.3.7.2 *ep0Δ*

Az *ep0* a PRV-ben egy korai gén, míg a HSV-1-ben a homológja (*icp0*) az IE kinetikai osztályba tartozik. Az EPO egy transzaktivátor, de kimutatták, hogy többek között a *vhs* promóter aktivitását is elnyomja⁴²⁹.

Azt találtuk, hogy a PRV gének átlagosan körülbelül 1,5-szer magasabb szinten íródnak le az *ep0Δ* vírusban, mint a *wt*-ban a fertőzést követő első órában, a legnagyobb mértékű hatást az E génekre gyakorolja. Az *ep0Δ* és a *wt* vírus közötti legjelentősebb génexpresszióbeli különbségeket az első 2 órában és a fertőzés késői szakaszában (12-től 24 óráig) figyeltünk meg. A 0-tól 2 óráig terjedő időszakban az E gének átlagosan jelentősen nagyobb mértékben fejeződnek ki a mutánsban, mint a *wt* vírusban, ami arra utalhat, hogy a tegument EPO fehérjének szuppresszív szerepe van a vírusgén-expresszióban a fertőzés e szakaszában.

Továbbá, az *ep0Δ* esetében a vírusgén (és különösen az L gének) expressziója jóval kisebb mértékű a *wt*-hoz viszonyítva, a 12-től 24 óráig tartó időszakban, ami arra utal, hogy az EPO fehérje stimuláló szerepet játszik ebben a fertőzési szakaszban. Az *ep0* gén deléciójának egy másik jellemző jelensége, hogy a mutáns vírus E transzkript szintjei jelentősen csökkennek 4 h-nál a 2 h-hoz képest, ami nem igaz az E/L és L génekre. Ezt a jelenséget azzal magyarázzuk, hogy a tegumentből származó EPO fehérje kicserélődik a *de novo* EPO-ra 4 h-ra, ami elősegítheti az E gének transzkripcióját. A két vírus expressziós görbéi közti jelentős különbség, hogy a mutáns vírus transzkript szintjei 6 h-nál csúcsosodnak, míg a *wt* vírus E/L és L génjei 8 h-nál, az E gének pedig 24 h-nál (helyi maximumok 2 és 8 h-nál).

Eredményeink szerint az E transzkriptek nettó változásának dinamikája (\bar{R}_d) szinte tökéletesen inverz mintát mutat a két genetikai háttérben, ami nem igaz az E/L és L génekre (53. ábra).



53. ábra. Az E, E/L és L gének átlagos R_d értékei időbeli változásai a *wt* és az *ep0*-mutáns vírusban²⁷⁹. Az E gének átlagos R_d expressziós görbéi fordított (inverz) lefutásúak a két vírusban, a másik két osztályban nincs ilyen egyértelmű mintázat. Eredményeink alapján az EPO fehérje egységesen stimuláló hatása az összes kinetikai osztályba tartozó gének expressziójára, a 12–24-h közti fertőzési periódusban.

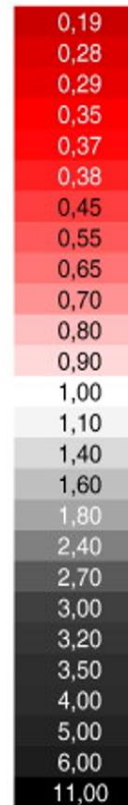
6.3.7.3 *ul41Δ*

A virion host shutoff (VHS) fehérje, amelyet a PRV *ul41* gén kódol, egy mRNS-specifikus endoribonukleáz⁴³⁰. Az *ul41* kései gén, amely beépül a tegumentbe^{431,432}. Korábban kimutattuk, hogy VHS fehérje nem alapvető jelentőségű a vírus működése szempontjából, mivel a vhs-knockout vírus megőrzi a növekedési képességét mind tenyésztett sejteken, mind pedig *in vivo*³⁵. Azonban a VHS deléciója 5-10-szeres csökkenést eredményez a vírushozamban sejt kultúrában^{431,433}, valamint jelentős virulencia-vesztést okoz⁴³⁴, ami arra utal, hogy a VHS fontos szerepet játszik a vírusfertőzésben^{435,436}. A VHS fehérje a virális és gazdasejt mRNS-ek degradációját okozza⁴¹⁰. A VHS további funkciói: a vírus transzkripciójához szükséges ribonukleotidok biztosítása, a fertőzésre adott gazdaválaszok megakadályozása bizonyos sejtes gének, mint például a TNF-receptor 1 gén mRNS-einek eliminálásával⁴³¹.

Eredményeink szerint az E gének expressziója alapvetően megváltozik a *vhsΔ* vírusban. Azt tapasztaltuk, hogy a vizsgált PRV transzkriptek mindegyike alacsonyabb mértékben expresszálódik a *vhsΔ* esetén, mint a *wt*-nál ($R_r < 1$) 1 h-val a fertőzés után (**54. ábra**). Ennek a jelenségnek a magyarázata a következő. Habár a virális mRNS-ek is a VHS fehérje célpontjai^{2,437}, a celluláris mRNS-ek a vírus RNáz általi preferenciális degradációjára⁴³⁶ ebben a fertőzési szakaszban kompenzálhatja ezt a hatást két különböző módon: azáltal, hogy megszünteti a gazda és a vírus közötti versenyt a sejt translációs apparátusáért a vírus javára, és az újonnan szintetizált vírus RNS-ek nyersanyagaként szolgáló új ribonukleotidok előállításával. Eredményeink szerint ebben a fertőzési időpontban a VHS-nek nincs a PRV gének különböző kinetikai osztályaira nézve eltérő hatása. A fertőzés után 2 h-val a VHS gátló hatást fejt ki a PRV RNS-ekre mindhárom kinetikai osztályban (**54. ábra**), néhány kivétellel. Úgy tűnik tehát, hogy 1-től 2 h-ig átvált gátló hatásúvá a vírus mRNS-ekre nézve. A 2-től 4 h-ig tartó időszakban a VHS hatásának csökkenése figyelhető meg. Az E/L és L gének átlagos expressziós értékei a mutánsban gyakorlatilag megegyeznek a *wt* víruséval 4 h-kor, ellentétben az E transzkriptekkel, amelyek átlagosan nagyobb expressziót mutatnak a mutáns törzsben. Ezt a jelenséget a tegument VHS fehérje csökkenő szintjével magyarázzuk, amely egyenesen hat a vírusgének különböző kinetikai osztályaira, míg az újonnan szintetizált VHS fehérje szintje növekszik, s ennek úgy tűnik, hogy differenciált hatása van a különféle kinetikai osztályba tartozó génekre. Az E gének kifejeződése jelentősen magasabb a mutánsban, mint a *wt* vírusban 4-től 24 h-ig, ami 8 és 12 h-kor a legjelentősebb. Az L gének expressziója nem változik jelentősen a mutáns vírusban a *wt*-hoz képest 4 h-kor, egyes L gének expressziója még alacsonyabb a *vhsΔ*-ben, mint a *wt*-ben a fertőzés számos időpontjában, pl. az *ul1*, *ul5*, *ul10*, *ul19*, *ul44*, *ul48* és *us1* gének. Az E/L RNS-ek átlagos szintjei mindig az E és L szintek között esnek a 6-tól 24 h-ig tartó intervallumban. Négy és 6 h-kor nem figyelhető meg a VHS szelektív hatása az E/L és L kinetikai osztályú PRV géneken. A VHS fehérje az E/L és L transzkriptekre gyakorolt hatása hasonlóan alacsony az első 6 h alatt, míg ez nem igaz a vírus E génjeire. Az E gének expressziós értéke gyorsan nő a 4 h-tól 12 h-ig tartó időszakban, de 18 h-ra élesen csökken; azonban ezen transzkriptek mennyisége még ebben az időpontban is magasabb a *vhs*-deletált vírusban, mint a *wt* vírusban (18 h). Az E/L transzkriptek értéke minden időpontban nagyobb, mint 1, 6 és 18 h között, csúcspontját 8 h-nál közepes szinten érve el az E mRNS-ekhez képest. Adataink azt jelzik, hogy 24 h-nál nincs szelektív hatása a VHS fehérjének az E/L RNS-ekre. Az L gének transzkriptjeire csak enyhe hatást fejt ki a VHS fehérje, a 6-tól 24 h-ig terjedő időszakban. Kimutattuk, hogy VHS fehérje a legerősebb hatást az E és E/L génekre gyakorolja 8 és 12 h között.

Összességében a VHS fehérje nagy gátló hatást fejt ki az E gének RNS szintjeire ($r = 0.582$), közepes hatást az E/L ($r = 0.894$) génekre és alig van hatása az L génkifejeződésre ($r = 0.960$).

K	Gén	1h	2h	4h	6h	8h	12h	18h	24h
IE	<i>ie180</i>	0,529	0,951	0,849	1,346	3,172	3,379	2,007	1,580
E	<i>ep0</i>	0,400	1,184	1,871	5,192	6,064	11,070	3,749	3,532
E	<i>ul9</i>	0,398	1,582	1,087	1,111	1,586	1,999	0,976	1,067
E	<i>ul21</i>	0,779	3,144	1,779	2,393	2,629	5,605	1,823	2,400
E	<i>ul23</i>	0,665	1,219	0,827	1,988	4,876	3,886	1,745	2,969
E	<i>ul29</i>	0,567	1,743	1,331	1,754	5,507	3,306	1,348	1,196
E	<i>ul30</i>	0,646	4,154	2,277	1,835	3,660	3,987	1,627	1,147
E	<i>ul50</i>	0,385	3,337	1,038	1,829	4,614	2,509	3,517	1,933
E	<i>ul52</i>	0,505	2,071	1,957	2,111	4,142	2,585	1,121	0,541
E	<i>ul54</i>	0,294	1,893	0,917	1,774	3,986	3,566	1,836	2,633
E	<i>us3</i>	0,789	0,603	1,207	1,581	2,102	3,990	3,444	2,304
E/L	<i>ul14</i>	0,440	1,926	0,709	1,493	3,019	2,242	1,333	1,197
E/L	<i>ul15</i>	0,542	1,026	0,649	0,876	1,507	2,790	1,160	1,280
E/L	<i>ul20</i>	0,267	1,290	1,337	0,838	3,631	1,446	2,466	1,506
E/L	<i>ul43</i>	0,613	1,637	0,922	1,857	3,271	3,094	2,779	1,152
E/L	<i>ul53</i>	0,747	1,088	1,074	2,239	4,548	1,337	1,798	0,967
E/L	<i>us6</i>	0,418	3,170	0,301	0,531	0,527	0,675	0,697	0,908
E/L	<i>us9</i>	0,533	2,785	0,687	0,956	1,508	1,881	1,969	1,206
L	<i>ul1</i>	0,409	2,196	0,515	0,691	0,889	1,286	0,774	1,122
L	<i>ul5</i>	0,380	0,516	0,285	2,340	1,264	3,300	0,557	1,477
L	<i>ul6</i>	0,579	0,488	0,703	1,147	1,240	1,685	1,674	1,241
L	<i>ul10</i>	0,512	1,993	0,988	1,269	1,017	0,683	0,874	0,578
L	<i>ul17</i>	0,523	0,912	1,303	1,526	1,410	1,754	1,377	1,173
L	<i>ul19</i>	0,350	2,060	0,700	1,021	1,150	1,562	0,310	0,261
L	<i>ul22</i>	0,837	2,176	1,987	0,874	3,065	1,530	2,274	2,032
L	<i>ul24</i>	0,518	4,899	0,848	1,655	2,360	2,051	0,971	1,179
L	<i>ul32</i>	0,945	2,390	0,971	1,613	1,199	1,247	0,965	1,024
L	<i>ul33</i>	0,535	4,466	1,466	1,167	2,023	1,456	1,388	1,171
L	<i>ul36</i>	0,463	3,598	1,567	1,812	1,708	1,043	0,867	0,837
L	<i>ul37</i>	0,390	2,683	0,952	1,588	2,594	2,291	0,666	0,670
L	<i>ul42</i>	0,498	1,284	0,562	0,869	0,781	1,239	0,865	1,561
L	<i>ul44</i>	0,250	1,396	0,397	0,375	0,566	0,446	1,141	0,650
L	<i>ul48</i>	0,362	1,881	0,646	0,679	1,059	0,976	0,857	0,638
L	<i>ul49,5</i>	0,199	1,763	0,721	0,753	0,690	1,557	1,329	0,755
L	<i>ul51</i>	0,474	1,440	1,047	4,190	2,367	1,650	1,107	1,473
L	<i>us1</i>	0,250	0,941	0,645	0,815	0,752	2,333	2,035	1,191
	<i>AST</i>	0,288	1,298	0,409	0,776	1,087	1,229	1,176	0,997
	<i>LAT</i>	0,376	2,553	0,415	0,934	1,176	2,495	0,736	0,618



54. ábra A VHS mutáns és a wt vírus transzkriptek R értékei aránya hőterkép-szerű megjelenítése²⁸¹. A géneket a kinetikai osztályaik szerint csoportosítottuk. Az ábra az E és E/L gének expressziós szintjében bekövetkező drámai mértékű emelkedést mutat 8 és 12 h-nál, a mutáns vírusban. Színkód: fekete: legnagyobb Rr érték, sötétpiros: legkisebb Rr érték, fehér: 1-hez közeli Rr érték.

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

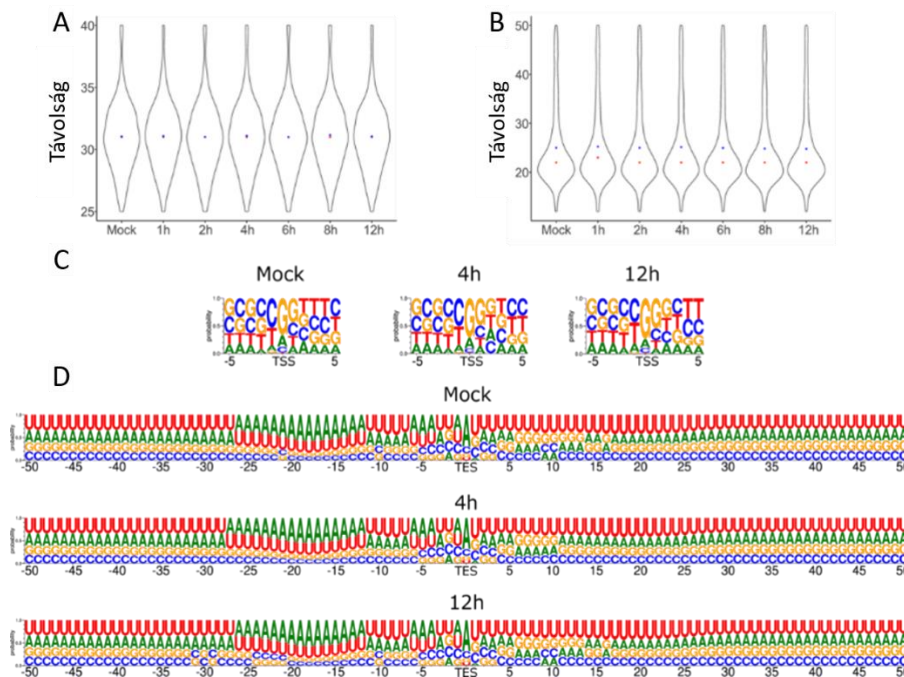
Számos vírus esetén elvégeztük a gazdasejtek expressziójára gyakorolt hatásuk elemzését is (pl. BoHV-1, EHV-1, VACV, VSIV, SARS-CoV-2). A disszertációban részletesen a BoHV-1 a *Bos taurus* transzkriptomra kifejtett hatásával kapcsolatos eredményeinkre fókuszálunk, továbbá a VACV gazdasejtjeinek a vírusfertőzés következtében megváltozó génexpressziós profilját mutatom be.

6.4.1 A *Bos taurus* transzkriptom és BoHV-1 fertőzés hatására bekövetkező változásai

BoHV-1 fertőzött sejtekből az alábbi technikákkal készítettünk könyvtárakat: az ONT dcDNS-Seq, ONT amplifikált cDNS szekvenálás random primerekkel és Illumina SLRS. Mindhárom technikát alkalmaztuk a szarvasmarha transzkriptom annotációjához, míg a dcDNS-Seq-et a BoHV-1 gazdasejt génexpresszióra gyakorolt hatásának időbeli változó elemzésére.

***Bos taurus* transzkriptek annotálása:** A transzkriptek detektálását és annotálását a LoRTIA szoftverrel végeztük. Az intronok, a TSS-ek és a TES-ek annotációjához azt a kritériumot állítottuk fel, hogy ezeket a szekvenciákat legalább két független mintában meghatározza a LoRTIA. Így összesen 11025 TSS-t, 21317 TES-t és 139771 intront azonosítottunk. Továbbá, a LoRTIA összesen 227672 *B. taurus* transzkriptet határozott meg, melyek medián hossza 1678 nt.

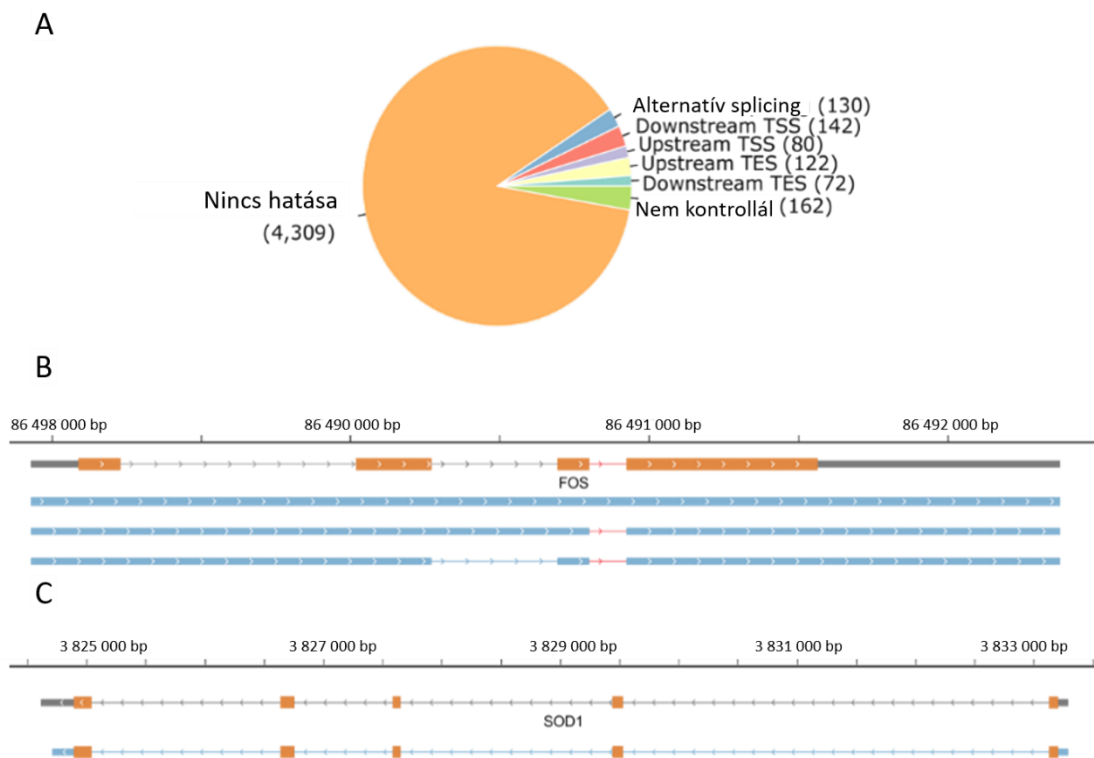
A konszenzus TATA boxokat a gének TSS-ei előtt, azoktól átlagosan 31,15 nt távolságban azonosítottuk, míg PAS-okat átlagosan 25,35 nt távolságra a gazdagének TES-eitől. Adataink azt mutatják, hogy a vírusfertőzés nem okoz jelentős változást a promóterek és a TSS-ek, valamint a PAS-ok és a TES-ek közötti távolságban (**55. ábra, A, B**). Nem találtunk szignifikáns módosulást a TSS ± 5 nt környező régiójában és a szarvasmarha gén TES-ek ± 50 nt környező régiójában a fertőzés során (**55. ábra, C, D**).



55. ábra A vírusfertőzés hatása a gazda transzkriptek TSS-eire és TES-eire²⁸³. (A) A TATA-boxok távolsága a TSS-ektől és (B) a poliadenilációs szignálók távolsága a TES-ektől bázispárokban. A (C) TSS-ek és a (D) TES-ek szekvenciamotívuma a mock-fertőzött sejtekben, ill. 4 és 12 h-val a fertőzés után.

A fertőzés során a gazdasejtekben bekövetkező splicing, TSS és TES változások értékeléséhez a LoRTIA által azonosított transzkriptek közül azokat vontuk be az analízisbe, melyekből több mint tíz leolvasást kaptunk. Összesen 130 transzkriptet detektáltunk, melyben alternatív splicing figyelhető meg (**56. ábra, A**).

Ismert, hogy a FOS fehérje (proto-onkogén), amely a stressz jelátviteli útvonal 'azonnali reagálója', gyorsan lebomlik, ha a harmadik intronja megmarad⁴³⁸. Kimutattuk a FOS transzkript nem splicingolt változatát nagyon alacsony gyakorisággal, és az RNS további, a fent említett exont nem tartalmazó splice variánsait, amelyek a fertőzés első órájától kezdve jelen voltak (**56. ábra, B**). Ez megerősíti a FOS a vírusfertőzés korai időpontjaiban való jelenlétéről szóló korábbi eredményeket^{439,440}.



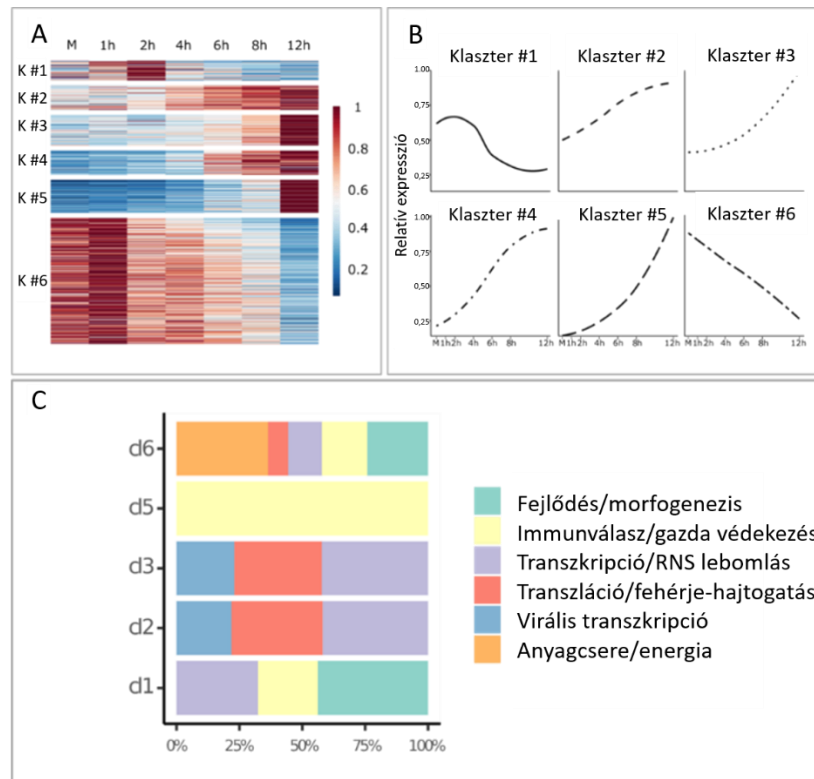
56. ábra **A vírusfertőzés hatása a gazdasejtek transzkript izoformáira**²⁸³. (A) A fertőzött sejtekben a mock sejtekhez képest összesen 546 olyan TI-t detektáltunk, melynek legalább tíz readje van. (B) A szürke téglalapok a korábban kimutatott TI-kat képviselik. A kék téglalapok az FOS transzkript egy új, nem spliceolt és két új spliceolt változatát jelzik. Az intronokat a téglalapok közötti vonalak ábrázolják. Az intron, amely a FOS megnövekedett lebomlási sebességéért felelős, pirossal látható. A narancssárga téglalapok az ORF-et jelölik. (C) A kék téglalapok a SOD1 új TES izoformáját, a vonalak az intronokat, a narancssárga téglalapok pedig az ORF-et ábrázolják. Egy upstream TES-t tartalmazó SOD1 izoforma a fertőzés során kimutatható, de a mock-mintákban nem azonosítható.

A gének 3'-UTR-jeiben gyakran előfordulnak miRNS-célpontok, hozzájárulva az mRNS lebomlásához. Így a rövidebb 3'-UTR a transzkript stabilitásának növekedéséhez vezethet, míg a hosszabb 3'-UTR-eket több miRNS és egyéb transz-ható elem is megcélozhatja, ezáltal eltérő szabályozási mintákat generálhat⁴⁴¹. Adataink elemzése során 72 olyan RNS-t észleltünk, amelyekben a TES-ek downstream helyezkedtek el, és 122 olyan variánst, amelyekben a TES-ek inkább upstream helyezkedtek el, mint a mock mintákban. A szuperoxid-dizmutáz 1 (SOD1) védelmet nyújt az oxidatív károsodások ellen⁴⁴², beleértve az interferon (IFN-I) jelátvitel által

kiváltott károsodásokat is⁴⁴³. A fertőzött sejtekben kimutatott SOD1 3'-UTR izoformája rövidebb volt, mint a mock sejtekben (56. ábra, C).

Egy korábbi tanulmányban kimutatták, hogy a HSV-1 fertőzés hatására a gazdaszervezetben a transzkripció termináció zavart szenved, s ez kiterjedt transzkripció átfedéseket eredményezett a szomszédos géntermékek között³⁶⁸. Eredményeink szerint a poliadenilált transzkriptek hossza jelentősen megváltozott a BoHV-1 fertőzés második órájától kezdődően. A RNS-ek hosszának legkifejezettebb növekedése 2 óra (+ 134 bp változás), míg a legnagyobb csökkenést 12 h-nál észleltük (- 193 bp) a mock-hoz képest.

Gazdagén expresszió változása a vírusfertőzés során: összesen 8342 gazdagént azonosítottunk, amelyek tíznél több transzkriptet produkáltak mindhárom biológiai párhuzamosban. Differenciál expressziós (DE) analízist alkalmazva 0,01-es hamis felfedezési arány (FDR) küszöbértékkel, a 8342 gazdagén közül 686 olyan gént azonosítottunk, amelyeknek jelentősen megváltozott az expressziós szintje vírusfertőzés hatására. A géneket expressziós profiljuk (és nem abszolút expressziós szintjük) alapján csoportosítottuk. Az elemzés ezen részében az expressziós szintek idősorait egy relatív skálára transzformáltuk, amely a mintavételi pontok közötti expressziós változásokat reprezentálja. Ez lehetővé tette a gének csoportosítását expressziós profiljuk alapján a vírusfertőzés során az abszolút abundanciájuk helyett. Hat génklasztert azonosítottunk jellegzetes expressziós profillal (57. ábra). Négy olyan géncsoportot (2–5. klaszter) azonosítottunk, amelyek állandóan overexpresszáltak voltak, azaz nagyobb mértékben fejeződtek ki a fertőzés minden időpontjában, mint a mock sejtek esetén, egy olyan géncsoportot, ahol az expressziós szintek folyamatosan csökkentek a vírusfertőzés teljes időszaka alatt (6. klaszter), és végül egy csoport, amely kezdetben overexpressziót, majd downregulációt mutatott (1. klaszter).



57. ábra **A gazdagének differenciál expressziós analízise**²⁸³. (A) A hőtérkép a hat azonosított, elkülönülő klaszter relatív expressziós értékét mutatja. (B) A hat klaszter normalizált átlagos relatív expressziós értéke. (C) A gén otológia (GO) szerint a gének eloszlása a hat funkcionális kategóriával társítva. K: klaszter

Overreprezentációs elemzést végeztünk a PANTHER szoftverrel. A felülreprezentált géneket a Gene Ontology (azaz gén-ontológia; GO) adatbázis szerint hat funkcionális csoportba soroltuk (**57. ábra**): 296 gén játszik szerepet a sejtanyagcserében, 257 részt vesz a transzkripcióban és az RNS lebomlásában, 242 a fejlődési és morfogenetikai folyamatokban, 187 a sejtmetabolizmusban, immunválaszban, 161 a translációban és a fehérje feltekeredésében, míg 61 gén specifikusan kapcsolódik a vírustranszkripcióval kapcsolatos folyamatokhoz.

Az első klaszter génjei ($n = 53$) közepes expressziót mutattak a fertőzést megelőzően (amely átmenetileg enyhén overexpresszált volt az 1 és 2 h időpontokban), majd a későbbi időpontokban downreguláció következett. Ebben a klaszterben a gének felülreprezentáltak a fejlődési és morfogenetikai folyamatok széles skáláját irányító útvonalakban. Számos, ebben a klaszterben jelenlévő transzkripciót szabályozó fehérjéket kódoló gén a fertőzés során csökkenő expressziót mutat. Az immunválasz és a gyulladáshoz kapcsolódó folyamatok citokin szabályozásában részt vevő gének is érintettek.

A második és a harmadik géncsoport ($n = 64$, $n = 82$) közepes expresszióval rendelkezett a fertőzést megelőzően, ami minden egymást követő időpontban emelkedett. Ezeknek a klasztereknek a génjei felülreprezentáltak olyan funkciókban és molekuláris folyamatokban, mint amelyek összefüggésbe hozhatók a virális génexpresszióval és a virion-összeszereléssel. A transzkripció és transláció folyamatokban részt vevő gének túlműködése, valamint az RNS lebomlása is megfigyelhető volt. Az RNS lebomlása lehet a gazdasejt azonnali válasza a virális transzkriptok felhalmozódásának ellensúlyozására, vagy a vírus arra irányuló erőfeszítése, hogy eliminálja a versengő gazda-mRNS-eket, hogy elősegítse a vírus RNS-ek translációját^{444,445}.

A negyedik klaszterbe tartozó gének ($n = 64$) relatív expressziója alacsony volt a fertőzést megelőzően. Ezek a gének a fertőzés során szigma alakú görbét követően felfelé szabályozottak, azonban ezek a gének nem voltak jelentősen felülreprezentáltak a GO molekuláris funkcióiban vagy a GO biológiai folyamataiban.

Az ötödik géncsoport ($n = 88$) a fertőzést megelőzően nulla vagy elhanyagolható expressziót mutatott, de exponenciális expresszió növekedés következett be a fertőzés során. Ennek a klaszternek a felülreprezentált génjei vírusellenes sejtes és védekező válaszokkal, például az I. típusú interferon jelátviteli útvonalakkal voltak összefüggésben.

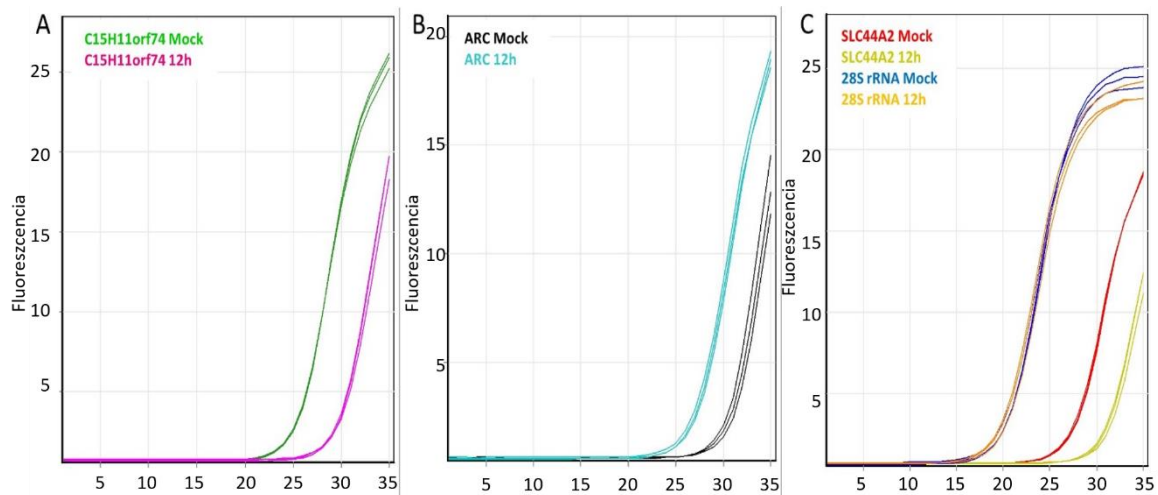
A hatodik klaszter génjei ($n = 335$) a vírust fertőzést megelőzően magas expressziójú gazdagének csoportja, amelyek a fertőzés során éles downregulációt mutattak. Ezek a gének felülreprezentáltak a fehérje feltekeredésével, a sejtciklus szabályozásával és a mitokondriális folyamatokkal kapcsolatosan, beleértve az aerob légzést is.

„Key response” gazdagének

A gazdasejtek azonnali válaszána elemzése céljából DE analízist végeztünk a Mock és az 1h expressziós értékeken. A három biológiai párhuzamosban hat szarvasmarha-gént azonosítottunk, amelyek expressziója szignifikánsan csökkent, valamint 19 gént, amelyek szignifikánsan túlexpresszáltak. A overexpressziós elemzés nem mutatott ki szignifikáns összefüggést sem a GO biológiai folyamataival, sem a GO molekuláris funkcióival, sem a túl-, sem az alulszabályozott gének részalmazásával, sem pedig a gének teljes készletével. A STRING asszociációs elemzés azonban négy hálózatot tárt fel e gének között. Az első génhálózat (GADD45B, GADD45A, DDIT3, ATF3, IFRD1, CARM1, SQSTM1) olyan géneket tartalmaz, amelyek a gazdaszervezet DNS-

károsodási válaszához, a transzkripció szabályozásához kapcsolódnak. Ezenkívül egy gén szerepet játszik a szelektív automakrofágiában. A második hálózat két interferon gamma-stimulált génből áll; Az IRF9, egy TF, amely alapvető szerepet játszik a vírusellenes aktivitásban, és az MT2A, egy metallotionein fehérje. A harmadik hálózat két génből áll (SRSF5 és HNRNPDL), amelyek a pre-mRNS-feldolgozáshoz, szállításhoz és splicing szabályozáshoz kapcsolódnak. A hematopoetikus sejteket szabályozó IL11 citokin a negyedik hálózat része volt. Azt találtuk, hogy az IL11 alulszabályozott. Ezzel szemben a CXCL5, a neutrofilek aktiválásával kapcsolatos gén, amely a 4-es hálózatban is jelen van, a vírusfertőzést követően túlszabályozott. A fennmaradó négy a hat alulszabályozott gén közül (LASP1, HDAC7, SLC44A2 HSPG2) és nyolc a 19 túlreprezentált/felfelé szabályozott gén közül (ID2, HMG3, TMEM190, TSC22D1, PRKAR2B, LOC100847759, LOC100847143, LOC100174924) a fejlődésben résztvevő gén.

Tíz gazdagén a vírusfertőzés hatására megváltozott expresszióját validáltuk qRT-PCR-rel. Az adott gének expresszióját két különböző sejtpopulációban mértük: mock és fertőzött sejtekben. A qPCR adatok igazolják azon eredményeinket, amelyeket LRS-megközelítéssel értünk el minden egyes vizsgált gén esetében. Megállapítottuk, hogy az ARC, DDIT3, FAM102A, GADD45G, MYLIP és TSC22D1 gének expressziója nőtt a vírussal fertőzött sejtekben, míg az ANXA1, C15H11orf74, LOC112446408 és SLC44A2 gének expressziós szintje csökkent a mock-fertőzött sejtekhez képest (58. ábra).



58. ábra A kiválasztott *Bos taurus* gének qPCR amplifikációs görbéi²⁸³. (A) A C15H11orf74 gén qPCR görbéi. BoHV-1 fertőzés drasztikus gátló hatással van e gén expressziójára. (B) Az ARC gén expressziója a fertőzést követően szignifikánsan megemelkedik. (C) Ez az ábra egyrészt a 28S rRNS háztartási gén (a vírusfertőzés nincs az expressziójára hatással), másrészt a SLC44A2 gén (a vírusfertőzés hatására csökken az expressziója) qPCR görbéit mutatja.

A gazdasejtek génexpressziós elemzése során kapott eredményeket *in silico* megközelítéssel is validáltuk. 30 háztartási gént választottunk ki, amelyek nagyon stabil expressziót mutatnak különböző szövetekben⁴⁴⁶. Bár ezeknek a géneknek a mennyisége változó, várható, hogy a normalizált adatok szerint expressziós szintjeik nagyon magas és szignifikáns korrelációt mutatnak a biológiai párhuzamosok és a különböző időpontokban lévő minták között. Kiszámoltuk a korrelációs együtthatót (R) és az egyes biológiai ismétlődések és mérési pontok közötti szignifikanciát (részletes adatok a publikációnk²⁸³ S3 kiegészítő táblázatában és S3 kiegészítő ábráján). Az a tény, hogy a háztartási (housekeeping) gének átírási arányai nagy szignifikanciával korrelálnak, a szekvenálási technikánk és az adatnormalizációs protokollunk megbízhatóságának belső megerősítése.

6.4.2 A VACV fertőzés hatása a CV-1 sejtek génexpressziós profiljára

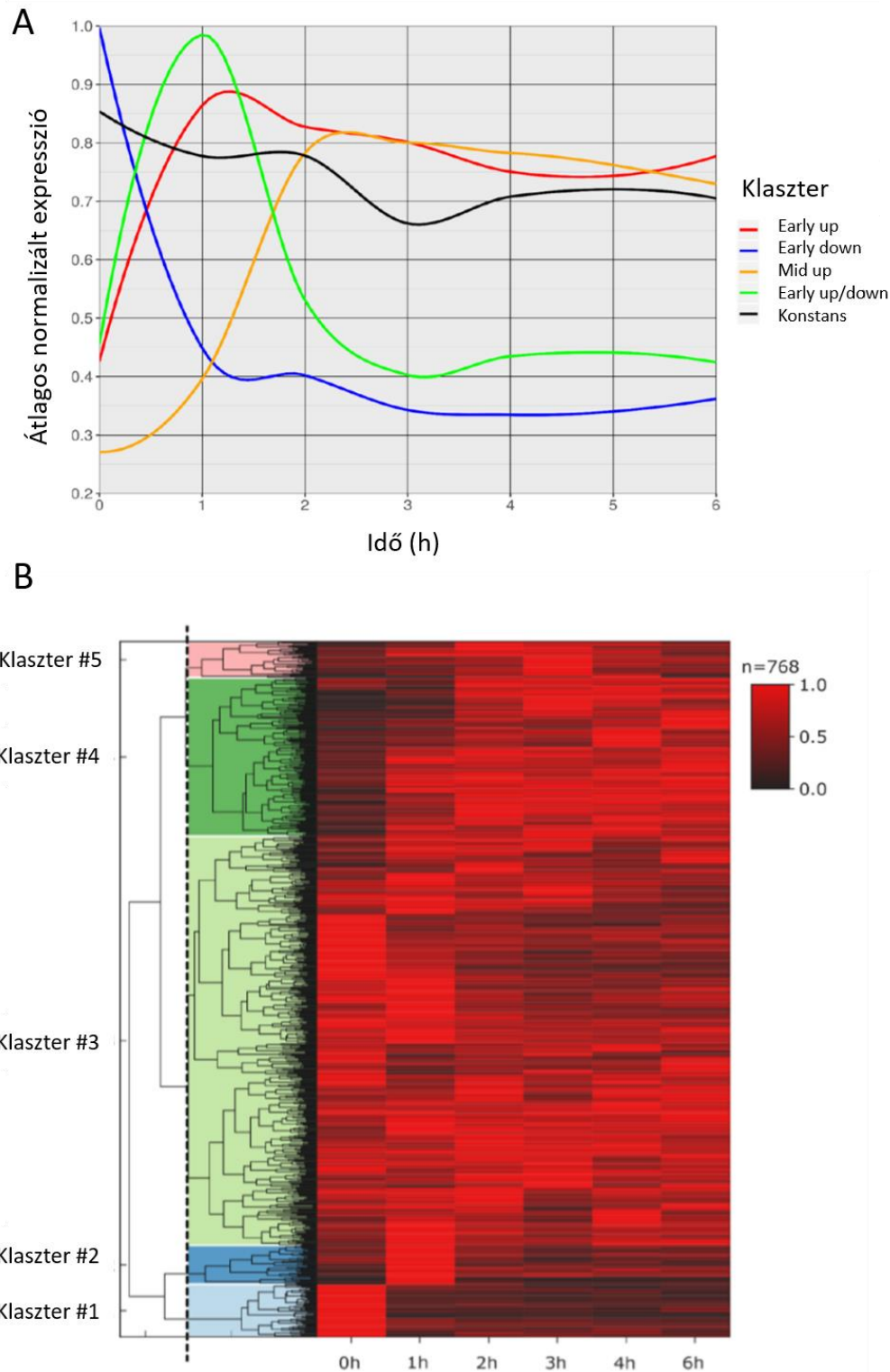
A VACV fertőzött és a mock-fertőzött sejtekből a PacBio Sequel platformjára amplifikált Iso-Seq és az ONT MinION platformjára amplifikált cDNS könyvtárakat készítettünk.

A *C. sabaeus* transzkriptom: a LoRTIA programmal 478 TSS-t és 2011 TES-t, valamint 24574 splice junctiont annotáltunk, annak a kritériumnak a figyelembe vételével, hogy mindegyiket legalább tíz leolvasás erősítsen meg. A TSS-ektől upstream 43 kanonikus CAAT boxot (átlagosan, 104,913 nt távolságra), 880 kanonikus GC boxot (átlagosan 33,374 nt távolságra), illetve 80 kanonikus TATA-boxot (átlagosan 31,13 nt távolságra) azonosítottunk. Egy korábbi publikációban leírták, hogy az iniciátor szekvenciák nagyobb valószínűséggel vannak jelen a humán gének TSS-ei környékén, ha nincsenek a TSS-től upstream TATA boxok⁴⁴⁷. Eredményünk ezt alátámasztja, a BBCABW (B = C/G/T, W = A/T) iniciátor konszenzus szekvencia sokkal gyakoribb a TATA-box nélküli gének TSS-ei körül, mint a TATA boxszal rendelkezőknél.

Összesen 1849, kanonikus upstream poliA szignállal rendelkező TES-t azonosítottunk (távolság átlagosan 26,681 nt), a ± 50 -nt-os upstream régió U-gazdag, míg a downstream régió G/U-gazdag. Továbbá, 12287 intront annotáltunk, melyek közül 12215 tartalmazza a kanonikus GT/AG, 65 a GC/AG, 7 az AT/AC splice junctionöket. Kizárólag azokat a LoRTIA által annotált transzkripteket fogadtuk el, melyek legalább két különböző technikával és három mintában jelen voltak. A transzkriptek átlag hosszúsága 693,661 nt és nem azonosítottunk szignifikáns eltérést a fertőzés során. Az 5'-UTR-ek átlagos hossza 52,956 nt (sd: 75.903), míg a 3'-UTR-eké 295,219. A minimum 10 readre vonatkozó kritériumot is figyelembe véve 758 új TI-t (207-et már korábban azonosított TSS és TES pozíciókkal ± 10 nt-os intervallumban), 692 hossz izoformát és 66 alternatívan spliceolt variánst azonosítottunk. Összesen 239 mRNS TSS pozíciója tér el a korábban annotálttól, 19 gén TES pozíciója, míg 56 gén mindkét vége. Harmincegy TI esetén a TSS downstream helyezkedik el a korábban azonosított transláció iniciáció helyétől, s így rövidebb ORF-fel rendelkeznek. Ezek az RNS-ek N-terminálisukban csonkolt fehérjét eredményezhetnek. Mindösszesen 177 ncRNS-t annotáltunk.

A gazdagének válasza a VACV fertőzés hatására: korábbi tanulmányok kevés differenciálisan expresszálódó gént azonosítottak a VACV fertőzés hatására^{448,449}, míg egy nemrégiben végzett proteomikai tanulmány szintén azt mutatta, hogy a VACV fertőzés nagyon kevés gazdagént befolyásol⁴⁵⁰. Az eredményeink szerint nagymértékben expresszálódó 768 gént öt klaszterbe csoportosítottuk a vírusfertőzésre mutatott válaszuk alapján (**59. 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 expresszálódtak, 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 (**59. ábra**). A legjobban jellemezett génklaszterek 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 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.



59. ábra **A nagyon abundáns gazdagének expressziós változásai a VACV fertőzés során**²⁹⁶. (A) A gazdagén klaszterek expressziós profilja. (B) Azonosított öt elkülönülő gazdagén klaszter hő térképes ábrája.

7 AZ ADATOK ELÉRHETŐSÉGE

7.1 AZ ÁLTALUNK ELŐÁLLÍTOTT SZEKVENCIÁK ELÉRHETŐSÉGE

7.1.1 Transzkriptom

A transzkriptom szekvenálás nyers adatainak jelentős részét az European Nucleotide Archive-ba (ENA) töltöttük fel, ahol elérhetőek az alábbi linkeken (azonosítók az egyes linkek utolsó / utáni kódjai):

- AcMNPV <https://www.ebi.ac.uk/ena/browser/view/PRJEB24943>
- AcMNPV <https://www.ebi.ac.uk/ena/browser/view/PRJEB27818>
- ASFV <https://www.ebi.ac.uk/ena/browser/view/PRJEB36723>
- BoHV-1 <https://www.ebi.ac.uk/ena/browser/view/PRJEB33511>
- CCHFV, WNV és ZIKA <https://www.ebi.ac.uk/ena/browser/view/PRJEB46591>
- CCHFV, IVA, VSIV WNV, ZIKA <https://www.ebi.ac.uk/ena/browser/view/PRJEB43615>
- CVB5 <https://www.ebi.ac.uk/ena/browser/view/PRJEB46598>
- EBV <https://www.ebi.ac.uk/ena/browser/view/PRJEB38992>
- EHV-1 <https://www.ebi.ac.uk/ena/browser/view/PRJEB52190>
- HCMV / RSII <https://www.ebi.ac.uk/ena/browser/view/PRJEB22072>
- HCMV / Sequel és MinION <https://www.ebi.ac.uk/ena/browser/view/PRJEB25680>
- HSV-1 <https://www.ebi.ac.uk/ena/browser/view/PRJEB25433>
- IVA <https://www.ebi.ac.uk/ena/browser/view/PRJEB46600>
- HSV-1 <https://www.ebi.ac.uk/ena/browser/view/PRJEB25433>
- KSHV <https://www.ebi.ac.uk/ena/browser/view/PRJEB60022>
- PRV / RSII <https://www.ebi.ac.uk/ena/browser/view/PRJEB17709>
- PRV / Illumina <https://www.ebi.ac.uk/ena/browser/view/PRJEB9526>
- PRV / RSII <https://www.ebi.ac.uk/ena/browser/view/PRJEB12867>
- PRV – RSII, Sequel, MinION <https://www.ebi.ac.uk/ena/browser/view/PRJEB24593>
- SARS-CoV-2 <https://www.ebi.ac.uk/ena/browser/view/PRJEB46891>
- VACV / Sequel és MinION <https://www.ebi.ac.uk/ena/browser/view/PRJEB26430>
- VACV / RSII és MinION <https://www.ebi.ac.uk/ena/browser/view/PRJEB26434>
- VACV / RSII és MinION <https://www.ebi.ac.uk/ena/browser/view/PRJEB23324>
- Vero cells <https://www.ebi.ac.uk/ena/browser/view/PRJEB46639>
- VSIV <https://www.ebi.ac.uk/ena/browser/view/PRJEB46127>

- VZV <https://www.ebi.ac.uk/ena/browser/view/PRJEB25401>

A HSV-1 egy kisebb hányada a Gene Expression Omnibus (GEO) adatbázisba került feltöltésre:

- HSV-1 <https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE97785>)

7.1.2 Genom

A genomszekvenálások adatai az alábbi linkeken érhetőek el

- PRV Ka <https://www.ncbi.nlm.nih.gov/nucore/KJ717942.1>
- PRV MdBio <https://www.ncbi.nlm.nih.gov/nucore/LT934125.1>
- PCV-1 <https://www.ncbi.nlm.nih.gov/nucore/KX816645>
- PERV <https://www.ncbi.nlm.nih.gov/nucore/KY484771>
- SARS-CoV-2 <https://www.ncbi.nlm.nih.gov/nucore/OM812693.1>
- VACV <https://www.ncbi.nlm.nih.gov/nucore/LT966077.1>

7.2 AZ ANALÍZISÜNKHÖZ FELHASZNÁLT, MÁSOK ÁLTAL KÖZÖLT SZEKVENCIAÍK ELÉRHETŐSÉGE

A GEO adatbázisból az alábbi adatokat használtuk

GSE79337 (<https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE79337>)

GSE59717 (<https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE59717>)

GSE128324 (<https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE128324>)

A Sequence Read Archive-ból a következőket:

PRJNA505045 (<https://www.ncbi.nlm.nih.gov/bioproject/PRJNA505045/>)

PRJNA482043 (<https://www.ncbi.nlm.nih.gov/bioproject/?term=PRJNA482043>)

PRJNA483305 (<https://www.ncbi.nlm.nih.gov/bioproject/?term=PRJNA483305>)

PRJNA533478 (<https://www.ncbi.nlm.nih.gov/bioproject/?term=PRJNA533478>)

Az ENA-ból pedig az alábbiakat:

PRJEB27861 (<https://www.ebi.ac.uk/ena/browser/view/PRJEB27861>)

PRJEB42868 (<https://www.ebi.ac.uk/ena/browser/view/PRJEB42868>)

PRJEB38829 (<https://www.ebi.ac.uk/ena/browser/view/PRJEB38829>)

PRJEB64684 (<https://www.ebi.ac.uk/ena/browser/view/PRJEB64684>)

8 AKTUÁLIS PROJEKTJEINK, JÖVŐBENI TERVEINK

Az OBI 3G csoportja Boldogkői Zsolt vezetésével folytatja a vírus transzkriptomok feltérképezését, az alábbiak szerint. Új projektjeikben, amely részben az értekezésünk alapját képező eredményeken alapul, fő célként egy új genetikai szabályozó rendszer felfedezését tűztük ki vírusmodellek alkalmazásával, a genetikai szabályozásra vonatkozó hipotézisünk alapján. Lényege, hogy a gének a leíródásuk során különféle transzkripció átfedéseket hoznak létre egymással. Feltételezésünk szerint az átfedő DNS szakaszok átírása során a transzkripció apparátusok kölcsönhatnak egymással és egy össz-genom szintű interferencia hálózatot hoznak létre. Az elmúlt években ennek a rendszernek feltártuk a strukturális hátterét, a jelen és jövő projektjei pedig a funkcionalitás analízisére fókuszálnak.

- Beavatkozunk a vírusok genomjának, vagy egyes genetikai moduljainak a működésébe CRISPR/Cas9 és homológ rekombinációs módszerekkel, majd azt vizsgáljuk, hogy milyen rendszerszintű változások következnek be a génexpressziót illetően.
- Az általunk leírt replikációs origó-hoz kapcsolt raRNS-ek szerepét vizsgáljuk. Az előzetes kutatásaink feltárták, hogy ezek az RNS-ek minden vizsgált DNS vírussal előfordulnak, de még a rokon fajok is különböző megoldásokat alkalmazhatnak a kódolásukat illetően. Az raRNS-ek funkcióját vizsgáljuk meg azáltal, hogy módosítjuk a kifejeződését, meghatározzuk a sejten belüli lokalizációját és hogy kapcsolódik-e nukleinsav molekulákhoz. Feltételezéseink szerint, e transzkriptek nem csupán RNS-ként funkcionálnak, hanem a leíródásuk szabályozza a DNS replikáció elkezdését és haladásának irányát is. A projekt során számos vírus transzkriptomjának szerkezetét, illetve dinamikájának törvényszerűségeit ismerjük meg, de a levont következtetések univerzálisak lehetnek. A kutatás során fejlesztésre kerülő kísérletes és bioinformatikai eszközöket pedig a kutató közösség rendelkezésére állnak majd.
- új megközelítéseket vezetünk be a transzkriptom funkciójának szekvenálás-alapú analíziséhez
 - o a SLAM-Seq technológiát⁴⁵¹ alkalmazzuk az újonnan szintetizált és a már kész, létező RNS-ek megkülönböztetésére;
 - o Okazaki fragmens szekvenálással⁴⁵² vizsgáljuk a replikációs villa irányultságát az raRNS-ek funkcionális vizsgálatához
- a génmódosított vírusokat transzneurális nyomjelzésre is alkalmazzuk

A vezetésemmel működő MTA-SZTE Lendület Genomika, Mikrobiomika, Nutri-omikák és Informatika (GeMiNI) kutatócsoport pedig a szekvenálás, főként az LRS, valamint a kapcsolódó laboratóriumi és bioinformatikai területeken szerzett tapasztalatokat felhasználva egy új területet térképez fel. Projektjeink az elhízás, 2-es típusú diabétesz, diéta, egészség, öregedés kérdéskör köré összpontosulnak, 'state-of-the-art' genomikai megközelítésekkel. Projektjeink központjában továbbra is kulcsszerepet kapnak a vírusok, hiszen együttesük, a virom egészségre gyakorolt hatása még kevésbé ismert. Kimutatták azonban, hogy többek közt a herpeszvírusok a humán bélrendszeri virom 'core' képviselői, összefüggést találtak a virom EBV tartalma és az elhízás között^{453,454}, illetve a HSV és HCMV szeropozitivitás, valamint a (pre)diabétesz kialakulása között⁴⁵⁵, így a herpeszvíruskutatás fontos szereppel bír majd e projektjeinkben is.

9 LEGFONTOSABB EREDMÉNYEINK ÖSSZEGZÉSE

A dolgozat az elmúlt tíz év, elsősorban hosszú-read szekvenáláson alapuló, vírus genomikával és transzkriptomikával kapcsolatos eredményeink egy szűk keresztmetszetét mutatja be. Egy külön fejezet pedig a korábbi, a qRT-PCR-technikát alkalmazó, a vírusgének funkcionális elemzésére fókuszáló projektjeinkről szól.

A projekteink fő célkitűzése a virális komplexitás feltérképezése volt, melyet előbb NGS, majd TGS módszerekkel vizsgáltunk. Továbbá, arra is kíváncsiak voltunk, hogy a vírusok fő szabályozó génjei milyen hatással vannak a virális gének expressziójára. Az új LRS szekvenálási technológiák lehetővé tették, hogy a vírusok gazdasejtekre kifejtett hatását is a korábbi lehetőséghez viszonyítva precízebben megvizsgáljuk.

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, ami elsősorban azt jelenti, hogy a genomok transzkripcionális aktivitása jelentősebb, mint azt az ORF-ek száma indokolná.** Az intergenikus régiókat hosszú UTR szekvenciák írják át, illetve önálló transzkriptek is leíródnak e DNS szakaszokról.

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 terminálódik a transzkripció.** Ebben a régióban **az antiszensz RNS-ek expressziója 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 transzkripció

átfedések alapvető jellemzői a virális RNS profiloknak. **A transzkripció 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 transzkripció 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 leíró különböző **transzkript izoformák egymáshoz viszonyított aránya hogyan változik a fertőzés egyes időpontjaiban**^{21,30}, továbbá megvizsgáltuk a transzkript izoformák kódoló kapacitását is. Megállapítottuk, hogy a 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.

Az LRS és a korábbi² qRT-PCR eredményeink összehasonlítása: 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 párhuzamosan 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 alkalmaztunk. A dinamikus transzkriptom analízishez **saját matematikai megközelítéseket** dolgoztunk ki.

Megszekvenáltuk a PRV-Ka törzset, saját izolátumunkat, s az MdBio törzset, valamint géndeléciós PRV genomot is. **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 laboratóriumunk 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 kimutattuk, hogy **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 génkiütött PRV-ket állítottunk elő, majd a vad típusú vírussal párhuzamosan az összes génjük expresszióját mértük real-time kvantitatív 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 csoportokat, 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 sabaues (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 sejtfelszíni 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.

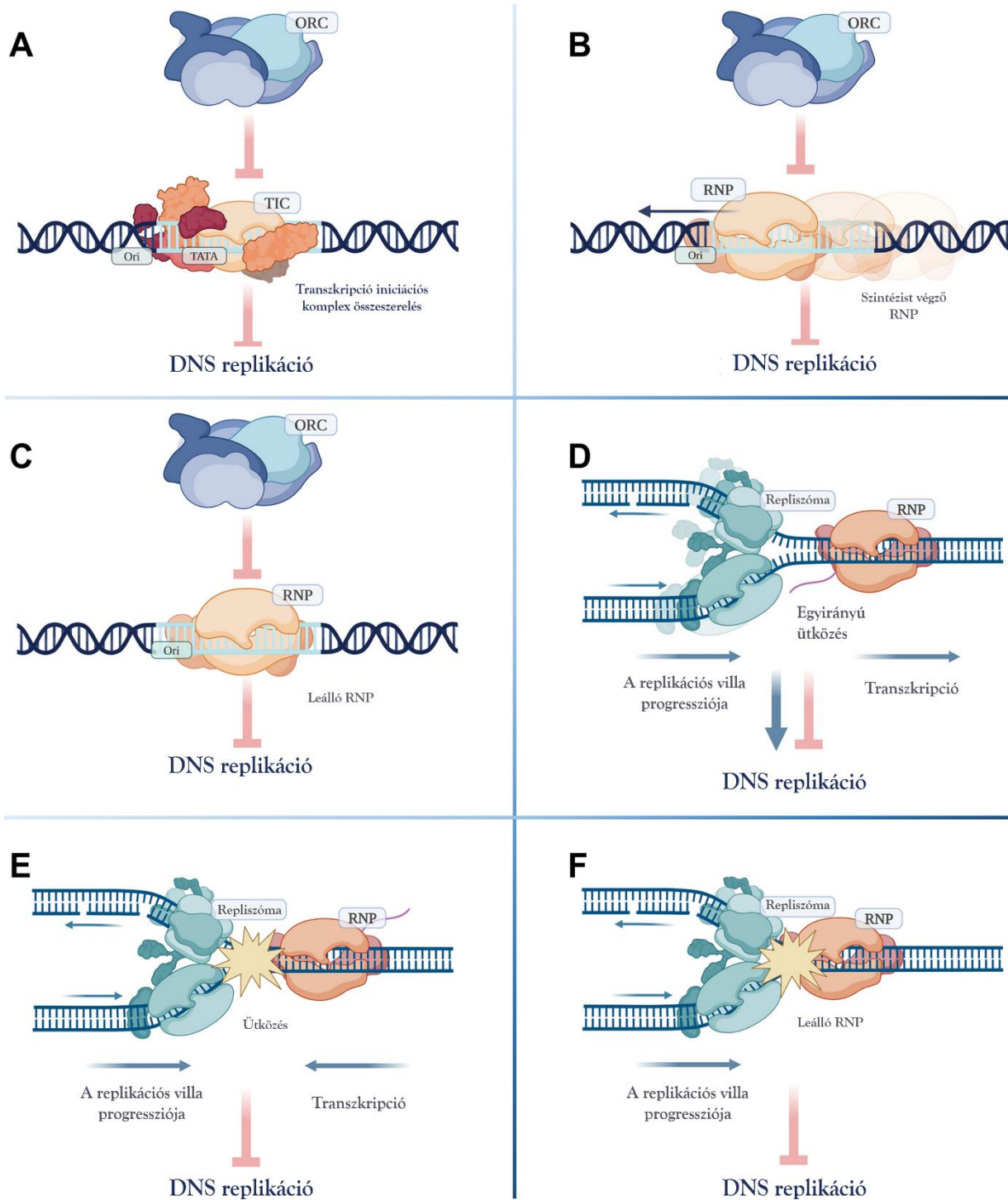
A VSIV fertőzés hatását két sejtvonalon (Vero és T98G) is megfigyeltük. Jelentős különbséget találtunk a két sejttypusban a vírusfertőzés hatékonyságában. **A gliasejtek sokkal érzékenyebbek a fertőzésre,** míg a fibroblaszt sejtek ellenállóbbak. Közel 1400 eltérően kifejeződő gént és 35 különbözően expresszálódó KEGG (Kyoto Encyclopedia of Genes and Genomes - Kyoto Gén- és Genomtudományi Enciklopédia) útvonalat észleltünk a sejtvonalak 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 veszettség fertőzés során az egér agysejtekben⁴⁵⁶. Ez arra utal, hogy ezek a gének valószínűleg más sejtvonalakban és gazdafajokban is érintettek, és ezért további vizsgálat tárgyai kell, hogy legyenek a kérődzők vezikuláris betegségeinek patogenezisével kapcsolatban.

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

A projektjeink során – illetve egyes vírusoknál mások által - kimutatott jelenségek, azaz a rendkívül összetett, genom-szerte átfedő transzkripció hálózat létezése^{262,297,299,344}, 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**²⁷⁵. Kimutattuk, hogy a legtöbb, általunk vizsgált vírusban **a replikációs origókat átfedik bizonyos IncRNS-ek, valamint az mRNS-ek hosszú 5' vagy 3' UTR izoformái.** Úgy véljük, hogy számos 5' UTR izoforma nem kódoló, mivel jelentős távolság van azok TSS-ei és a start kodonok között. Kivételek lehetnek azok a transzkriptek, amelyeknek az 5' UTR nagy részei splicing által ki vannak vágva²⁶². Funkcionális elemzések felfedték a mechanisztikus részleteket arról, hogy a replikációs RNS-ek hogyan irányítják a DNS-szintézis kezdetét az RNS:DNS hibridek kialakulása révén több vírusban is⁸⁶.

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 α HV-ok vizsgálatára, mivel esetükben ezekről az raRNS-ekről kevesebb információ állt rendelkezésre a többi herpeszvírushoz képest. Munkánk **során leírtunk új raRNS-eket és korrigáltuk már leírt raRNS-ek korábbi annotációit.** Továbbá **azonosítottunk egy bonyolult TO-hálózatot a replikációt és transzkripciót szabályozó gének által kódolt transzkriptek és körülöttük lévő specifikus IncRNS-ek között is.** Promóter konszenzus szekvenciákat is detektáltunk az összes vizsgált herpeszvírus OriS-ében. Bár a herpeszvírus OriS-ei 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 IncRNS 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 mentén⁴⁵⁷. Korábbi eredmények azt sugallják, hogy ezek az ütközések drámaibb következményekkel járnak, amikor konvergens orientációban történnek, nem pedig parallel⁴⁵⁸. Számos molekuláris mechanizmus alakult ki az RNP és a replikációs villa közötti konfliktus minimalizálására⁴⁵⁹. 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 replizóma 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**).



60. ábra **A transzkripció a DNS replikációra gyakorolt potenciális hatásai.**⁸² (A) Az Ori-ban található promóterekről induló transzkripció iniciációs komplexek toborzása gátolja az ORC és a repliszóma összeállítását. (B) Az RNS polimeráz II áthaladása az Ori-n gátolja az ORC és a repliszóma összeállítását. (C) Az RNS polimeráz II megakadása az Ori-n gátolja az ORC és a repliszóma összeállítását. (D) A replikáció és a transzkripció apparátusainak egyidejű mozgása lassíthatja vagy gyorsíthatja a replikációs villa előrehaladását. A transzkripció elősegítheti a DNS replikációját a két DNS-szál előzetes megnyitásával. (E) A replikáció és a transzkripció apparátusainak egymással való ütközése mindkét folyamatot gátolja. (F) Az RNS polimeráz II megakadása gátolja a replikációs villa előrehaladását.

Az α 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 IncRNS-ekkel is átfedésben vannak.** A TO-k valószínűsíthető funkciója, hogy további interakció formákat segítsenek 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 α 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 α HV-ok legösszetettebb genomi lokusza, amely IncRNS-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 IncRNS (LAT, LLT, L/ST) és miRNS is kifejeződik ebből a genomi régióból⁴⁵⁹.

Azonosítottuk a NOIR-1 transzkript családot, amelyet az US-IR régió kódol, és amely a Varicellovírus nemzetség evolúciós újítása. Ezeknek az IncRNS-eknek a pontos helye jelentős változatosságot mutat, azonban közös bennük, hogy 3'-koterminálisban végződnek a latenciában működő LLT transzkripttel, ugyanakkor a lítikus ciklusban exprezálódnak. 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 RNS-eket először PRV-ben azonosítottuk^{6,272}, majd detektáltuk az ortológjaikat az EHV-1-ben^{25,82}.** A kanonikus **CTO-S** nem fedi át az OriL-t, de **befolyásolhatja a replikációt** azáltal, hogy segíti szétválasztani a két DNS szálát, s így például 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 α HV-okban, összesen három kemény konvergencia 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 izofomái, vagy ezek kombináció alkotják őket. Ugyanakkor **az általános előfordulásuk a jelenség jelentőségét mutatja, a DNS replikáció és az általános transzkripció szabályozásában**. Míg az α 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.

Úgy véljük, hogy eredményeink következményei sokkal szélesebb perspektívát képviselnek azt illetően, hogy hogyan fejlődött együtt a herpeszvírusok replikációjának és transzkripciójának szabályozása. Ezek a bonyolult kölcsönhatások a különböző gének és szabályozó elemek között értékes betekintést nyújthatnak az általános szabályozási mechanizmusokba, amelyek a herpeszvírusok replikációját és génkifejeződésüket irányítják. **E potenciális kölcsönhatásokra és funkcionális jelentőségükre fókuszáló további vizsgálatok új utat nyithatnak a herpeszvírus-fertőzések kezelésére irányuló innovatív terápiás megközelítések felé.**

10 ZÁRSZÓ

A vírusok nagyon egyszerűnek tűnhetnek, hiszen rendkívül kicsik, sokak szerint még csak nem is élőlények. A dolgozatban bemutatott eredményeink a molekuláris mechanizmusaik egy szegmensére, s annak általunk feltárt bonyolultságára világítanak rá. A vírusokkal kapcsolatos kutatások jelentőségét a Covid-19 járványnál jobban semmi sem támasztotta alá az utóbbi évtizedekben, azonban a magyar származású virológus-immunológus, George Klein (Klein György) fogalmazta meg talán a legfrappánsabban:

“The stupidest virus is cleverer than the cleverest virologist⁴⁶⁰.”

– maradt még megismerni való.



11 IRODALOMJEGYZÉK

1. 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 Feb;6(2):127-30. doi: 10.1038/nmeth.1292.
2. 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 Oct 23;10:491. doi: 10.1186/1471-2164-10-491.
3. 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. doi: 10.1186/s12866-015-0470-0.
4. 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. doi: 10.1128/genomeA.00628-14.
5. 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. doi: 10.1155/2019/1806842.
6. 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. doi: 10.1371/journal.pone.0162868.
7. 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. doi: 10.1038/srep43751.
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. *Front Microbiol*. 2017 Jun 20;8:1079. doi: 10.3389/fmicb.2017.01079.
9. 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. doi: 10.1038/s41598-017-16262-z.
10. 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. *Sci Rep*. 2017 Aug 2;7(1):7106. doi: 10.1038/s41598-017-06522-3.
11. 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. *Sci Data*. 2019 Feb 5;6:190010. doi: 10.1038/sdata.2019.10.
12. 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. *Nat Biotechnol*. 2022 Jul;40(7):1082-1092. doi: 10.1038/s41587-022-01231-3.

13. 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. doi: 10.3389/fmicb.2017.02708.
14. 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. doi: 10.1186/s12864-018-5267-8. PMID: 30514211; PMCID: PMC6280550.
15. Balázs, Z. (2018). LoRTIA. Available at: <https://github.com/zsolt-balazs/LoRTIA> [Accessed January 16, 2019].
16. 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. doi: 10.1038/sdata.2018.266.
17. 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. doi: 10.1038/s41598-020-70794-5.
18. 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. *Sci Rep.* 2020 Nov 24;10(1):20496. doi: 10.1038/s41598-020-77520-1.
19. 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 Feb 20;10(2):242. doi: 10.3390/pathogens10020242.
20. 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.
21. 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. *Sci Rep.* 2021 Jul 14;11(1):14487. doi: 10.1038/s41598-021-93593-y.
22. 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. doi: 10.1186/s12985-021-01734-6.
23. Torma G, Tombácz D, Moldován N, Fülöp Á, Prazsák I, Csabai Z, Snyder M, Boldogkői Z. Dual isoform sequencing reveals complex transcriptomic and epitranscriptomic landscapes of a prototype baculovirus. *Sci Rep.* 2022 Jan 25;12(1):1291. doi: 10.1038/s41598-022-05457-8.
24. 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. doi: 10.1093/gigascience/giac094.
25. 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 Jun 28;9(7):e17716. doi: 10.1016/j.heliyon.2023.e17716.
26. Prazsák I, Tombácz D, Fülöp Á, Torma G, Gulyás G, Dörmő Á, Kakuk B, McKenzie Spire 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. doi: 10.1128/msystems.01007-23.
27. Roossinck MJ. The good viruses: viral mutualistic symbioses. *Nat Rev Microbiol*. 2011 Feb;9(2):99-108. doi: 10.1038/nrmicro2491.
 28. Hershey AD, Chase M. Independent functions of viral protein and nucleic acid in growth of bacteriophage. *J Gen Physiol*. 1952 May;36(1):39-56. doi: 10.1085/jgp.36.1.39.
 29. Palù G, Parolin C, Takeuchi Y, Pizzato M. Progress with retroviral gene vectors. *Rev Med Virol*. 2000 May-Jun;10(3):185-202. doi: 10.1002/(sici)1099-1654(200005/06)10:3<185::aid-rmv285>3.0.co;2-8.
 30. Marshall E. Gene therapy. Viral vectors still pack surprises. *Science*. 2001 Nov 23;294(5547):1640. doi: 10.1126/science.294.5547.1640.
 31. 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. *J Biomed Biotechnol*. 2009;2009:361795. doi: 10.1155/2009/361795.
 32. Smith AE. Viral vectors in gene therapy. *Annu Rev Microbiol*. 1995;49:807-38. doi: 10.1146/annurev.mi.49.100195.004111.
 33. Boldogkői Z, Nógrádi A. Gene and cancer therapy--pseudorabies virus: a novel research and therapeutic tool? *Curr Gene Ther*. 2003 Apr;3(2):155-82. doi: 10.2174/1566523034578393.
 34. Zhang C, Zhou D. Adenoviral vector-based strategies against infectious disease and cancer. *Hum Vaccin Immunother*. 2016 Aug 2;12(8):2064-2074. doi: 10.1080/21645515.2016.1165908.
 35. Lundstrom K. Viral Vectors in Gene Therapy. *Diseases*. 2018 May 21;6(2):42. doi: 10.3390/diseases6020042.
 36. Li J, Liu T, Dong Y, Kondoh K, Lu Z. Trans-synaptic Neural Circuit-Tracing with Neurotropic Viruses. *Neurosci Bull*. 2019 Oct;35(5):909-920. doi: 10.1007/s12264-019-00374-9.
 37. Xu X, Holmes TC, Luo MH, Beier KT, Horwitz GD, Zhao F, Zeng W, Hui M, Semler BL, Sandri-Goldin RM. Viral Vectors for Neural Circuit Mapping and Recent Advances in Trans-synaptic Anterograde Tracers. *Neuron*. 2020 Sep 23;107(6):1029-1047. doi: 10.1016/j.neuron.2020.07.010.
 38. Travieso T, Li J, Mahesh S, Mello JDFRE, Blasi M. The use of viral vectors in vaccine development. *NPJ Vaccines*. 2022 Jul 4;7(1):75. doi: 10.1038/s41541-022-00503-y.
 39. 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. doi: 10.1038/s41392-023-01408-5.
 40. 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. doi: 10.1186/s13027-018-0218-1.
 41. Melcher A, Harrington K, Vile R. Oncolytic virotherapy as immunotherapy. *Science*. 2021 Dec 10;374(6573):1325-1326. doi: 10.1126/science.abk3436.
 42. Jahan N, Ghouse SM, Martuza RL, Rabkin SD. In Situ Cancer Vaccination and Immunovirotherapy Using Oncolytic HSV. *Viruses*. 2021 Aug 31;13(9):1740. doi: 10.3390/v13091740.
 43. Russell SJ, Bell JC, England CE, McFadden G. Advances in oncolytic virotherapy. *Commun Med (Lond)*. 2022 Apr 7;2:33. doi: 10.1038/s43856-022-00098-4.
 44. Thiel K. Old dogma, new tricks--21st Century phage therapy. *Nat Biotechnol*. 2004 Jan;22(1):31-6. doi: 10.1038/nbt0104-31.

45. Paul VD, Sundarrajan S, Rajagopalan SS, Hariharan S, Kempashanaiah N, Padmanabhan S, Sriram B, Ramachandran J. Lysis-deficient phages as novel therapeutic agents for controlling bacterial infection. *BMC Microbiol.* 2011 Aug 31;11:195. doi: 10.1186/1471-2180-11-195.
46. Federici S, Kredo-Russo S, Valdés-Mas R, Kviatcovsky D, Weinstock E, Matiuhin Y, Silberberg Y, Atarashi K, Furuichi M, Oka A, Liu B, Fibelman M, Weiner IN, Khabra E, Cullin N, Ben-Yishai N, Inbar D, Ben-David H, Nicenboim J, Kowalsman N, Lieb W, Kario E, Cohen T, Geffen YF, Zelcbuch L, Cohen A, Rappo U, Gahali-Sass I, Golemo M, Lev V, Dori-Bachash M, Shapiro H, Moresi C, Cuevas-Sierra A, Mohapatra G, Kern L, Zheng D, Nobs SP, Suez J, Stettner N, Harmelin A, Zak N, Puttagunta S, Bassan M, Honda K, Sokol H, Bang C, Franke A, Schramm C, Maharshak N, Sartor RB, Sorek R, Elinav E. Targeted suppression of human IBD-associated gut microbiota commensals by phage consortia for treatment of intestinal inflammation. *Cell.* 2022 Aug 4;185(16):2879-2898.e24. doi: 10.1016/j.cell.2022.07.003.
47. Fujiki J, Schnabl B. Phage therapy: Targeting intestinal bacterial microbiota for the treatment of liver diseases. *JHEP Rep.* 2023 Sep 23;5(12):100909. doi: 10.1016/j.jhepr.2023.100909.
48. Chen X, Mendes BG, Alves BS, Duan Y. Phage therapy in gut microbiome. *Prog Mol Biol Transl Sci.* 2023;201:93-118. doi: 10.1016/bs.pmbts.2023.04.005.
49. Villarreal LP. Evolution of Viruses. *Encyclopedia of Virology.* 2008:174–84. doi: 10.1016/B978-012374410-4.00706-8.
50. Lefkowitz EJ, Dempsey DM, Hendrickson RC, Orton RJ, Siddell SG, Smith DB. Virus taxonomy: the database of the International Committee on Taxonomy of Viruses (ICTV). *Nucleic Acids Res.* 2018 Jan 4;46(D1):D708-D717. doi: 10.1093/nar/gkx932.
51. Baltimore D. Expression of animal virus genomes. *Bacteriol Rev.* 1971 Sep;35(3):235-41. doi: 10.1128/br.35.3.235-241.1971.
52. Baltimore D. Viral genetic systems. *Trans N Y Acad Sci.* 1971 Mar;33(3):327-32. doi: 10.1111/j.2164-0947.1971.tb02600.x.
53. Baltimore D. The strategy of RNA viruses. *Harvey Lect.* 1974-1975;70 Series:57-74.
54. Walker PJ, Siddell SG, Lefkowitz EJ, Mushegian AR, Adriaenssens EM, Alfenas-Zerbini P, Dempsey DM, Dutilh BE, García ML, Curtis Hendrickson R, Junglen S, Krupovic M, Kuhn JH, Lambert AJ, Łobocka M, Oksanen HM, Orton RJ, Robertson DL, Rubino L, Sabanadzovic S, Simmonds P, Smith DB, Suzuki N, Van Doorslaer K, Vandamme AM, Varsani A, Zerbini FM. Recent changes to virus taxonomy ratified by the International Committee on Taxonomy of Viruses (2022). *Arch Virol.* 2022 Nov;167(11):2429-2440. doi: 10.1007/s00705-022-05516-5.
55. 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. doi: 10.13140/RG.2.2.34193.94562
56. Gaïa M, Meng L, Pelletier E, Forterre P, Vanni C, Fernandez-Guerra A, Jaillon O, Wincker P, Ogata H, Krupovic M, Delmont TO. Mirusviruses link herpesviruses to giant viruses. *Nature.* 2023 Apr;616(7958):783-789. doi: 10.1038/s41586-023-05962-4.
57. Koonin EV, Dolja VV, Krupovic M, Varsani A, Wolf YI, Yutin N, Zerbini M, Kuhn JH. Create a megataxonomic framework, filling all principal taxonomic ranks, for ssDNA viruses. Report number: ICTV Proposal No. 2019.005G. doi: 10.13140/RG.2.2.36696.85760
58. Krupovic M, Kuhn JH, Wang F, Baquero DP, Dolja VV, Egelman EH, Prangishvili D, Koonin EV. Adnaviria: a New Realm for Archaeal Filamentous Viruses with Linear A-Form Double-Stranded DNA Genomes. *J Virol.* 2021 Jul 12;95(15):e0067321. doi: 10.1128/JVI.00673-21.

59. Hepojoki J, Hetzel U, Paraskevopoulou S, Drosten C, Harrach B, Zerbini M, Koonin EV, Krupovic M, Dolja V and Kuhn JH. Create one new realm (Ribozyviria) including one new family (Kolmioviridae) including genus Deltavirus and seven new genera for a total of 15 species. Report number: ICTV Proposal No. 2020.012D. doi: 10.13140/RG.2.2.31235.43041.
60. Zheludev IN, Edgar RC, Lopez-Galiano MJ, de la Peña M, Babaian A, Bhatt AS, Fire AZ. Viroid-like colonists of human microbiomes. *bioRxiv* 2024.01.20.576352; doi: 10.1101/2024.01.20.576352-
61. Howley PM, Knipe DM, Enquist LW: *Fields Virology: Fundamentals*. 7th ed. 2023. Wolters Kluwer.
62. Weir JP. Regulation of herpes simplex virus gene expression. *Gene*. 2001 Jun 27;271(2):117-30. doi: 10.1016/s0378-1119(01)00512-1.
63. DeLuca NA, McCarthy AM, Schaffer PA. Isolation and characterization of deletion mutants of herpes simplex virus type 1 in the gene encoding immediate-early regulatory protein ICP4. *J Virol*. 1985 Nov;56(2):558-70. doi: 10.1128/JVI.56.2.558-570.1985.
64. Fox HL, Dembowski JA, DeLuca NA. A Herpesviral Immediate Early Protein Promotes Transcription Elongation of Viral Transcripts. *mBio*. 2017 Jun 13;8(3):e00745-17. doi: 10.1128/mBio.00745-17.
65. Hagglund R, Roizman B. Role of ICP0 in the strategy of conquest of the host cell by herpes simplex virus 1. *J Virol*. 2004 Mar;78(5):2169-78. doi: 10.1128/jvi.78.5.2169-2178.2004.
66. Zhou C, Knipe DM. Association of herpes simplex virus type 1 ICP8 and ICP27 proteins with cellular RNA polymerase II holoenzyme. *J Virol*. 2002 Jun;76(12):5893-904. doi: 10.1128/jvi.76.12.5893-5904.2002.
67. Mark F. Stinski and Jeffery L. Meier. *Human Herpesviruses: Biology, Therapy, and Immunoprophylaxis*. (Cambridge University Press, 2007).
68. Liu P, Speck SH. Synergistic autoactivation of the Epstein-Barr virus immediate-early BRLF1 promoter by Rta and Zta. *Virology*. 2003 Jun 5;310(2):199-206. doi: 10.1016/s0042-6822(03)00145-4.
69. Perng GC, Dunkel EC, Geary PA, Slanina SM, Ghiasi H, Kaiwar R, Nesburn AB, Wechsler SL. The latency-associated transcript gene of herpes simplex virus type 1 (HSV-1) is required for efficient in vivo spontaneous reactivation of HSV-1 from latency. *J Virol*. 1994 Dec;68(12):8045-55. doi: 10.1128/JVI.68.12.8045-8055.1994.
70. Wang QY, Zhou C, Johnson KE, Colgrove RC, Coen DM, Knipe DM. Herpesviral latency-associated transcript gene promotes assembly of heterochromatin on viral lytic-gene promoters in latent infection. *Proc Natl Acad Sci U S A*. 2005 Nov 1;102(44):16055-9. doi: 10.1073/pnas.0505850102.
71. Boldogkői Z, Murvai J, Fodor I. G and C accumulation at silent positions of codons produces additional ORFs. *Trends Genet*. 1995 Apr;11(4):125-6. doi: 10.1016/s0168-9525(00)89019-8. PMID: 7732585.
72. Tormanen K, Allen S, Mott KR, Ghiasi H. The Latency-Associated Transcript Inhibits Apoptosis via Downregulation of Components of the Type I Interferon Pathway during Latent Herpes Simplex Virus 1 Ocular Infection. *J Virol*. 2019 May 1;93(10):e00103-19. doi: 10.1128/JVI.00103-19.
73. Lee LY, Schaffer PA. A virus with a mutation in the ICP4-binding site in the L/ST promoter of herpes simplex virus type 1, but not a virus with a mutation in open reading frame P, exhibits

- cell-type-specific expression of gamma(1)34.5 transcripts and latency-associated transcripts. *J Virol*. 1998 May;72(5):4250-64. doi: 10.1128/JVI.72.5.4250-4264.1998.
74. Vashee S, Cvetič C, Lu W, Simanček P, Kelly TJ, Walter JC. Sequence-independent DNA binding and replication initiation by the human origin recognition complex. *Genes Dev*. 2003 Aug 1;17(15):1894-908. doi: 10.1101/gad.1084203.
75. 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. doi: 10.1073/pnas.85.9.2959.
76. Santocanale C, Diffley JF. ORC- and Cdc6-dependent complexes at active and inactive chromosomal replication origins in *Saccharomyces cerevisiae*. *EMBO J*. 1996 Dec 2;15(23):6671-9.
77. Weller SK, Coen DM. Herpes simplex viruses: mechanisms of DNA replication. *Cold Spring Harb Perspect Biol*. 2012 Sep 1;4(9):a013011. doi: 10.1101/cshperspect.a013011.
78. Packard JE, Dembowski JA. HSV-1 DNA Replication-Coordinated Regulation by Viral and Cellular Factors. *Viruses*. 2021 Oct 7;13(10):2015. doi: 10.3390/v13102015.
79. Hammerschmidt W, Sugden B. Identification and characterization of oriLyt, a lytic origin of DNA replication of Epstein-Barr virus. *Cell*. 1988 Nov 4;55(3):427-33. doi: 10.1016/0092-8674(88)90028-1.
80. Ballestas ME, Chatis PA, Kaye KM. Efficient persistence of extrachromosomal KSHV DNA mediated by latency-associated nuclear antigen. *Science*. 1999 Apr 23;284(5414):641-4. doi: 10.1126/science.284.5414.641.
81. Wang Y, Tang Q, Maul GG, Yuan Y. Kaposi's sarcoma-associated herpesvirus ori-Lyt-dependent DNA replication: dual role of replication and transcription activator. *J Virol*. 2006 Dec;80(24):12171-86. doi: 10.1128/JVI.00990-06.
82. 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. doi: 10.1038/s41598-023-43344-y.
83. Dellino GI, Cittaro D, Piccioni R, Luzi L, Banfi S, Segalla S, Cesaroni M, Mendoza-Maldonado R, Giacca M, Pelicci PG. Genome-wide mapping of human DNA-replication origins: levels of transcription at ORC1 sites regulate origin selection and replication timing. *Genome Res*. 2013 Jan;23(1):1-11. doi: 10.1101/gr.142331.112.
84. Tikhanovich I, Liang B, Seoighe C, Folk WR, Nasheuer HP. Inhibition of human BK polyomavirus replication by small noncoding RNAs. *J Virol*. 2011 Jul;85(14):6930-40. doi: 10.1128/JVI.00547-11.
85. Gatherer D, Seirafian S, Cunningham C, Holton M, Dargan DJ, Baluchova K, Hector RD, Galbraith J, Herzyk P, Wilkinson GW, Davison AJ. High-resolution human cytomegalovirus transcriptome. *Proc Natl Acad Sci U S A*. 2011 Dec 6;108(49):19755-60. doi: 10.1073/pnas.1115861108.
86. Tai-Schmiedel J, Karniely S, Lau B, Ezra A, Eliyahu E, Nachshon A, Kerr K, Suárez N, Schwartz M, Davison AJ, Stern-Ginossar N. Human cytomegalovirus long noncoding RNA4.9 regulates viral DNA replication. *PLoS Pathog*. 2020 Apr 15;16(4):e1008390. doi: 10.1371/journal.ppat.1008390.
87. Rossetto CC, Tarrant-Elorza M, Pari GS. Cis and trans acting factors involved in human cytomegalovirus experimental and natural latent infection of CD14 (+) monocytes and CD34 (+) cells. *PLoS Pathog*. 2013;9(5):e1003366. doi: 10.1371/journal.ppat.1003366.

88. Huang L, Zhu Y, Anders DG. The variable 3' ends of a human cytomegalovirus oriLyt transcript (SRT) overlap an essential, conserved replicator element. *J Virol*. 1996 Aug;70(8):5272-81. doi: 10.1128/JVI.70.8.5272-5281.1996.
89. Prichard MN, Jairath S, Penfold ME, St Jeor S, Bohlman MC, Pari GS. Identification of persistent RNA-DNA hybrid structures within the origin of replication of human cytomegalovirus. *J Virol*. 1998 Sep;72(9):6997-7004. doi: 10.1128/JVI.72.9.6997-7004.1998.
90. Rennekamp AJ, Lieberman PM. Initiation of Epstein-Barr virus lytic replication requires transcription and the formation of a stable RNA-DNA hybrid molecule at OriLyt. *J Virol*. 2011 Mar;85(6):2837-50. doi: 10.1128/JVI.02175-10.
91. Xu Y, Cei SA, Rodriguez Huete A, Colletti KS, Pari GS. Human cytomegalovirus DNA replication requires transcriptional activation via an IE2- and UL84-responsive bidirectional promoter element within oriLyt. *J Virol*. 2004 Nov;78(21):11664-77. doi: 10.1128/JVI.78.21.11664-11677.2004.
92. Norseen J, Thomae A, Sridharan V, Aiyar A, Schepers A, Lieberman PM. RNA-dependent recruitment of the origin recognition complex. *EMBO J*. 2008 Nov 19;27(22):3024-35. doi: 10.1038/emboj.2008.221.
93. Voss JH, Roizman B. Properties of two 5'-coterminal RNAs transcribed part way and across the S component origin of DNA synthesis of the herpes simplex virus 1 genome. *Proc Natl Acad Sci U S A*. 1988 Nov;85(22):8454-8. doi: 10.1073/pnas.85.22.8454.
94. Carter JB, Saunders VA. (2013). *Virology: Principles and Applications*. Hoboken, NJ: John Wiley & Sons.
95. Roizman B, Knipe DM, Whitley RJ. (2013). *Herpes Simplex Viruses*. In D. M. Knipe & P. M. Howley (Eds.), *Fields' Virology* (6th ed., pp. 1823–1897). Wolters Kluwer Health/Lippincott Williams & Wilkins.
96. Davison AJ. (2007). Comparative analysis of the genomes, in *Source Human Herpesviruses: Biology, Therapy, and Immunoprophylaxis*, eds A. Arvin, G. Campadelli-Fiume, E. Mocarski, P. S. Moore, B. Roizman, R. Whitley, et al. (Cambridge: Cambridge University Press).
97. Davison AJ, Clements JB. (2010). Herpesviruses: general properties, in *Topley & Wilson's Microbiology and Microbial Infections*, eds B. W. J. Mahy, V. ter Meulen, S. P. Borriello, P. R. Murray, G. Funke, W. G. Merz, et al. (Chichester: John Wiley & Sons, Ltd.), doi: 10.1002/9780470688618.taw0231
98. Knipe DM, Howley PM. (2013). *Fields' Virology* (6th ed.). Wolters Kluwer Health/Lippincott Williams & Wilkins.
99. Looker KJ, Magaret AS, May MT, Turner KM, Vickerman P, Gottlieb SL, Newman LM. Global and Regional Estimates of Prevalent and Incident Herpes Simplex Virus Type 1 Infections in 2012. *PLoS One*. 2015 Oct 28;10(10):e0140765. doi: 10.1371/journal.pone.0140765.
100. Kennedy PG. Varicella-zoster virus latency in human ganglia. *Rev Med Virol*. 2002 Sep-Oct;12(5):327-34. doi: 10.1002/rmv.362.
101. Weinmann, S., Chun, C., Schmid, D. S., Roberts, M., Vandermeer, M., Riedlinger, K., ... & Irving, S. A. (2019). Incidence of Herpes Zoster Among Children: 2003–2014. *Pediatrics*, 144(1), e20182917.
102. Stevenson EV, Collins-McMillen D, Kim JH, Cieply SJ, Bentz GL, Yurochko AD. HCMV reprogramming of infected monocyte survival and differentiation: a Goldilocks phenomenon. *Viruses*. 2014 Feb 13;6(2):782-807. doi: 10.3390/v6020782.

- 103.** Wattré P, Dewilde A, Lobert PE. Actualités sur la pathologie du cytomégalo­virus humain [Current status of human cytomegalovirus disease]. *Rev Med Interne*. 1995;16(5):354-67. French. doi: 10.1016/0248-8663(96)80723-5.
- 104.** Wen LZ, Xing W, Liu LQ, Ao LM, Chen SH, Zeng WJ. Cytomegalovirus infection in pregnancy. *Int J Gynaecol Obstet*. 2002 Nov;79(2):111-6. doi: 10.1016/s0020-7292(02)00239-4.
- 105.** Gerna G, Lilleri D. Human cytomegalovirus (HCMV) infection/re-infection: development of a protective HCMV vaccine. *New Microbiol*. 2019 Jan;42(1):1-20. Epub 2019 Jan 21. PMID: 30671581.
- 106.** Dewhurst S. Human herpesvirus type 6 and human herpesvirus type 7 infections of the central nervous system. *Herpes*. 2004 Jun;11 Suppl 2:105A-111A.
- 107.** Mullins TB, Krishnamurthy K. Roseola Infantum. [Updated 2023 Jul 4]. In: StatPearls. Treasure Island (FL): StatPearls Publishing; 2024 Jan-. <https://www.ncbi.nlm.nih.gov/books/NBK448190/>
- 108.** Mozhgani SH, Rajabi F, Qurbani M, Erfani Y, Yaslianifard S, Moosavi A, Pourrostami K, Baradaran Bagheri A, Soleimani A, Behzadian F, Safavi M, Rezaei F. Human Herpesvirus 6 Infection and Risk of Chronic Fatigue Syndrome: A Systematic Review and Meta-Analysis. *Intervirology*. 2022;65(1):49-57. doi: 10.1159/000517930.
- 109.** Santpere G, Telford M, Andrés-Benito P, Navarro A, Ferrer I. The Presence of Human Herpesvirus 6 in the Brain in Health and Disease. *Biomolecules*. 2020 Nov 6;10(11):1520. doi: 10.3390/biom10111520.
- 110.** Hogestyn JM, Salois G, Xie L, Apa C, Youngyunpipatkul J, Pröschel C, Mayer-Pröschel M. Expression of the human herpesvirus 6A latency-associated transcript U94A impairs cytoskeletal functions in human neural cells. *Mol Cell Neurosci*. 2022 Dec;123:103770. doi: 10.1016/j.mcn.2022.103770.
- 111.** Cohen JI. Epstein-Barr virus infection. *N Engl J Med*. 2000 Aug 17;343(7):481-92. doi: 10.1056/NEJM200008173430707.
- 112.** Rickinson AK. Epstein-Barr virus. In: Knipe DM, Howley PM, editors. *Fields virology*. 5th ed. Philadelphia: Lippincott, Williams and Wilkins; 2007. p. 2655-700.
- 113.** Young LS, Yap LF, Murray PG. Epstein-Barr virus: more than 50 years old and still providing surprises. *Nat Rev Cancer*. 2016 Dec;16(12):789-802. doi: 10.1038/nrc.2016.92.
- 114.** Shannon-Lowe C, Rickinson A. The Global Landscape of EBV-Associated Tumors. *Front Oncol*. 2019 Aug 6;9:713. doi: 10.3389/fonc.2019.00713.
- 115.** Gold JE, Okyay RA, Licht WE, Hurley DJ. Investigation of Long COVID Prevalence and Its Relationship to Epstein-Barr Virus Reactivation. *Pathogens*. 2021 Jun 17;10(6):763. doi: 10.3390/pathogens10060763.
- 116.** Uldrick TS, Whitby D. Update on KSHV epidemiology, Kaposi Sarcoma pathogenesis, and treatment of Kaposi Sarcoma. *Cancer Lett*. 2011 Jun 28;305(2):150-62. doi: 10.1016/j.canlet.2011.02.006.
- 117.** Goncalves PH, Ziegelbauer J, Uldrick TS, Yarchoan R. Kaposi sarcoma herpesvirus-associated cancers and related diseases. *Curr Opin HIV AIDS*. 2017 Jan;12(1):47-56. doi: 10.1097/COH.0000000000000330.
- 118.** Chang Y, Cesarman E, Pessin MS, Lee F, Culpepper J, Knowles DM, Moore PS. Identification of herpesvirus-like DNA sequences in AIDS-associated Kaposi's sarcoma. *Science*. 1994 Dec 16;266(5192):1865-9. doi: 10.1126/science.7997879.

- 119.** Ganem D. KSHV and the pathogenesis of Kaposi sarcoma: listening to human biology and medicine. *J Clin Invest*. 2010 Apr;120(4):939-49. doi: 10.1172/JCI40567.
- 120.** Cesarman E, Knowles DM. The role of Kaposi's sarcoma-associated herpesvirus (KSHV/HHV-8) in lymphoproliferative diseases. *Semin Cancer Biol*. 1999 Jun;9(3):165-74. doi: 10.1006/scbi.1998.0118.
- 121.** Azab W, Dayaram A, Greenwood AD, Osterrieder N. How Host Specific Are Herpesviruses? Lessons from Herpesviruses Infecting Wild and Endangered Mammals. *Annu Rev Virol*. 2018 Sep 29;5(1):53-68. doi: 10.1146/annurev-virology-092917-043227.
- 122.** Pomeranz LE, Reynolds AE, Hengartner CJ. Molecular biology of pseudorabies virus: impact on neurovirology and veterinary medicine. *Microbiol Mol Biol Rev*. 2005 Sep;69(3):462-500. doi: 10.1128/MMBR.69.3.462-500.2005.
- 123.** Müller T, Hahn EC, Tottewitz F, Kramer M, Klupp BG, Mettenleiter TC, Freuling C. Pseudorabies virus in wild swine: a global perspective. *Arch Virol*. 2011 Oct;156(10):1691-705. doi: 10.1007/s00705-011-1080-2.
- 124.** Pusic I, Prodanov-Radulovic J, Dosen R, Stojanov I, Stojanovic D, Petrovic T. Epizootical characteristics of Aujeszky's disease in Vojvodina region and biosecurity concerns. *Biotech Animal Husbandry*, 2783,9.875–882, 2011.
- 125.** Prodanov-Radulović J, Došen R, Pušić I et al. Emergence of pseudorabies virus (morbus aujeszky) infection at large swine farms in AP Vojvodina (Serbia). *Contem Agriculture*. 64:105–111, 2015.
- 126.** Lazić S, Milićević V, Lazić G et al. A serological survey on Aujeszky's disease in wild boars in the region of Vojvodina in Serbia. in Proceedings of the First International Symposium of Veterinary Medicine "One Health-New Challenges" (ISVM2015), pp. 290–294, Vrdnik, Serbia, May 2015.
- 127.** Köppel C, Knopf L, Ryser MP, Miserez R, Thür B, Stärk KDC. Serosurveillance for selected infectious disease agents in wild boars (*Sus scrofa*) and outdoor pigs in Switzerland. *Eur J Wildlife Res*, 53:3,212–220, 2007.
- 128.** Masot AJ, Gil M, Risco D, Jiménez OM, Núñez JI, Redondo E. Pseudorabies virus infection (Aujeszky's disease) in an Iberian lynx (*Lynx pardinus*) in Spain: a case report. *BMC Vet Res*. 2017 Jan 5;13(1):6. doi: 10.1186/s12917-016-0938-7.
- 129.** Papageorgiou KV, Suárez NM, Wilkie GS, Filioussis G, Papaioannou N, Nauwynck HJ, Davison AJ, Kritas SK. Genome Sequences of Two Pseudorabies Virus Strains Isolated in Greece. *Genome Announc*. 2016 Jan 21;4(1):e01624-15. doi: 10.1128/genomeA.01624-15.
- 130.** Gray A: Aujeszky's disease found in Europe for first time since 2019. *Veterinary Record*. 2023. <https://doi.org/10.1002/vetr.00100076>
- 131.** 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. doi: 10.1155/2019/1806842.
- 132.** Tikoo SK, Campos M, Babiuk LA. Bovine herpesvirus 1 (BHV-1): biology, pathogenesis, and control. *Adv Virus Res*. 1995;45:191-223. doi: 10.1016/s0065-3527(08)60061-5.
- 133.** Kydd JH, Smith KC, Hannant D, Livesay GJ, Mumford JA. Distribution of equid herpesvirus-1 (EHV-1) in the respiratory tract of ponies: implications for vaccination strategies. *Equine Vet J*. 1994 Nov;26(6):466-9. doi: 10.1111/j.2042-3306.1994.tb04051.x.

- 134.** European Food Safety Authority (EFSA); Carvelli A, Nielsen SS, Paillot R, Broglia A, Kohnle L. Clinical impact, diagnosis and control of Equine Herpesvirus-1 infection in Europe. *EFSA J.* 2022 Apr 6;20(4):e07230. doi: 10.2903/j.efsa.2022.7230.
- 135.** Mönttinen HAM, Bicep C, Williams TA, Hirt RP. The genomes of nucleocytoplasmic large DNA viruses: viral evolution writ large. *Microb Genom.* 2021 Sep;7(9):000649. doi: 10.1099/mgen.0.000649.
- 136.** Baxby D. 1996. Poxviruses, chapter 69. In Baron S (ed), *Medical microbiology*, 4th ed. University of Texas Medical Branch at Galveston, Galveston, TX.
- 137.** Hendrickson RC, Wang C, Hatcher EL, Lefkowitz EJ. Orthopoxvirus genome evolution: the role of gene loss. *Viruses.* 2010 Sep;2(9):1933-1967. doi: 10.3390/v2091933.
- 138.** Condit RC, Moussatche N, Traktman P. In a nutshell: structure and assembly of the vaccinia virion. *Adv Virus Res.* 2006;66:31-124. doi: 10.1016/S0065-3527(06)66002-8.
- 139.** Reeves PM, Smith SK, Olson VA, Thorne SH, Bornmann W, Damon IK, Kalman D. Variola and monkeypox viruses utilize conserved mechanisms of virion motility and release that depend on abl and SRC family tyrosine kinases. *J Virol.* 2011 Jan;85(1):21-31. doi: 10.1128/JVI.01814-10.
- 140.** Walsh D. Poxviruses: Slipping and sliding through transcription and translation. *PLoS Pathog.* 2017 Nov 16;13(11):e1006634. doi: 10.1371/journal.ppat.1006634.
- 141.** Diven DG. An overview of poxviruses. *J Am Acad Dermatol.* 2001 Jan;44(1):1-16. doi: 10.1067/mjd.2001.109302.
- 142.** Moss B, Smith GL. Poxviridae: The Viruses and Their Replication. in *Field's Virology* 573–613 (2021).
- 143.** Elwood JM. Smallpox and its eradication. *J Epidemiol Community Health.* 1989 Mar;43(1):92.
- 144.** Moss B. Vaccinia virus: a tool for research and vaccine development. *Science.* 1991 Jun 21;252(5013):1662-7. doi: 10.1126/science.2047875.
- 145.** Paran N, Suezzer Y, Lustig S, Israely T, Schwantes A, Melamed S, Katz L, Preuss T, Hanschmann KM, Kalinke U, Erez N, Levin R, Velan B, Löwer J, Shafferman A, Sutter G. Postexposure immunization with modified vaccinia virus Ankara or conventional Lister vaccine provides solid protection in a murine model of human smallpox. *J Infect Dis.* 2009 Jan 1;199(1):39-48. doi: 10.1086/595565.
- 146.** Jacobs BL, Langland JO, Kibler KV, Denzler KL, White SD, Holechek SA, Wong S, Huynh T, Baskin CR. Vaccinia virus vaccines: past, present and future. *Antiviral Res.* 2009 Oct;84(1):1-13. doi: 10.1016/j.antiviral.2009.06.006.
- 147.** Davidson T. 2017. *Vaccines: history, science, and issues.* Greenwood, Santa Barbara, CA.
- 148.** Chung CS, Chen CH, Ho MY, Huang CY, Liao CL, Chang W. Vaccinia virus proteome: identification of proteins in vaccinia virus intracellular mature virion particles. *J Virol.* 2006 Mar;80(5):2127-40. doi: 10.1128/JVI.80.5.2127-2140.2006.
- 149.** Zeh HJ, Downs-Canner S, McCart JA, Guo ZS, Rao UN, Ramalingam L, Thorne SH, Jones HL, Kalinski P, Wieckowski E, O'Malley ME, Daneshmand M, Hu K, Bell JC, Hwang TH, Moon A, Breitbach CJ, Kirn DH, Bartlett DL. First-in-man study of western reserve strain oncolytic vaccinia virus: safety, systemic spread, and antitumor activity. *Mol Ther.* 2015 Jan;23(1):202-14. doi: 10.1038/mt.2014.194.
- 150.** Adler H, Gould S, Hine P, Snell LB, Wong W, Houlihan CF, Osborne JC, Rampling T, Beadsworth MB, Duncan CJ, Dunning J, Fletcher TE, Hunter ER, Jacobs M, Khoo SH, Newsholme W, Porter D, Porter RJ, Ratcliffe L, Schmid ML, Semple MG, Tunbridge AJ, Wingfield T, Price NM;

NHS England High Consequence Infectious Diseases (Airborne) Network. Clinical features and management of human monkeypox: a retrospective observational study in the UK. *Lancet Infect Dis*. 2022 Aug;22(8):1153-1162. doi: 10.1016/S1473-3099(22)00228-6.

151. Noe S, Zange S, Seilmaier M, Antwerpen MH, Fenzl T, Schneider J, Spinner CD, Bugert JJ, Wendtner CM, Wölfel R. Clinical and virological features of first human monkeypox cases in Germany. *Infection*. 2023 Feb;51(1):265-270. doi: 10.1007/s15010-022-01874-z.

152. Luna N, Ramírez AL, Muñoz M, Ballesteros N, Patiño LH, Castañeda SA, Bonilla-Aldana DK, Paniz-Mondolfi A, Ramírez JD. Phylogenomic analysis of the monkeypox virus (MPXV) 2022 outbreak: Emergence of a novel viral lineage? *Travel Med Infect Dis*. 2022 Sep-Oct;49:102402. doi: 10.1016/j.tmaid.2022.102402.

153. Isidro J, Borges V, Pinto M, Sobral D, Santos JD, Nunes A, Mixão V, Ferreira R, Santos D, Duarte S, Vieira L, Borrego MJ, Nuncio S, de Carvalho IL, Pelerito A, Cordeiro R, Gomes JP. Phylogenomic characterization and signs of microevolution in the 2022 multi-country outbreak of monkeypox virus. *Nat Med*. 2022 Aug;28(8):1569-1572. doi: 10.1038/s41591-022-01907-y.

154. Kumar N, Acharya A, Gendelman HE, Byrareddy SN. The 2022 outbreak and the pathobiology of the monkeypox virus. *J Autoimmun*. 2022 Jul;131:102855. doi: 10.1016/j.jaut.2022.102855.

155. Davison AJ, Moss B. Structure of vaccinia virus early promoters. *J Mol Biol*. 1989 Dec 20;210(4):749-69. doi: 10.1016/0022-2836(89)90107-1.

156. Davison AJ, Moss B. Structure of vaccinia virus late promoters. *J Mol Biol*. 1989 Dec 20;210(4):771-84. doi: 10.1016/0022-2836(89)90108-3.

157. Baldick CJ Jr, Keck JG, Moss B. Mutational analysis of the core, spacer, and initiator regions of vaccinia virus intermediate-class promoters. *J Virol*. 1992 Aug;66(8):4710-9. doi: 10.1128/JVI.66.8.4710-4719.1992.

158. Broyles SS. Vaccinia virus transcription. *J Gen Virol*. 2003 Sep;84(Pt 9):2293-2303. doi: 10.1099/vir.0.18942-0.

159. Yang Z, Bruno DP, Martens CA, Porcella SF, Moss B. Simultaneous high-resolution analysis of vaccinia virus and host cell transcriptomes by deep RNA sequencing. *Proc Natl Acad Sci U S A*. 2010 Jun 22;107(25):11513-8. doi: 10.1073/pnas.1006594107.

160. Yang Z, Reynolds SE, Martens CA, Bruno DP, Porcella SF, Moss B. Expression profiling of the intermediate and late stages of poxvirus replication. *J Virol*. 2011 Oct;85(19):9899-908. doi: 10.1128/JVI.05446-11.

161. Galindo I, Alonso C. African Swine Fever Virus: A Review. *Viruses*. 2017 May 10;9(5):103. doi: 10.3390/v9050103.

162. Gallardo C, Casado N, Soler A, Djadjovski I, Krivko L, Madueño E, Nieto R, Perez C, Simon A, Ivanova E, Donescu D, Milicevik V, Chondrokouki E, Nurmoja I, Frant M, Feliziani F, Václavěk P, Pileviciene S, Marisa A. A multi gene-approach genotyping method identifies 24 genetic clusters within the genotype II-European African swine fever viruses circulating from 2007 to 2022. *Front Vet Sci*. 2023 Jan 25;10:1112850. doi: 10.3389/fvets.2023.1112850.

163. Department for Environment, Food and Rural Affairs Updated Outbreak Assessment #33 African swine fever in Europe. Disease report. 2023 <https://assets.publishing.service.gov.uk/media/64e72de7db1c07001422b391/Update ASF in Europe 33.pdf>

164. Sánchez-Vizcaíno JM, Mur L, Martínez-López B. African swine fever (ASF): five years around Europe. *Vet Microbiol*. 2013 Jul 26;165(1-2):45-50. doi: 10.1016/j.vetmic.2012.11.030.

165. Rodríguez JM, Salas ML. African swine fever virus transcription. *Virus Res.* 2013 Apr;173(1):15-28. doi: 10.1016/j.virusres.2012.09.014.
166. Dixon LK, Sun H, Roberts H. African swine fever. *Antiviral Res.* 2019 May;165:34-41. doi: 10.1016/j.antiviral.2019.02.018.
167. Rohrmann, G. F. (2013). *Baculovirus Molecular Biology [Third Edition]*. National Center for Biotechnology Information (US)
168. King LA, Possee RD. (1992). *The Baculovirus Expression System: A laboratory guide*. London: Chapman & Hall.
169. van Oers MM, Pijlman GP, Vlak JM. Thirty years of baculovirus-insect cell protein expression: from dark horse to mainstream technology. *J Gen Virol.* 2015 Jan;96(Pt 1):6-23. doi: 10.1099/vir.0.067108-0.
170. Kogan PH, Chen X, Blissard GW. Overlapping TATA-dependent and TATA-independent early promoter activities in the baculovirus gp64 envelope fusion protein gene. *J Virol.* 1995 Mar;69(3):1452-61. doi: 10.1128/JVI.69.3.1452-1461.1995.
171. Chen YR, Zhong S, Fei Z, Hashimoto Y, Xiang JZ, Zhang S, Blissard GW. The transcriptome of the baculovirus *Autographa californica* multiple nucleopolyhedrovirus in *Trichoplusia ni* cells. *J Virol.* 2013 Jun;87(11):6391-405. doi: 10.1128/JVI.00194-13.
172. Garrity DB, Chang MJ, Blissard GW. Late promoter selection in the baculovirus gp64 envelope fusion protein gene. *Virology.* 1997 May 12;231(2):167-81. doi: 10.1006/viro.1997.8540.
173. Jin J, Guarino LA. 3'-end formation of baculovirus late RNAs. *J Virol.* 2000 Oct;74(19):8930-7. doi: 10.1128/jvi.74.19.8930-8937.2000.
174. Campillo-Balderas JA, Lazcano A, Becerra A: Viral Genome Size Distribution Does not Correlate with the Antiquity of the Host Lineages. *Front Ecol Evol.* 2015. 3:143. <https://doi.org/10.3389/fevo.2015.00143>
175. Tischer I, Rasch R, Tochtermann G. Characterization of papovavirus-and picornavirus-like particles in permanent pig kidney cell lines. *Zentralbl Bakteriol Orig A.* 1974 Feb;226(2):153-67.
176. Tischer I, Gelderblom H, Vettermann W, Koch MA. A very small porcine virus with circular single-stranded DNA. *Nature.* 1982 Jan 7;295(5844):64-6. doi: 10.1038/295064a0.
177. LeCann P, Albina E, Madec F, Cariolet R, Jestin A. Piglet wasting disease. *Vet Rec.* 1997 141, 660.
178. 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. doi: 10.1128/genomeA.01539-16.
179. Bratanich AC, Blanchetot A. PCV2 replicase transcripts in infected porcine kidney (PK15) cells. *Virus Genes.* 2002 Dec;25(3):323-8. doi: 10.1023/a:1020940411267.
180. Chaiyakul M, Hsu K, Dardari R, Marshall F, Czub M. Cytotoxicity of ORF3 proteins from a nonpathogenic and a pathogenic porcine circovirus. *J Virol.* 2010 Nov;84(21):11440-7. doi: 10.1128/JVI.01030-10.
181. Cheung AK. Comparative analysis of the transcriptional patterns of pathogenic and nonpathogenic porcine circoviruses. *Virology.* 2003 May 25;310(1):41-9. doi: 10.1016/s0042-6822(03)00096-5.
182. Belshaw R, Pereira V, Katzourakis A, Talbot G, Paces J, Burt A, Tristem M. Long-term reinfection of the human genome by endogenous retroviruses. *Proc Natl Acad Sci U S A.* 2004 Apr 6;101(14):4894-9. doi: 10.1073/pnas.0307800101.

- 183.** Nelson PN, Hooley P, Roden D, Davari Ejtchadi H, Rylance P, Warren P, Martin J, Murray PG; Molecular Immunology Research Group. Human endogenous retroviruses: transposable elements with potential? *Clin Exp Immunol*. 2004 Oct;138(1):1-9. doi: 10.1111/j.1365-2249.2004.02592.x.
- 184.** Tönjes RR, Löwer R, Boller K, Denner J, Hasenmaier B, Kirsch H, König H, Korbmacher C, Limbach C, Lugert R, Phelps RC, Scherer J, Thelen K, Löwer J, Kurth R. HERV-K: the biologically most active human endogenous retrovirus family. *J Acquir Immune Defic Syndr Hum Retrovirol*. 1996;13 Suppl 1:S261-7. doi: 10.1097/00042560-199600001-00039.
- 185.** Prangishvili D, Bamford DH, Forterre P, Iranzo J, Koonin EV, Krupovic M. The enigmatic archaeal virosphere. *Nat Rev Microbiol*. 2017 Nov 10;15(12):724-739. doi: 10.1038/nrmicro.2017.125.
- 186.** Bolduc B, Shaughnessy DP, Wolf YI, Koonin EV, Roberto FF, Young M. Identification of novel positive-strand RNA viruses by metagenomic analysis of archaea-dominated Yellowstone hot springs. *J Virol*. 2012 May;86(10):5562-73. doi: 10.1128/JVI.07196-11.
- 187.** Bolduc B, Wirth JF, Mazurie A, Young MJ. Viral assemblage composition in Yellowstone acidic hot springs assessed by network analysis. *ISME J*. 2015 Oct;9(10):2162-77. doi: 10.1038/ismej.2015.28.
- 188.** Fields BN, Knipe DM, Howley PM. (2013). *Fields Virology*. Wolters Kluwer Health/Lippincott Williams & Wilkins.
- 189.** Clark DP, Pazdernik NJ, McGehee MR. *Molecular Biology (Third Edition)*, Chapter 24 - Viruses, Viroids, and Prions, Editor(s): David P. Clark, Nanette J. Pazdernik, Michelle R. McGehee, Academic Cell, 2019, Pages 749-792, ISBN 9780128132883, <https://doi.org/10.1016/B978-0-12-813288-3.00024-0>.
- 190.** Cong Y, Sun Y, Deng X, Yu H, Lian X, Cong Y. A SYBR Green-based real-time RT-PCR assay to differentiate the H1N1 influenza virus lineages. *J Virol Methods*. 2022 Feb;300:114387. doi: 10.1016/j.jviromet.2021.114387.
- 191.** Bouvier NM, Palese P. The biology of influenza viruses. *Vaccine*. 2008 Sep 12;26 Suppl 4(Suppl 4):D49-53. doi: 10.1016/j.vaccine.2008.07.039.
- 192.** Chen W, Calvo PA, Malide D, Gibbs J, Schubert U, Bacik I, Basta S, O'Neill R, Schickli J, Palese P, Henklein P, Bennink JR, Yewdell JW. A novel influenza A virus mitochondrial protein that induces cell death. *Nat Med*. 2001 Dec;7(12):1306-12. doi: 10.1038/nm1201-1306.
- 193.** Lamb RA, Lai CJ, Choppin PW. Sequences of mRNAs derived from genome RNA segment 7 of influenza virus: colinear and interrupted mRNAs code for overlapping proteins. *Proc Natl Acad Sci U S A*. 1981 Jul;78(7):4170-4. doi: 10.1073/pnas.78.7.4170.
- 194.** Dauber B, Heins G, Wolff T. The influenza B virus nonstructural NS1 protein is essential for efficient viral growth and antagonizes beta interferon induction. *J Virol*. 2004 Feb;78(4):1865-72. doi: 10.1128/jvi.78.4.1865-1872.2004.
- 195.** Kochs G, García-Sastre A, Martínez-Sobrido L. Multiple anti-interferon actions of the influenza A virus NS1 protein. *J Virol*. 2007 Jul;81(13):7011-21. doi: 10.1128/JVI.02581-06.
- 196.** Briedis DJ, Lamb RA. Influenza B virus genome: sequences and structural organization of RNA segment 8 and the mRNAs coding for the NS1 and NS2 proteins. *J Virol*. 1982 Apr;42(1):186-93. doi: 10.1128/JVI.42.1.186-193.1982.
- 197.** Lamb RA, Choppin PW, Chanock RM, Lai CJ. Mapping of the two overlapping genes for polypeptides NS1 and NS2 on RNA segment 8 of influenza virus genome. *Proc Natl Acad Sci U S A*. 1980 Apr;77(4):1857-61. doi: 10.1073/pnas.77.4.1857.

- 198.** Wasik BR, Bhushan A, Ogbunugafor CB, Turner PE. Delayed transmission selects for increased survival of vesicular stomatitis virus. *Evolution*. 2015 Jan;69(1):117-25. doi: 10.1111/evo.12544.
- 199.** Rihn SJ, Aziz MA, Stewart DG, Hughes J, Turnbull ML, Varela M, Sugrue E, Herd CS, Stanifer M, Sinkins SP, Palmarini M, Wilson SJ. TRIM69 Inhibits Vesicular Stomatitis Indiana Virus. *J Virol*. 2019 Sep 30;93(20):e00951-19. doi: 10.1128/JVI.00951-19.
- 200.** Johnson KM, Vogel JE, Peralta PH. Clinical and serological response to laboratory-acquired human infection by Indiana type vesicular stomatitis virus (VSV). *Am J Trop Med Hyg*. 1966 Mar;15(2):244-6. doi: 10.4269/ajtmh.1966.15.244.
- 201.** Cargnelutti JF, Olinda RG, Maia LA, de Aguiar GM, Neto EG, Simões SV, de Lima TG, Dantas AF, Weiblen R, Flores EF, Riet-Correa F. Outbreaks of Vesicular stomatitis Alagoas virus in horses and cattle in northeastern Brazil. *J Vet Diagn Invest*. 2014 Nov;26(6):788-94. doi: 10.1177/1040638714553428.
- 202.** Liang B, Li Z, Jenni S, Rahmeh AA, Morin BM, Grant T, Grigorieff N, Harrison SC, Whelan SPJ. Structure of the L Protein of Vesicular Stomatitis Virus from Electron Cryomicroscopy. *Cell*. 2015 Jul 16;162(2):314-327. doi: 10.1016/j.cell.2015.06.018.
- 203.** Heilmann E, Kimpel J, Geley S, Naschberger A, Urbiola C, Nolden T, von Laer D, Wollmann G. The Methyltransferase Region of Vesicular Stomatitis Virus L Polymerase Is a Target Site for Functional Intramolecular Insertion. *Viruses*. 2019 Oct 26;11(11):989. doi: 10.3390/v11110989.
- 204.** Patil G, Xu L, Wu Y, Song K, Hao W, Hua F, Wang L, Li S. TRIM41-Mediated Ubiquitination of Nucleoprotein Limits Vesicular Stomatitis Virus Infection. *Viruses*. 2020 Jan 22;12(2):131. doi: 10.3390/v12020131.
- 205.** Banerjee AD, Abraham G, Colonno RJ. Vesicular stomatitis virus: mode of transcription. *J Gen Virol*. 1977 Jan;34(1):1-8. doi: 10.1099/0022-1317-34-1-1.
- 206.** Pan W, Song D, He W, Lu H, Lan Y, Tong J, Gao F, Zhao K. The matrix protein of vesicular stomatitis virus inhibits host-directed transcription of target genes via interaction with the TFIIF subunit p8. *Vet Microbiol*. 2017 Sep;208:82-88. doi: 10.1016/j.vetmic.2017.07.020.
- 207.** Nikolic J, Belot L, Raux H, Legrand P, Gaudin Y, Albertini A. Structural basis for the recognition of LDL-receptor family members by VSV glycoprotein. *Nat Commun*. 2018 Mar 12;9(1):1029. doi: 10.1038/s41467-018-03432-4.
- 208.** Hawman DW, Feldmann H. Crimean-Congo haemorrhagic fever virus. *Nat Rev Microbiol*. 2023 Jul;21(7):463-477. doi: 10.1038/s41579-023-00871-9.
- 209.** Simpson DI, Knight EM, Courtois G, Williams MC, Weinbren MP, Kibukamusoke JW. Congo virus: a hitherto undescribed virus occurring in Africa. I. Human isolations--clinical notes. *East Afr Med J*. 1967 Feb;44(2):86-92.
- 210.** Messina JP, Pigott DM, Golding N, Duda KA, Brownstein JS, Weiss DJ, Gibson H, Robinson TP, Gilbert M, William Wint GR, Nuttall PA, Gething PW, Myers MF, George DB, Hay SI. The global distribution of Crimean-Congo hemorrhagic fever. *Trans R Soc Trop Med Hyg*. 2015 Aug;109(8):503-13. doi: 10.1093/trstmh/trv050.
- 211.** Magyar N, Kis Z, Barabás É, Nagy A, Henczkó J, Damjanova I, Takács M, Pályi B. New geographical area on the map of Crimean-Congo hemorrhagic fever virus: First serological evidence in the Hungarian population. *Ticks Tick Borne Dis*. 2021 Jan;12(1):101555. doi: 10.1016/j.ttbdis.2020.101555.

- 212.** Földvári G, Szabó É, Tóth GE, Lanszki Z, Zana B, Varga Z, Kemenesi G. Emergence of *Hyalomma marginatum* and *Hyalomma rufipes* adults revealed by citizen science tick monitoring in Hungary. *Transbound Emerg Dis*. 2022 Sep;69(5):e2240-e2248. doi: 10.1111/tbed.14563.
- 213.** Celina SS, Černý J, Samy AM. Mapping the potential distribution of the principal vector of Crimean-Congo haemorrhagic fever virus *Hyalomma marginatum* in the Old World. *PLoS Negl Trop Dis*. 2023 Nov 27;17(11):e0010855. doi: 10.1371/journal.pntd.0010855.
- 214.** Rainey T, Occi JL, Robbins RG, Egizi A. Discovery of *Haemaphysalis longicornis* (Ixodida: Ixodidae) Parasitizing a Sheep in New Jersey, United States. *J Med Entomol*. 2018 May 4;55(3):757-759. doi: 10.1093/jme/tjy006.
- 215.** Hayes CG. West Nile virus: Uganda, 1937, to New York City, 1999. *Ann N Y Acad Sci*. 2001 Dec;951:25-37. doi: 10.1111/j.1749-6632.2001.tb02682.x.
- 216.** Hachid A, Beloufa MA, Seghier M, Bahoura N, Dia M, Fall G, Sall AA. Evidence of West Nile virus circulation among humans in central northern Algeria. *New Microbes New Infect*. 2019 Mar 8;29:100512. doi: 10.1016/j.nmni.2019.01.008.
- 217.** Clark MB, Schaefer TJ. West Nile Virus. 2023 Aug 8. In: StatPearls [Internet]. Treasure Island (FL): StatPearls Publishing; 2024 Jan–.
- 218.** Brinton MA. Replication cycle and molecular biology of the West Nile virus. *Viruses*. 2013 Dec 27;6(1):13-53. doi: 10.3390/v6010013.
- 219.** Lindenbach B.D., Thiel H.J., Rice C.M. Flaviviridae: The viruses and their replication. In: Knipe D.M., Howley P.M., editors. *Fields Virology*. Lippincott Williams and Wilkins; Philadelphia, PA, USA: 2007. pp. 1101–1152.
- 220.** Basu R, Tumban E. Zika Virus on a Spreading Spree: what we now know that was unknown in the 1950's. *Virol J*. 2016 Oct 6;13(1):165. doi: 10.1186/s12985-016-0623-2.
- 221.** Duffy MR, Chen TH, Hancock WT, Powers AM, Kool JL, Lanciotti RS, Pretrick M, Marfel M, Holzbauer S, Dubray C, Guillaumot L, Griggs A, Bel M, Lambert AJ, Laven J, Kosoy O, Panella A, Biggerstaff BJ, Fischer M, Hayes EB. Zika virus outbreak on Yap Island, Federated States of Micronesia. *N Engl J Med*. 2009 Jun 11;360(24):2536-43. doi: 10.1056/NEJMoa0805715.
- 222.** Lanciotti RS, Kosoy OL, Laven JJ, Velez JO, Lambert AJ, Johnson AJ, Stanfield SM, Duffy MR. Genetic and serologic properties of Zika virus associated with an epidemic, Yap State, Micronesia, 2007. *Emerg Infect Dis*. 2008 Aug;14(8):1232-9. doi: 10.3201/eid1408.080287.
- 223.** Del Campo M, Feitosa IM, Ribeiro EM, Horovitz DD, Pessoa AL, França GV, García-Alix A, Doriqui MJ, Wanderley HY, Sanseverino MV, Neri JJ, Pina-Neto JM, Santos ES, Verçosa I, Cernach MC, Medeiros PF, Kerbage SC, Silva AA, van der Linden V, Martelli CM, Cordeiro MT, Dhalia R, Vianna FS, Victora CG, Cavalcanti DP, Schuler-Faccini L; Zika Embryopathy Task Force-Brazilian Society of Medical Genetics ZETF-SBGM. The phenotypic spectrum of congenital Zika syndrome. *Am J Med Genet A*. 2017 Apr;173(4):841-857. doi: 10.1002/ajmg.a.38170.
- 224.** Oehler E, Watrin L, Larre P, Leparc-Goffart I, Lastere S, Valour F, Baudouin L, Mallet H, Musso D, Ghawche F. Zika virus infection complicated by Guillain-Barre syndrome--case report, French Polynesia, December 2013. *Euro Surveill*. 2014 Mar 6;19(9):20720. doi: 10.2807/1560-7917.es2014.19.9.20720.
- 225.** Aubry M, Finke J, Teissier A, Roche C, Brout J, Paulous S, Desprès P, Cao-Lormeau VM, Musso D. Seroprevalence of arboviruses among blood donors in French Polynesia, 2011-2013. *Int J Infect Dis*. 2015 Dec;41:11-2. doi: 10.1016/j.ijid.2015.10.005.

- 226.** Besnard M, Lastere S, Teissier A, Cao-Lormeau V, Musso D. Evidence of perinatal transmission of Zika virus, French Polynesia, December 2013 and February 2014. *Euro Surveill.* 2014 Apr 3;19(13):20751.
- 227.** Bonaldo MC, Ribeiro IP, Lima NS, Dos Santos AA, Menezes LS, da Cruz SO, de Mello IS, Furtado ND, de Moura EE, Damasceno L, da Silva KA, de Castro MG, Gerber AL, de Almeida LG, Lourenço-de-Oliveira R, Vasconcelos AT, Brasil P. Isolation of Infective Zika Virus from Urine and Saliva of Patients in Brazil. *PLoS Negl Trop Dis.* 2016 Jun 24;10(6):e0004816. doi: 10.1371/journal.pntd.0004816.
- 228.** Lindenbach BD, Rice CM. Molecular biology of flaviviruses. *Adv Virus Res.* 2003;59:23-61. doi: 10.1016/s0065-3527(03)59002-9.
- 229.** Shi W, Zhang Z, Ling C, Carr MJ, Tong Y, Gao GF. Increasing genetic diversity of Zika virus in the Latin American outbreak. *Emerg Microbes Infect.* 2016 Jul 6;5(7):e68. doi: 10.1038/emi.2016.68.
- 230.** Enjuanes L, Gorbalenya AE, de Groot RJ, Cowley JA, Ziebuhr J, Snijder EJ. Nidovirales. *Encyclopedia of Virology.* 2008:419–30. doi: 10.1016/B978-012374410-4.00775-5.
- 231.** Naqvi AAT, Fatima K, Mohammad T, Fatima U, Singh IK, Singh A, Atif SM, Hariprasad G, Hasan GM, Hassan MI. Insights into SARS-CoV-2 genome, structure, evolution, pathogenesis and therapies: Structural genomics approach. *Biochim Biophys Acta Mol Basis Dis.* 2020 Oct 1;1866(10):165878. doi: 10.1016/j.bbadis.2020.165878.
- 232.** Lu R, Zhao X, Li J, Niu P, Yang B, Wu H, Wang W, Song H, Huang B, Zhu N, Bi Y, Ma X, Zhan F, Wang L, Hu T, Zhou H, Hu Z, Zhou W, Zhao L, Chen J, Meng Y, Wang J, Lin Y, Yuan J, Xie Z, Ma J, Liu WJ, Wang D, Xu W, Holmes EC, Gao GF, Wu G, Chen W, Shi W, Tan W. Genomic characterisation and epidemiology of 2019 novel coronavirus: implications for virus origins and receptor binding. *Lancet.* 2020 Feb 22;395(10224):565-574. doi: 10.1016/S0140-6736(20)30251-8.
- 233.** Zhou P, Yang XL, Wang XG, Hu B, Zhang L, Zhang W, Si HR, Zhu Y, Li B, Huang CL, Chen HD, Chen J, Luo Y, Guo H, Jiang RD, Liu MQ, Chen Y, Shen XR, Wang X, Zheng XS, Zhao K, Chen QJ, Deng F, Liu LL, Yan B, Zhan FX, Wang YY, Xiao GF, Shi ZL. A pneumonia outbreak associated with a new coronavirus of probable bat origin. *Nature.* 2020 Mar;579(7798):270-273. doi: 10.1038/s41586-020-2012-7.
- 234.** Mathieu E, Ritchie H, Rodés-Guirao L, Appel C, Giattino C, Hasell J, Macdonald B, Dattani S, Beltekian D, Ortiz-Ospina E, Roser M (2020) - "Coronavirus Pandemic (COVID-19)". Published online at OurWorldInData.org. Retrieved from: 'https://ourworldindata.org/coronavirus'
- 235.** Worldometer 2024: <https://www.worldometers.info/coronavirus/>
- 236.** Woo PCY, Huang Y, Lau SKP, Yuen KY. Coronavirus genomics and bioinformatics analysis. *Viruses.* 2010 Aug;2(8):1804-1820. doi: 10.3390/v2081803.
- 237.** Sola I, Moreno JL, Zúñiga S, Alonso S, Enjuanes L. Role of nucleotides immediately flanking the transcription-regulating sequence core in coronavirus subgenomic mRNA synthesis. *J Virol.* 2005 Feb;79(4):2506-16. doi: 10.1128/JVI.79.4.2506-2516.2005.
- 238.** Hussain S, Pan J, Chen Y, Yang Y, Xu J, Peng Y, Wu Y, Li Z, Zhu Y, Tien P, Guo D. Identification of novel subgenomic RNAs and noncanonical transcription initiation signals of severe acute respiratory syndrome coronavirus. *J Virol.* 2005 May;79(9):5288-95. doi: 10.1128/JVI.79.9.5288-5295.2005.
- 239.** Sola I, Almazán F, Zúñiga S, Enjuanes L. Continuous and Discontinuous RNA Synthesis in Coronaviruses. *Annu Rev Virol.* 2015 Nov;2(1):265-88. doi: 10.1146/annurev-virology-100114-055218.

- 240.** Hu YF, Zhao R, Xue Y, Yang F, Jin Q. Full genome sequence of a novel coxsackievirus B5 strain isolated from neurological hand, foot, and mouth disease patients in China. *J Virol.* 2012 Oct;86(20):11408-9. doi: 10.1128/JVI.01709-12.
- 241.** Koonin EV, Dolja VV, Krupovic M, Varsani A, WolfYI, Yutin N, Zerbini M, Kuhn JH. Create a megataxonomic framework, filling all principal taxonomic ranks, for realm Riboviria. Report number: ICTV Proposal No. 2019.006G. doi: 10.13140/RG.2.2.20234.21444
- 242.** Denner J. 2010. Endogenous retroviruses, p 35–69 In Kurth R, Bannert N. (ed), *Retroviruses: molecular biology, genomics and pathogenesis*. Caister Academic Press, Hethersett, Norwich, United Kingdom
- 243.** Lieber MM, Sherr CJ, Benveniste RE, Todaro GJ. Biologic and immunologic properties of porcine type C viruses. *Virology.* 1975 Aug;66(2):616-9. doi: 10.1016/0042-6822(75)90234-2.
- 244.** Todaro GJ, Benveniste RE, Lieber MM, Sherr CJ. Characterization of a type C virus released from the porcine cell line PK(15). *Virology.* 1974 Mar;58(1):65-74. doi: 10.1016/0042-6822(74)90141-x.
- 245.** Czauderna F, Fischer N, Boller K, Kurth R, Tönjes RR. Establishment and characterization of molecular clones of porcine endogenous retroviruses replicating on human cells. *J Virol.* 2000 May;74(9):4028-38. doi: 10.1128/jvi.74.9.4028-4038.2000.
- 246.** Akiyoshi DE, Denaro M, Zhu H, Greenstein JL, Banerjee P, Fishman JA. Identification of a full-length cDNA for an endogenous retrovirus of miniature swine. *J Virol.* 1998 May;72(5):4503-7. doi: 10.1128/JVI.72.5.4503-4507.1998.
- 247.** Blusch JH, Seelmeir S, von der Helm K. Molecular and enzymatic characterization of the porcine endogenous retrovirus protease. *J Virol.* 2002 Aug;76(15):7913-7. doi: 10.1128/jvi.76.15.7913-7917.2002.
- 248.** Vernet G: *Genomics of Infectious Diseases and Private Industry*. Editor(s): Michel Tibayrenc, *Genetics and Evolution of Infectious Diseases (Second Edition)*, Elsevier, 2017, Pages 421-434, <https://doi.org/10.1016/B978-0-12-799942-5.00018-4>.
- 249.** Hughes GM, Gang L, Murphy WJ, Higgins DG, Teeling EC. Using Illumina next generation sequencing technologies to sequence multigene families in de novo species. *Mol Ecol Resour.* 2013 May;13(3):510-21. doi: 10.1111/1755-0998.12087.
- 250.** Minoche AE, Dohm JC, Himmelbauer H. Evaluation of genomic high-throughput sequencing data generated on Illumina HiSeq and genome analyzer systems. *Genome Biol.* 2011 Nov 8;12(11):R112. doi: 10.1186/gb-2011-12-11-r112.
- 251.** Domcke S, Hill AJ, Daza RM, Cao J, O'Day DR, Pliner HA, Aldinger KA, Pokholok D, Zhang F, Milbank JH, Zager MA, Glass IA, Steemers FJ, Doherty D, Trapnell C, Cusanovich DA, Shendure J. A human cell atlas of fetal chromatin accessibility. *Science.* 2020 Nov 13;370(6518):eaba7612. doi: 10.1126/science.aba7612.
- 252.** Garalde DR, Snell EA, Jachimowicz D, Sipos B, Lloyd JH, Bruce M, Pantic N, Admassu T, James P, Warland A, Jordan M, Ciccone J, Serra S, Keenan J, Martin S, McNeill L, Wallace EJ, Jayasinghe L, Wright C, Blasco J, Young S, Brocklebank D, Juul S, Clarke J, Heron AJ, Turner DJ. Highly parallel direct RNA sequencing on an array of nanopores. *Nat Methods.* 2018 Mar;15(3):201-206. doi: 10.1038/nmeth.4577.
- 253.** Schatz MC. Nanopore sequencing meets epigenetics. *Nat Methods.* 2017 Mar 31;14(4):347-348. doi: 10.1038/nmeth.4240.

- 254.** Warburton PE, Sebra RP. Long-Read DNA Sequencing: Recent Advances and Remaining Challenges. *Annu Rev Genomics Hum Genet.* 2023 Aug 25;24:109-132. doi: 10.1146/annurev-genom-101722-103045.
- 255.** 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. doi: 10.3389/fmicb.2023.1179966.
- 256.** Gorcenco S, Ilinca A, Almasoudi W, Kafantari E, Lindgren AG, Puschmann A. New generation genetic testing entering the clinic. *Parkinsonism Relat Disord.* 2020 Apr;73:72-84. doi: 10.1016/j.parkreldis.2020.02.015.
- 257.** Pourmohammadi R, Abouei J, Anpalagan A. Error analysis of the PacBio sequencing CCS reads. *Int J Biostat.* 2023 May 8;19(2):439-453. doi: 10.1515/ijb-2021-0091.
- 258.** Zaragoza-Solas A, Haro-Moreno JM, Rodriguez-Valera F, López-Pérez M. Long-Read Metagenomics Improves the Recovery of Viral Diversity from Complex Natural Marine Samples. *mSystems.* 2022 Jun 28;7(3):e0019222. doi: 10.1128/msystems.00192-22.
- 259.** Kim C, Pongpanich M, Porntaveetus T. Unraveling metagenomics through long-read sequencing: a comprehensive review. *J Transl Med.* 2024 Jan 28;22(1):111. doi: 10.1186/s12967-024-04917-1.
- 260.** Agostinho DP, Fu Y, Menon VK, Metcalf GA, Treangen TJ, Sedlazeck FJ. Unveiling microbial diversity: harnessing long-read sequencing technology. *Nat Methods.* 2024 Apr 30. doi: 10.1038/s41592-024-02262-1.
- 261.** Gibson G, Muse SV. A Primer of Genome Science. Third Edition. Published: 01 January 2009. 344 Pages ISBN: 9780878932368
- 262.** 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. doi: 10.1016/j.tim.2019.01.010.
- 263.** Oikonomopoulos S, Bayega A, Fahiminiya S, Djambazian H, Berube P, Ragoussis J. Methodologies for Transcript Profiling Using Long-Read Technologies. *Front Genet.* 2020 Jul 7;11:606. doi: 10.3389/fgene.2020.00606.
- 264.** Eid J, Fehr A, Gray J, Luong K, Lyle J, Otto G, Peluso P, Rank D, Baybayan P, Bettman B, Bibillo A, Bjornson K, Chaudhuri B, Christians F, Cicero R, Clark S, Dalal R, Dewinter A, Dixon J, Foquet M, Gaertner A, Hardenbol P, Heiner C, Hester K, Holden D, Kearns G, Kong X, Kuse R, Lacroix Y, Lin S, Lundquist P, Ma C, Marks P, Maxham M, Murphy D, Park I, Pham T, Phillips M, Roy J, Sebra R, Shen G, Sorenson J, Tomaney A, Travers K, Trulson M, Vieceli J, Wegener J, Wu D, Yang A, Zaccarin D, Zhao P, Zhong F, Korf J, Turner S. Real-time DNA sequencing from single polymerase molecules. *Science.* 2009 Jan 2;323(5910):133-8. doi: 10.1126/science.1162986.
- 265.** Scott D, Ely B. Comparison of genome sequencing technology and assembly methods for the analysis of a GC-rich bacterial genome. *Curr Microbiol.* 2015 Mar;70(3):338-44. doi: 10.1007/s00284-014-0721-6.
- 266.** Miyamoto M, Motooka D, Gotoh K, Imai T, Yoshitake K, Goto N, Iida T, Yasunaga T, Horii T, Arakawa K, Kasahara M, Nakamura S. Performance comparison of second- and third-generation sequencers using a bacterial genome with two chromosomes. *BMC Genomics.* 2014 Aug 21;15(1):699. doi: 10.1186/1471-2164-15-699.
- 267.** Tilgner H, Jahanbani F, Blauwkamp T, Moshrefi A, Jaeger E, Chen F, Harel I, Bustamante CD, Rasmussen M, Snyder MP. Comprehensive transcriptome analysis using synthetic long-read

sequencing reveals molecular co-association of distant splicing events. *Nat Biotechnol.* 2015 Jul;33(7):736-42. doi: 10.1038/nbt.3242.

268. <https://www.elementbiosciences.com/blog/what-is-loopseq-long-read-chemistry>

269. 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. doi: 10.3389/fgene.2018.00259.

270. 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. doi: 10.1038/s41592-019-0617-2.

271. McGeoch DJ, Dalrymple MA, Dolan A, McNab D, Perry LJ, Taylor P, Challberg MD. Structures of herpes simplex virus type 1 genes required for replication of virus DNA. *J Virol.* 1988 Feb;62(2):444-53. doi: 10.1128/JVI.62.2.444-453.1988.

272. 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. doi: 10.3390/v7052727.

273. 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. doi: 10.1038/s41598-018-26955-8.

274. 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. doi: 10.1093/gigascience/giy139.

275. Boldogkői Z. Transcriptional interference networks coordinate the expression of functionally related genes clustered in the same genomic loci. *Front Genet.* 2012 Jul 5;3:122. doi: 10.3389/fgene.2012.00122.

276. Olasz F, Tombácz D, Torma G, Csabai Z, Moldován N, Dörmő Á, Prazsák 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. *Front Genet.* 2020 Jul 28;11:758. doi: 10.3389/fgene.2020.00758.

277. Boldogkői Z, Moldován N, Szűcs A, Tombácz D. Transcriptome-wide analysis of a baculovirus using nanopore sequencing. *Sci Data.* 2018 Dec 4;5:180276. doi: 10.1038/sdata.2018.276.

278. 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. *Sci Data.* 2018 Jun 19;5:180119. doi: 10.1038/sdata.2018.119.

279. 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 Feb 10;493(2):235-42. doi: 10.1016/j.gene.2011.11.049.

280. 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. doi: 10.1007/s00705-017-3420-3.

281. 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 Jul;98(1):15-25. doi: 10.1016/j.ygeno.2011.03.010.

- 282.** 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. doi: 10.3389/fgene.2021.619056.
- 283.** 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. *Sci Rep.* 2021 Jul 9;11(1):14219. doi: 10.1038/s41598-021-93142-7.
- 284.** 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.
- 285.** 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. doi: 10.1016/j.dib.2022.108386.
- 286.** 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. doi: 10.1038/sdata.2017.194.
- 287.** 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. doi: 10.3389/fgene.2018.00432.
- 288.** 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. doi: 10.3389/fgene.2019.00834.
- 289.** 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. doi: 10.1038/s41597-020-0558-8.
- 290.** 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 α promote lytic de novo KSHV infection. *J Virol.* 2023 Nov 30;97(11):e0097223. doi: 10.1128/jvi.00972-23.
- 291.** 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. doi: 10.1038/s41597-023-02149-4.
- 292.** 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. doi: 10.1016/j.virusres.2017.05.010.
- 293.** 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. doi: 10.1093/femsle/fny013.
- 294.** 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. doi: 10.1128/genomeA.01040-17.

- 295.** 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. doi: 10.1128/genomeA.01570-17.
- 296.** 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. doi: 10.1186/s13104-021-05657-x.
- 297.** 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. doi: 10.3390/pathogens10080919.
- 298.** Tombácz D, Prazsák I, Moldován N, Szűcs A, Boldogkői Z. Lytic Transcriptome Dataset of Varicella Zoster Virus Generated by Long-Read Sequencing. *Front Genet.* 2018 Oct 16;9:460. doi: 10.3389/fgene.2018.00460.
- 299.** 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. doi: 10.1128/JVI.02989-13.
- 300.** Travers KJ, Chin CS, Rank DR, Eid JS, Turner SW. A flexible and efficient template format for circular consensus sequencing and SNP detection. *Nucleic Acids Res.* 2010 Aug;38(15):e159. doi: 10.1093/nar/gkq543.
- 301.** Clark TA, Murray IA, Morgan RD, Kislyuk AO, Spittle KE, Boitano M, Fomenkov A, Roberts RJ, Korlach J. Characterization of DNA methyltransferase specificities using single-molecule, real-time DNA sequencing. *Nucleic Acids Res.* 2012 Feb;40(4):e29. doi: 10.1093/nar/gkr1146.
- 302.** Tcherepanov V, Ehlers A, Upton C. Genome Annotation Transfer Utility (GATU): rapid annotation of viral genomes using a closely related reference genome. *BMC Genomics.* 2006 Jun 13;7:150. doi: 10.1186/1471-2164-7-150.
- 303.** Wu TD, Watanabe CK. GMAP: a genomic mapping and alignment program for mRNA and EST sequences. *Bioinformatics.* 2005 May 1;21(9):1859-75. doi: 10.1093/bioinformatics/bti310.
- 304.** Kearse M, Moir R, Wilson A, Stones-Havas S, Cheung M, Sturrock S, Buxton S, Cooper A, Markowitz S, Duran C, Thierer T, Ashton B, Meintjes P, Drummond A. Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics.* 2012 Jun 15;28(12):1647-9. doi: 10.1093/bioinformatics/bts199.
- 305.** Shafin K, Pesout T, Lorig-Roach R, Haukness M, Olsen HE, Bosworth C, Armstrong J, Tigyi K, Maurer N, Koren S, Sedlazeck FJ, Marschall T, Mayes S, Costa V, Zook JM, Liu KJ, Kilburn D, Sorensen M, Munson KM, Vollger MR, Monlong J, Garrison E, Eichler EE, Salama S, Haussler D, Green RE, Akeson M, Phillippy A, Miga KH, Carnevali P, Jain M, Paten B. Nanopore sequencing and the Shasta toolkit enable efficient de novo assembly of eleven human genomes. *Nat Biotechnol.* 2020 Sep;38(9):1044-1053. doi: 10.1038/s41587-020-0503-6.
- 306.** Aksamentov I, Roemer C, Hodcroft EB, et al. Nextclade: clade assignment, mutation calling and quality control for viral genomes. *J Open Source Softw.* 2021;6:3773.
- 307.** Liao H, Cai D, Sun Y. VirStrain: a strain identification tool for RNA viruses. *Genome Biol.* 2022 Jan 31;23(1):38. doi: 10.1186/s13059-022-02609-x.
- 308.** Kakuk B. Github repository of scripts used in "High Temporal-Resolution Nanopore Sequencing Dataset of SARS-CoV-2 and Host Cell RNAs." 2022. https://github.com/Balays/SARS-CoV2_ONT_data

- 309.** Wickham H. *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York, 2016.
- 310.** Morgan M, Pagès H, Obenchain V, Hayden N (2023). Rsamtools: Binary alignment (BAM), FASTA, variant call (BCF), and tabix file import. doi:10.18129/B9.bioc.Rsamtools, R package version 2.18.0, <https://bioconductor.org/packages/Rsamtools>.
- 311.** Lawrence M, Huber W, Pagès H, Aboyoun P, Carlson M, Gentleman R, Morgan MT, Carey VJ. Software for computing and annotating genomic ranges. *PLoS Comput Biol*. 2013;9(8):e1003118. doi: 10.1371/journal.pcbi.1003118.
- 312.** Ahlmann-Eltze C. <https://const-ae.name/project/tidygenomics/>
- 313.** Wickham H, François R, Henry L, Müller K, Vaughan D (2023). dplyr: A Grammar of Data Manipulation. R package version 1.1.4, <https://github.com/tidyverse/dplyr>, <https://dplyr.tidyverse.org>.
- 314.** Chaisson MJ, Tesler G. Mapping single molecule sequencing reads using basic local alignment with successive refinement (BLASR): application and theory. *BMC Bioinformatics*. 2012 Sep 19;13:238. doi: 10.1186/1471-2105-13-238.
- 315.** Cock PJ, Antao T, Chang JT, Chapman BA, Cox CJ, Dalke A, Friedberg I, Hamelryck T, Kauff F, Wilczynski B, de Hoon MJ. Biopython: freely available Python tools for computational molecular biology and bioinformatics. *Bioinformatics*. 2009 Jun 1;25(11):1422-3. doi: 10.1093/bioinformatics/btp163.
- 316.** Bryne JC, Valen E, Tang MH, Marstrand T, Winther O, da Piedade I, Krogh A, Lenhard B, Sandelin A. JASPAR, the open access database of transcription factor-binding profiles: new content and tools in the 2008 update. *Nucleic Acids Res*. 2008 Jan;36(Database issue):D102-6. doi: 10.1093/nar/gkm955.
- 317.** Ahmed F, Kumar M, Raghava GP. Prediction of polyadenylation signals in human DNA sequences using nucleotide frequencies. *In Silico Biol*. 2009;9(3):135-48.
- 318.** Lee TY, Chang WC, Hsu JB, Chang TH, Shien DM. GPMIner: an integrated system for mining combinatorial cis-regulatory elements in mammalian gene group. *BMC Genomics*. 2012;13 Suppl 1(Suppl 1):S3. doi: 10.1186/1471-2164-13-S1-S3.
- 319.** Li H. Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics*. 2018 Sep 15;34(18):3094-3100. doi: 10.1093/bioinformatics/bty191.
- 320.** Drost, H.-G. & Gogleva, A. metablast: Perform Massive Local BLAST Searches. (2020).
- 321.** Shumate A, Salzberg SL. Liftoff: accurate mapping of gene annotations. *Bioinformatics*. 2021 Jul 19;37(12):1639-1643. doi: 10.1093/bioinformatics/btaa1016.
- 322.** Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R; 1000 Genome Project Data Processing Subgroup. The Sequence Alignment/Map format and SAMtools. *Bioinformatics*. 2009 Aug 15;25(16):2078-9. doi: 10.1093/bioinformatics/btp352.
- 323.** Tjeldnes, H. et al. ORFik: a comprehensive R toolkit for the analysis of translation. *BioRxiv* (2021).
- 324.** Hahne F, Ivanek R. Visualizing Genomic Data Using Gviz and Bioconductor. *Methods Mol Biol*. 2016;1418:335-51. doi: 10.1007/978-1-4939-3578-9_16.
- 325.** Trapnell C, Pachter L, Salzberg SL. TopHat: discovering splice junctions with RNA-Seq. *Bioinformatics*. 2009 May 1;25(9):1105-11. doi: 10.1093/bioinformatics/btp120
- 326.** Kim D, Pertea G, Trapnell C, Pimentel H, Kelley R, Salzberg SL. TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. *Genome Biol*. 2013 Apr 25;14(4):R36. doi: 10.1186/gb-2013-14-4-r36.

- 327.** Martin M. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet J.* 2011, 17, 10–12.
- 328.** Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. *Nat Methods.* 2012 Mar 4;9(4):357-9. doi: 10.1038/nmeth.1923.
- 329.** Heinz S, Benner C, Spann N, Bertolino E, Lin YC, Laslo P, Cheng JX, Murre C, Singh H, Glass CK. Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. *Mol Cell.* 2010 May 28;38(4):576-89. doi: 10.1016/j.molcel.2010.05.004.
- 330.** Rutherford K, Parkhill J, Crook J, Horsnell T, Rice P, Rajandream MA, Barrell B. Artemis: sequence visualization and annotation. *Bioinformatics.* 2000 Oct;16(10):944-5. doi: 10.1093/bioinformatics/16.10.944.
- 331.** Robinson JT, Thorvaldsdóttir H, Winckler W, Guttman M, Lander ES, Getz G, Mesirov JP. Integrative genomics viewer. *Nat Biotechnol.* 2011 Jan;29(1):24-6. doi: 10.1038/nbt.1754.
- 332.** Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, Batut P, Chaisson M, Gingeras TR. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics.* 2013 Jan 1;29(1):15-21. doi: 10.1093/bioinformatics/bts635. Epub 2012 Oct 25.
- 333.** Thodberg M, Thieffry A, Vitting-Seerup K, Andersson R, Sandelin A. CAGEfightR: analysis of 5'-end data using R/Bioconductor. *BMC Bioinformatics.* 2019 Oct 4;20(1):487. doi: 10.1186/s12859-019-3029-5.
- 334.** Carver T, Harris SR, Berriman M, Parkhill J, McQuillan JA. Artemis: an integrated platform for visualization and analysis of high-throughput sequence-based experimental data. *Bioinformatics.* 2012 Feb 15;28(4):464-9. doi: 10.1093/bioinformatics/btr703.
- 335.** Crooks GE, Hon G, Chandonia JM, Brenner SE. WebLogo: a sequence logo generator. *Genome Res.* 2004 Jun;14(6):1188-90. doi: 10.1101/gr.849004. PMID: 15173120; PMCID: PMC419797.
- 336.** Gu Z, Eils R, Schlesner M. Complex heatmaps reveal patterns and correlations in multidimensional genomic data. *Bioinformatics.* 2016 Sep 15;32(18):2847-9. doi: 10.1093/bioinformatics/btw313.
- 337.** Wang L, Park HJ, Dasari S, Wang S, Kocher JP, Li W. CPAT: Coding-Potential Assessment Tool using an alignment-free logistic regression model. *Nucleic Acids Res.* 2013 Apr 1;41(6):e74. doi: 10.1093/nar/gkt006.
- 338.** 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. doi: 10.1093/bioinformatics/btz247.
- 339.** Almagro Armenteros JJ, Tsirigos KD, Sønderby CK, Petersen TN, Winther O, Brunak S, von Heijne G, Nielsen H. SignalP 5.0 improves signal peptide predictions using deep neural networks. *Nat Biotechnol.* 2019 Apr;37(4):420-423. doi: 10.1038/s41587-019-0036-z.
- 340.** Mészáros B, Erdos G, Dosztányi Z. IUPred2A: context-dependent prediction of protein disorder as a function of redox state and protein binding. *Nucleic Acids Res.* 2018 Jul 2;46(W1):W329-W337. doi: 10.1093/nar/gky384.
- 341.** Reich M, Ohm K, Angelo M, Tamayo P, Mesirov JP: GeneCluster 2.0: an advanced toolset for bioarray analysis. *Bioinformatics.* 2004, 20(11):1797-1798.
- 342.** Saffran HA, Pare JM, Corcoran JA, Weller SK, Smiley JR. Herpes simplex virus eliminates host mitochondrial DNA. *EMBO Rep.* 2007 Feb;8(2):188-93. doi: 10.1038/sj.embor.7400878.

- 343.** Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* 2014;15(12):550. doi: 10.1186/s13059-014-0550-8.
- 344.** Depledge DP, Srinivas KP, Sadaoka T, Bready D, Mori Y, Placantonakis DG, Mohr I, Wilson AC. Direct RNA sequencing on nanopore arrays redefines the transcriptional complexity of a viral pathogen. *Nat Commun.* 2019 Feb 14;10(1):754. doi: 10.1038/s41467-019-08734-9.
- 345.** Patro R, Duggal G, Love MI, Irizarry RA, Kingsford C. Salmon provides fast and bias-aware quantification of transcript expression. *Nat Methods.* 2017 Apr;14(4):417-419. doi: 10.1038/nmeth.4197. g
- 346.** Sander J, Schultze JL, Yosef N. ImpulseDE: detection of differentially expressed genes in time series data using impulse models. *Bioinformatics.* 2017 Mar 1;33(5):757-759. doi: 10.1093/bioinformatics/btw665.
- 347.** McCarthy DJ, Chen Y, Smyth GK. Differential expression analysis of multifactor RNA-Seq experiments with respect to biological variation. *Nucleic Acids Res.* 2012 May;40(10):4288-97. doi: 10.1093/nar/gks042.
- 348.** Mi H, Muruganujan A, Casagrande JT, Thomas PD. Large-scale gene function analysis with the PANTHER classification system. *Nat Protoc.* 2013 Aug;8(8):1551-66. doi: 10.1038/nprot.2013.092.
- 349.** Caliński T, Harabasz J. A dendrite method for cluster analysis. *Commun Stat.* 1974 3 (1): 1–27. g
- 350.** Mestdagh P, Van Vlierberghe P, De Weer A, Muth D, Westermann F, Speleman F, Vandesompele J. A novel and universal method for microRNA RT-qPCR data normalization. *Genome Biol.* 2009;10(6):R64. doi: 10.1186/gb-2009-10-6-r64.
- 351.** Ausubel FM, Brent R, Kingston RE, Moore DD, Seidman JG, Smith JA, et al., editors. Current Protocols in Molecular Biology. 4th ed. Hoboken, NJ, USA: John Wiley & Sons, Inc.; 1999.
- 352.** Boldogkői Z, Braun A, Antal J, Fodor I. A restriction cleavage and transfection system for introducing foreign DNA sequences into the genome of a herpesvirus. *Res Virol.* 1998 Mar-Apr;149(2):87-97. doi: 10.1016/s0923-2516(98)80084-8.
- 353.** Boldogkői Z, Braun A, Fodor I. Replication and virulence of early protein 0 and long latency transcript deficient mutants of the Aujeszky's disease (pseudorabies) virus. *Microbes Infect.* 2000 Sep;2(11):1321-8. doi: 10.1016/s1286-4579(00)01285-5.
- 354.** Elhai J, Wolk CP. A versatile class of positive-selection vectors based on the nonviability of palindrome-containing plasmids that allows cloning into long polylinkers. *Gene.* 1988 Aug 15;68(1):119-38. doi: 10.1016/0378-1119(88)90605-1.
- 355.** Ibrahim F, Oppelt J, Maragkakis M, Mourelatos Z. TERA-Seq: true end-to-end sequencing of native RNA molecules for transcriptome characterization. *Nucleic Acids Res.* 2021 Nov 18;49(20):e115. doi: 10.1093/nar/gkab713.
- 356.** Kulpa D, Topping R, Telesnitsky A. Determination of the site of first strand transfer during Moloney murine leukemia virus reverse transcription and identification of strand transfer-associated reverse transcriptase errors. *EMBO J.* 1997 Feb 17;16(4):856-65. doi: 10.1093/emboj/16.4.856.
- 357.** Zhu YY, Machleder EM, Chenchik A, Li R, Siebert PD. Reverse transcriptase template switching: a SMART approach for full-length cDNA library construction. *Biotechniques.* 2001 Apr;30(4):892-7. doi: 10.2144/01304pf02.

- 358.** Ramsköld D, Luo S, Wang YC, Li R, Deng Q, Faridani OR, Daniels GA, Khrebtukova I, Loring JF, Laurent LC, Schroth GP, Sandberg R. Full-length mRNA-Seq from single-cell levels of RNA and individual circulating tumor cells. *Nat Biotechnol*. 2012 Aug;30(8):777-82. doi: 10.1038/nbt.2282.
- 359.** Minshall, N., Git, A. Enzyme- and gene-specific biases in reverse transcription of RNA raise concerns for evaluating gene expression. *Sci Rep*. 2020 10, 8151. doi: 10.1038/s41598-020-65005-0.
- 360.** Shagin DA, Lukyanov KA, Vagner LL, Matz MV. Regulation of average length of complex PCR product. *Nucleic Acids Res*. 1999 Sep 15;27(18):e23. doi: 10.1093/nar/27.18.e23.
- 361.** Bayega A, Oikonomopoulos S, Wang YC, Ragoussis J. Improved Nanopore full-length cDNA sequencing by PCR-suppression. *Front Genet*. 2022 Oct 17;13:1031355. doi: 10.3389/fgene.2022.1031355.
- 362.** 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. doi: 10.1186/s12864-019-6199-7.
- 363.** Klupp BG, Hengartner CJ, Mettenleiter TC, Enquist LW. Complete, annotated sequence of the pseudorabies virus genome. *J Virol*. 2004 Jan;78(1):424-40. doi: 10.1128/jvi.78.1.424-440.2004.
- 364.** Grimm KS, Klupp BG, Granzow H, Müller FM, Fuchs W, Mettenleiter TC. Analysis of viral and cellular factors influencing herpesvirus-induced nuclear envelope breakdown. *J Virol*. 2012 Jun;86(12):6512-21. doi: 10.1128/JVI.00068-12.
- 365.** Szpara ML, Tafuri YR, Parsons L, Shamim SR, Verstrepen KJ, Legendre M, Enquist LW. A wide extent of inter-strain diversity in virulent and vaccine strains of alphaherpesviruses. *PLoS Pathog*. 2011 Oct;7(10):e1002282. doi: 10.1371/journal.ppat.1002282.
- 366.** Yu T, Chen F, Ku X, Zhu Y, Ma H, Li S, He Q. Complete Genome Sequence of Novel Pseudorabies Virus Strain HNB Isolated in China. *Genome Announc*. 2016 Jan 28;4(1):e01641-15. doi: 10.1128/genomeA.01641-15.
- 367.** Tang S, Patel A, Krause PR. Hidden regulation of herpes simplex virus 1 pre-mRNA splicing and polyadenylation by virally encoded immediate early gene ICP27. *PLoS Pathog*. 2019 Jun 17;15(6):e1007884. doi: 10.1371/journal.ppat.1007884.
- 368.** Rutkowski AJ, Erhard F, L'Hernault A, Bonfert T, Schilhabel M, Crump C, Rosenstiel P, Efstathiou S, Zimmer R, Friedel CC, Dölken L. Widespread disruption of host transcription termination in HSV-1 infection. *Nat Commun*. 2015 May 20;6:7126. doi: 10.1038/ncomms8126.
- 369.** Whisnant AW, Jürges CS, Hennig T, Wyler E, Prusty B, Rutkowski AJ, L'hernault A, Djakovic L, Göbel M, Döring K, Menegatti J, Antrobus R, Matheson NJ, Künzig FWH, Mastrobuoni G, Bielow C, Kempa S, Liang C, Dandekar T, Zimmer R, Landthaler M, Grässer F, Lehner PJ, Friedel CC, Erhard F, Dölken L. Integrative functional genomics decodes herpes simplex virus 1. *Nat Commun*. 2020 Apr 27;11(1):2038. doi: 10.1038/s41467-020-15992-5.
- 370.** Pheasant K, Möller-Levet CS, Jones J, Depledge D, Breuer J, Elliott G. Nuclear-cytoplasmic compartmentalization of the herpes simplex virus 1 infected cell transcriptome is co-ordinated by the viral endoribonuclease vhs and cofactors to facilitate the translation of late proteins. *PLoS Pathog*. 2018 Nov 26;14(11):e1007331. doi: 10.1371/journal.ppat.1007331.
- 371.** Xing K, Deng R, Wang J, Feng J, Huang M, Wang X. Analysis and prediction of baculovirus promoter sequences. *Virus Research*. 2005 113(1):64-71. doi: 10.1016/j.virusres.2005.04.016.
- 372.** Cherbas L, Cherbas P. The arthropod initiator: the capsite consensus plays an important role in transcription. *Insect Biochem. Mol. Biol*. 23, 81–90 (1993).

- 373.** Palazzo AF, Lee ES. Non-coding RNA: what is functional and what is junk? *Front Genet.* 2015 Jan 26;6:2. doi: 10.3389/fgene.2015.00002.
- 374.** Barkley LR, Santocanale C. MicroRNA-29a regulates the benzo[a]pyrene dihydrodiol epoxide-induced DNA damage response through Cdc7 kinase in lung cancer cells. *Oncogenesis.* 2013 Jul 22;2(7):e57. doi: 10.1038/oncsis.2013.20.
- 375.** Marchese FP, Huarte M. A long noncoding RNA in DNA replication and chromosome dynamics. *Cell Cycle.* 2017 Jan 17;16(2):151-152. doi: 10.1080/15384101.2016.1241604.
- 376.** Pal S, Gupta R, Kim H, Wickramasinghe P, Baubet V, Showe LC, Dahmane N, Davuluri RV. Alternative transcription exceeds alternative splicing in generating the transcriptome diversity of cerebellar development. *Genome Res.* 2011 Aug;21(8):1260-72. doi: 10.1101/gr.120535.111.
- 377.** Isomura H, Stinski MF, Kudoh A, Murata T, Nakayama S, Sato Y, Iwahori S, Tsurumi T. Noncanonical TATA sequence in the UL44 late promoter of human cytomegalovirus is required for the accumulation of late viral transcripts. *J Virol.* 2008 Feb;82(4):1638-46. doi: 10.1128/JVI.01917-07.
- 378.** Leppek K, Das R, Barna M. Functional 5' UTR mRNA structures in eukaryotic translation regulation and how to find them. *Nat Rev Mol Cell Biol.* 2018 Mar;19(3):158-174. doi: 10.1038/nrm.2017.103. Epub 2017 Nov 22.
- 379.** Akirtava C, McManus CJ. Control of translation by eukaryotic mRNA transcript leaders- Insights from high-throughput assays and computational modeling. *Wiley Interdiscip Rev RNA.* 2021 May;12(3):e1623. doi: 10.1002/wrna.1623.
- 380.** Vincent HA, Ziehr B, Moorman NJ. Human Cytomegalovirus Strategies to Maintain and Promote mRNA Translation. *Viruses.* 2016 Apr 13;8(4):97. doi: 10.3390/v8040097.
- 381.** Wethmar K. The regulatory potential of upstream open reading frames in eukaryotic gene expression. *Wiley Interdiscip Rev RNA.* 2014 Nov-Dec;5(6):765-78. doi: 10.1002/wrna.1245.
- 382.** Stern-Ginossar N, Weisburd B, Michalski A, Le VT, Hein MY, Huang SX, Ma M, Shen B, Qian SB, Hengel H, Mann M, Ingolia NT, Weissman JS. Decoding human cytomegalovirus. *Science.* 2012 Nov 23;338(6110):1088-93. doi: 10.1126/science.1227919.
- 383.** Mainguy G, Koster J, Woltering J, Jansen H, Durston A. Extensive polycistronism and antisense transcription in the mammalian Hox clusters. *PLoS One.* 2007 Apr 4;2(4):e356. doi: 10.1371/journal.pone.0000356.
- 384.** Gullerova M, Proudfoot NJ. Cohesin complex promotes transcriptional termination between convergent genes in *S. pombe*. *Cell.* 2008 Mar 21;132(6):983-95. doi: 10.1016/j.cell.2008.02.040.
- 385.** Majerciak V, Ni T, Yang W, Meng B, Zhu J, Zheng ZM. A viral genome landscape of RNA polyadenylation from KSHV latent to lytic infection. *PLoS Pathog.* 2013;9(11):e1003749. doi: 10.1371/journal.ppat.1003749.
- 386.** Nam DK, Lee S, Zhou G, Cao X, Wang C, Clark T, Chen J, Rowley JD, Wang SM. Oligo(dT) primer generates a high frequency of truncated cDNAs through internal poly(A) priming during reverse transcription. *Proc Natl Acad Sci U S A.* 2002 Apr 30;99(9):6152-6. doi: 10.1073/pnas.092140899.
- 387.** Abdel-Ghany SE, Hamilton M, Jacobi JL, Ngam P, Devitt N, Schilkey F, Ben-Hur A, Reddy AS. A survey of the sorghum transcriptome using single-molecule long reads. *Nat Commun.* 2016 Jun 24;7:11706. doi: 10.1038/ncomms11706.
- 388.** Gonzalez-Garay ML. Introduction to Isoform Sequencing Using Pacific Biosciences Technology (Iso-Seq). Springer, 2016: 141-160

- 389.** Wang B, Tseng E, Regulski M, Clark TA, Hon T, Jiao Y, Lu Z, Olson A, Stein JC, Ware D. Unveiling the complexity of the maize transcriptome by single-molecule long-read sequencing. *Nat Commun*. 2016 Jun 24;7:11708. doi: 10.1038/ncomms11708.
- 390.** Shih SR, Nemeroff ME, Krug RM. The choice of alternative 5' splice sites in influenza virus M1 mRNA is regulated by the viral polymerase complex. *Proc Natl Acad Sci U S A*. 1995 Jul 3;92(14):6324-8. doi: 10.1073/pnas.92.14.6324. PMID: 7541537; PMCID: PMC41510.
- 391.** O'Grady T, Wang X, Höner Zu Bentrop K, Baddoo M, Concha M, Flemington EK. Global transcript structure resolution of high gene density genomes through multi-platform data integration. *Nucleic Acids Res*. 2016 Oct 14;44(18):e145. doi: 10.1093/nar/gkw629.
- 392.** Sommer G, Heise T. Posttranscriptional control of HBV gene expression. *Front Biosci*. 2008 May 1;13:5533-47. doi: 10.2741/3097.
- 393.** Wirth UV, Vogt B, Schwyzer M. The three major immediate-early transcripts of bovine herpesvirus 1 arise from two divergent and spliced transcription units. *J Virol*. 1991 Jan;65(1):195-205. doi: 10.1128/JVI.65.1.195-205.1991.
- 394.** Fan D, Wang M, Cheng A, Jia R, Yang Q, Wu Y, Zhu D, Zhao X, Chen S, Liu M, Zhang S, Ou X, Mao S, Gao Q, Sun D, Wen X, Liu Y, Yu Y, Zhang L, Tian B, Pan L, Chen X. The Role of VP16 in the Life Cycle of Alphaherpesviruses. *Front Microbiol*. 2020 Aug 18;11:1910. doi: 10.3389/fmicb.2020.01910.
- 395.** Boldogkői Z, Balázs Z, Moldován N, Prazsák I, Tombácz D. Novel classes of replication-associated transcripts discovered in viruses. *RNA Biol*. 2019 Feb;16(2):166-175. doi: 10.1080/15476286.2018.1564468.
- 396.** Ye X, Zhao Y, Karijovich J. The landscape of transcription initiation across latent and lytic KSHV genomes. *PLoS Pathog*. 2019 Jun 12;15(6):e1007852. doi: 10.1371/journal.ppat.1007852.
- 397.** Gaffney DF, McLauchlan J, Whitton JL, Clements JB. A modular system for the assay of transcription regulatory signals: the sequence TAATGARAT is required for herpes simplex virus immediate early gene activation. *Nucleic Acids Res*. 1985 Nov 11;13(21):7847-63. doi: 10.1093/nar/13.21.7847.
- 398.** 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 Jun;55(3):274-279. doi: 10.1007/s11262-019-01643-5.
- 399.** Hein MY, Weissman JS. Functional single-cell genomics of human cytomegalovirus infection. *Nat Biotechnol*. 2022 Mar;40(3):391-401. doi: 10.1038/s41587-021-01059-3.
- 400.** Koons MD, Van Scoy S, Hearing J. The replicator of the Epstein-Barr virus latent cycle origin of DNA replication, oriP, is composed of multiple functional elements. *J Virol*. 2001 Nov;75(22):10582-92. doi: 10.1128/JVI.75.22.10582-10592.2001.
- 401.** Purushothaman P, Uppal T, Verma SC. Molecular biology of KSHV lytic reactivation. *Viruses*. 2015 Jan 14;7(1):116-53. doi: 10.3390/v7010116.
- 402.** Purushothaman P, Thakker S, Verma SC. Transcriptome analysis of Kaposi's sarcoma-associated herpesvirus during de novo primary infection of human B and endothelial cells. *J Virol*. 2015 Mar;89(6):3093-111. doi: 10.1128/JVI.02507-14.
- 403.** Zhu FX, Cusano T, Yuan Y. Identification of the immediate-early transcripts of Kaposi's sarcoma-associated herpesvirus. *J Virol*. 1999 Jul;73(7):5556-67. doi: 10.1128/JVI.73.7.5556-5567.1999.

404. Majerciak V, Alvarado-Hernandez B, Lobanov A, Cam M, Zheng ZM. Genome-wide regulation of KSHV RNA splicing by viral RNA-binding protein ORF57. *PLoS Pathog*. 2022 Jul 14;18(7):e1010311. doi: 10.1371/journal.ppat.1010311.
405. Arias C, Weisburd B, Stern-Ginossar N, Mercier A, Madrid AS, Bellare P, Holdorf M, Weissman JS, Ganem D. KSHV 2.0: a comprehensive annotation of the Kaposi's sarcoma-associated herpesvirus genome using next-generation sequencing reveals novel genomic and functional features. *PLoS Pathog*. 2014 Jan;10(1):e1003847. doi: 10.1371/journal.ppat.1003847.
406. Bruce AG, Barcy S, DiMaio T, Gan E, Garrigues HJ, Lagunoff M, Rose TM. Quantitative Analysis of the KSHV Transcriptome Following Primary Infection of Blood and Lymphatic Endothelial Cells. *Pathogens*. 2017 Mar 19;6(1):11. doi: 10.3390/pathogens6010011.
407. Pearce M, Matsumura S, Wilson AC. Transcripts encoding K12, v-FLIP, v-cyclin, and the microRNA cluster of Kaposi's sarcoma-associated herpesvirus originate from a common promoter. *J Virol*. 2005 Nov;79(22):14457-64. doi: 10.1128/JVI.79.22.14457-14464.2005.
408. Schifano JM, Corcoran K, Kelkar H, Dittmer DP. Expression of the Antisense-to-Latency Transcript Long Noncoding RNA in Kaposi's Sarcoma-Associated Herpesvirus. *J Virol*. 2017 Jan 31;91(4):e01698-16. doi: 10.1128/JVI.01698-16.
409. 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. doi: 10.1016/j.virol.2005.12.036.
410. Flori L, Rogel-Gaillard C, Cochet M, Lemonnier G, Hugot K, Chardon P, Robin S, Lefèvre F. Transcriptomic analysis of the dialogue between Pseudorabies virus and porcine epithelial cells during infection. *BMC Genomics*. 2008 Mar 10;9:123. doi: 10.1186/1471-2164-9-123.
411. Harkness JM, Kader M, DeLuca NA. Transcription of the herpes simplex virus 1 genome during productive and quiescent infection of neuronal and nonneuronal cells. *J Virol*. 2014 Jun;88(12):6847-61. doi: 10.1128/JVI.00516-14.
412. Tóth JS, Tombácz D, Takács IF, Boldogkoi Z. The effects of viral load on pseudorabies virus gene expression. *BMC Microbiol*. 2010 Dec 6;10:311. doi: 10.1186/1471-2180-10-311. PMID: 21134263; PMCID: PMC3016322.
413. Wirth UV, Gunkel K, Engels M, Schwyzer M. Spatial and temporal distribution of bovine herpesvirus 1 transcripts. *J Virol*. 1989 Nov;63(11):4882-9. doi: 10.1128/JVI.63.11.4882-4889.1989.
414. Barber KA, Daugherty HC, Ander SE, Jefferson VA, Shack LA, Pechan T, Nanduri B, Meyer F. Protein Composition of the Bovine Herpesvirus 1.1 Virion. *Vet Sci*. 2017 Feb 20;4(1):11. doi: 10.3390/vetsci4010011.
415. Fraefel C, Wirth UV, Vogt B, Schwyzer M. Immediate-early transcription over covalently joined genome ends of bovine herpesvirus 1: the circ gene. *J Virol*. 1993 Mar;67(3):1328-33. doi: 10.1128/JVI.67.3.1328-1333.1993.
416. Some Genetic Functions Encoded by Herpes Simplex Virus Type 1. Available online: <http://darwin.bio.uci.edu/~faculty/wagner/table.html> (accessed on 3 October 2003).
417. The Genetic and Transcription Map of the HSV-1 Genome. Available online: <http://darwin.bio.uci.edu/~faculty/wagner/hsvimg04z.jpg> (accessed on 3 October 2003).
418. Roizman B. The function of herpes simplex virus genes: a primer for genetic engineering of novel vectors. *Proc Natl Acad Sci U S A*. 1996 Oct 15;93(21):11307-12. doi: 10.1073/pnas.93.21.11307.

- 419.** Roizman B, Campadelli-Fiume G. Alphaherpes viral genes and their functions. In: Arvin A, Campadelli-Fiume G, Mocarski E, Moore PS, Roizman B, Whitley R, Yamanishi K, editors. *Human Herpesviruses: Biology, Therapy, and Immunoprophylaxis*. Cambridge: Cambridge University Press; 2007. Chapter 6. PMID: 21348062.
- 420.** Vitting-Seerup K, Sandelin A. The Landscape of Isoform Switches in Human Cancers. *Mol Cancer Res*. 2017 Sep;15(9):1206-1220. doi: 10.1158/1541-7786.MCR-16-0459.
- 421.** Kakuk B, Kiss AA, Torma G, Csabai Z, Prazsák 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 Sep 15;10(9):1196. doi: 10.3390/pathogens10091196. PMID: 34578228; PMCID: PMC8468008.
- 422.** Osenberg S, Dominissini D, Rechavi G, Eisenberg E. Widespread cleavage of A-to-I hyperediting substrates. *RNA*. 2009 Sep;15(9):1632-9. doi: 10.1261/rna.1581809.
- 423.** Matthews MM, Thomas JM, Zheng Y, Tran K, Phelps KJ, Scott AI, Havel J, Fisher AJ, Beal PA. Structures of human ADAR2 bound to dsRNA reveal base-flipping mechanism and basis for site selectivity. *Nat Struct Mol Biol*. 2016 May;23(5):426-33. doi: 10.1038/nsmb.3203.
- 424.** Baumeister J, Klupp BG, Mettenleiter TC. Pseudorabies virus and equine herpesvirus 1 share a nonessential gene which is absent in other herpesviruses and located adjacent to a highly conserved gene cluster. *J Virol*. 1995 Sep;69(9):5560-7. doi: 10.1128/JVI.69.9.5560-5567.1995.
- 425.** Sacks WR, Greene CC, Aschman DP, Schaffer PA. Herpes simplex virus type 1 ICP27 is an essential regulatory protein. *J Virol*. 1985 Sep;55(3):796-805. doi: 10.1128/JVI.55.3.796-805.1985.
- 426.** Hardwicke MA, Sandri-Goldin RM. The herpes simplex virus regulatory protein ICP27 contributes to the decrease in cellular mRNA levels during infection. *J Virol*. 1994 Aug;68(8):4797-810. doi: 10.1128/JVI.68.8.4797-4810.1994.
- 427.** Hayashi ML, Blankenship C, Shenk T. Human cytomegalovirus UL69 protein is required for efficient accumulation of infected cells in the G1 phase of the cell cycle. *Proc Natl Acad Sci U S A*. 2000 Mar 14;97(6):2692-6. doi: 10.1073/pnas.050587597.
- 428.** Schwartz JA, Brittle EE, Reynolds AE, Enquist LW, Silverstein SJ. UL54-null pseudorabies virus is attenuated in mice but productively infects cells in culture. *J Virol*. 2006 Jan;80(2):769-84. doi: 10.1128/JVI.80.2.769-784.2006.
- 429.** Chang YY, Lin HW, Wong ML, Chang TJ. Regulation of the vhs gene promoter of pseudorabies virus by IE180 and EP0, and the requirement of a Sp1 Site for the promoter function. *Virus Genes*. 2004 Apr;28(3):247-58. doi: 10.1023/b:viru.0000025772.67359.f8.
- 430.** Lin HW, Hsu WL, Chang YY, Jan MS, Wong ML, Chang TJ. Role of the UL41 protein of pseudorabies virus in host shutoff, pathogenesis and induction of TNF- α expression. *J Vet Med Sci*. 2010 Sep;72(9):1179-87. doi: 10.1292/jvms.10-0059.
- 431.** Liang L, Roizman B. Herpes simplex virus 1 precludes replenishment of the short-lived receptor of tumor necrosis factor alpha by virion host shutoff-dependent degradation of its mRNA. *J Virol*. 2006 Aug;80(15):7756-9. doi: 10.1128/JVI.00587-06.
- 432.** Zelus BD, Stewart RS, Ross J. The virion host shutoff protein of herpes simplex virus type 1: messenger ribonucleolytic activity in vitro. *J Virol*. 1996 Apr;70(4):2411-9. doi: 10.1128/JVI.70.4.2411-2419.1996.
- 433.** Smibert CA, Popova B, Xiao P, Capone JP, Smiley JR. Herpes simplex virus VP16 forms a complex with the virion host shutoff protein vhs. *J Virol*. 1994 Apr;68(4):2339-46. doi: 10.1128/JVI.68.4.2339-2346.1994.

- 434.** Smith TJ, Ackland-Berglund CE, Leib DA. Herpes simplex virus virion host shutoff (vhs) activity alters periocular disease in mice. *J Virol.* 2000 Apr;74(8):3598-604. doi: 10.1128/jvi.74.8.3598-3604.2000.
- 435.** Read GS, Frenkel N. Herpes simplex virus mutants defective in the virion-associated shutoff of host polypeptide synthesis and exhibiting abnormal synthesis of alpha (immediate early) viral polypeptides. *J Virol.* 1983 May;46(2):498-512. doi: 10.1128/JVI.46.2.498-512.1983.
- 436.** Smibert CA, Smiley JR. Differential regulation of endogenous and transduced beta-globin genes during infection of erythroid cells with a herpes simplex virus type 1 recombinant. *J Virol.* 1990 Aug;64(8):3882-94. doi: 10.1128/JVI.64.8.3882-3894.1990.
- 437.** Oroskar AA, Read GS. Control of mRNA stability by the virion host shutoff function of herpes simplex virus. *J Virol.* 1989 May;63(5):1897-906. doi: 10.1128/JVI.63.5.1897-1906.1989.
- 438.** Jurado J, Fuentes-Almagro CA, Prieto-Alamo MJ, Pueyo C. Alternative splicing of c-fos pre-mRNA: contribution of the rates of synthesis and degradation to the copy number of each transcript isoform and detection of a truncated c-Fos immunoreactive species. *BMC Mol Biol.* 2007 Sep 21;8:83. doi: 10.1186/1471-2199-8-83.
- 439.** Hu B, Li X, Huo Y, Yu Y, Zhang Q, Chen G, Zhang Y, Fraser NW, Wu D, Zhou J. Cellular responses to HSV-1 infection are linked to specific types of alterations in the host transcriptome. *Sci Rep.* 2016 Jun 29;6:28075. doi: 10.1038/srep28075.
- 440.** Rubio N, Martin-Clemente B. Theiler's murine encephalomyelitis virus infection induces early expression of c-fos in astrocytes. *Virology.* 1999 May 25;258(1):21-9. doi: 10.1006/viro.1999.9684. PMID: 10329564.
- 441.** Pereira LA, Munita R, González MP, Andrés ME. Long 3'UTR of Nurr1 mRNAs is targeted by miRNAs in mesencephalic dopamine neurons. *PLoS One.* 2017 Nov 16;12(11):e0188177. doi: 10.1371/journal.pone.0188177.
- 442.** Miao L, St Clair DK. Regulation of superoxide dismutase genes: implications in disease. *Free Radic Biol Med.* 2009 Aug 15;47(4):344-56. doi: 10.1016/j.freeradbiomed.2009.05.018.
- 443.** Bhattacharya A, Hegazy AN, Deigendesch N, Kosack L, Cupovic J, Kandasamy RK, Hildebrandt A, Merkler D, Kühl AA, Vilagos B, Schliehe C, Panse I, Khamina K, Baazim H, Arnold I, Flatz L, Xu HC, Lang PA, Aderem A, Takaoka A, Superti-Furga G, Colinge J, Ludewig B, Löhning M, Bergthaler A. Superoxide Dismutase 1 Protects Hepatocytes from Type I Interferon-Driven Oxidative Damage. *Immunity.* 2015 Nov 17;43(5):974-86. doi: 10.1016/j.immuni.2015.10.013.
- 444.** Smiley JR. Herpes simplex virus virion host shutoff protein: immune evasion mediated by a viral RNase? *J Virol.* 2004 Feb;78(3):1063-8. doi: 10.1128/jvi.78.3.1063-1068.2004.
- 445.** Moon SL, Wilusz J. Cytoplasmic viruses: rage against the (cellular RNA decay) machine. *PLoS Pathog.* 2013;9(12):e1003762. doi: 10.1371/journal.ppat.1003762.
- 446.** Eisenberg E, Levanon EY. Human housekeeping genes, revisited. *Trends Genet.* 2013 Oct;29(10):569-74. doi: 10.1016/j.tig.2013.05.010. Epub 2013 Jun 27.
- 447.** Vo Ngoc L, Kassavetis GA, Kadonaga JT. The RNA Polymerase II Core Promoter in *Drosophila*. *Genetics.* 2019 May;212(1):13-24. doi: 10.1534/genetics.119.302021.
- 448.** 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. doi: 10.1371/journal.pone.0015615.
- 449.** Bourquain D, Dabrowski PW, Nitsche A. Comparison of host cell gene expression in cowpox, monkeypox or vaccinia virus-infected cells reveals virus-specific regulation of immune response genes. *Virol J.* 2013 Feb 20;10:61. doi: 10.1186/1743-422X-10-61.

- 450.** 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. doi: 10.1016/j.celrep.2019.04.042.
- 451.** Herzog VA, Reichholf B, Neumann T, Rescheneder P, Bhat P, Burkard TR, Wlotzka W, von Haeseler A, Zuber J, Ameres SL. Thiol-linked alkylation of RNA to assess expression dynamics. *Nat Methods.* 2017 Dec;14(12):1198-1204. doi: 10.1038/nmeth.4435.
- 452.** Kit Leng Lui S, Keegan S, Tonzi P, Kahli M, Chen YH, Chalhoub N, Coleman KE, Fenyo D, Smith DJ, Huang TT. Monitoring genome-wide replication fork directionality by Okazaki fragment sequencing in mammalian cells. *Nat Protoc.* 2021 Feb;16(2):1193-1218. doi: 10.1038/s41596-020-00454-5.
- 453.** Rampelli S, Turrone S, Schnorr SL, Soverini M, Quercia S, Barone M, Castagnetti A, Biagi E, Gallinella G, Brigidi P, Candela M. Characterization of the human DNA gut virome across populations with different subsistence strategies and geographical origin. *Environ Microbiol.* 2017 Nov;19(11):4728-4735. doi: 10.1111/1462-2920.13938.
- 454.** Maya-Lucas O, Murugesan S, Nirmalkar K, Alcaraz LD, Hoyo-Vadillo C, Pizano-Zárate ML, García-Mena J. The gut microbiome of Mexican children affected by obesity. *Anaerobe.* 2019 Feb;55:11-23. doi: 10.1016/j.anaerobe.2018.10.009.
- 455.** Woelfle T, Linkohr B, Waterboer T, Thorand B, Seissler J, Chadeau-Hyam M, Peters A. Health impact of seven herpesviruses on (pre)diabetes incidence and HbA1c: results from the KORA cohort. *Diabetologia.* 2022 Aug;65(8):1328-1338. doi: 10.1007/s00125-022-05704-7.
- 456.** 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. doi: 10.1038/nrm.2016.88.
- 457.** 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.
- 458.** 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.
- 459.** Umbach JL, Kramer MF, Jurak I, Karnowski HW, Coen DM, Cullen BR. MicroRNAs expressed by herpes simplex virus 1 during latent infection regulate viral mRNAs. *Nature.* 2008 Aug 7;454(7205):780-3. doi: 10.1038/nature07103. Epub 2008 Jul 2.
- 460.** Kärre K. Clever, cleverer, cleverest. *Nat Immunol.* 2002 Jun;3(6):505-6. doi: 10.1038/ni0602-505.

12 PUBLIKÁCIÓS LISTA

12.1 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 α 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**[✉]: 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**[✉]: 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**[✉]: 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***, 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.

12.2 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.

12.3 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: 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, Holubek 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**[✉], 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.

12.3.1 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.

12.4 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[✉]**: Ultra-deep, hybrid metagenomic sequencing of canine gut microbiome. *Scientific Data* 2024. – bírálólat 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. – bírálólat alatt, 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-TiO₂-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.

12.4.1 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, 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. 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.

13 TUDOMÁNYMETRIAI ADATOK

A 2024. május 13-i, az MTMT adminisztrátora által ellenőrzött állapot szerint

- Összegzett impakt faktor (IF): 324,144
 - első/utolsó szerzős IF: 182,366
- Összes idézetek száma: 1409
 - független idézetek száma: 818

MTMT közlemény és idéző összefoglaló táblázat				
Tombácz Dóra adatai (2024.05.22)				
Közlemény típusok	Szám		Hivatkozások 1	
	Összes	Részletezve	Független	Összes
Tudományos közlemények				
I. Tudományos folyóiratcikk	<u>69</u>	---	---	---
külföldi kiadású szakfolyóiratban idegen nyelven	---	<u>69</u>	<u>665</u>	<u>1200</u>
külföldi kiadású szakfolyóiratban magyar nyelven	---	0	0	0
hazai kiadású szakfolyóiratban idegen nyelven	---	0	0	0
hazai kiadású szakfolyóiratban magyar nyelven	---	0	0	0
II. Könyvek	<u>1</u>	---	---	---
a) Könyv, szerzőként	<u>1</u>	---	---	---
idegen nyelvű	---	<u>1</u>	0	0
magyar nyelvű	---	0	0	0
b) Könyv, szerkesztőként2	0	---	---	---
idegen nyelvű	---	0	---	---
magyar nyelvű	---	0	---	---
III. Könyvrészlet	0	---	---	---
idegen nyelvű	---	0	0	0
magyar nyelvű	---	0	0	0
IV. Konferenciaközlemény folyóiratban vagy konferenciakötetben	<u>1</u>	---	---	---
idegen nyelvű	---	<u>1</u>	<u>18</u>	<u>21</u>
magyar nyelvű	---	0	0	0
Közlemények összesen (I.-IV.)	<u>71</u>	---	<u>683</u>	<u>1221</u>
Absztrakt3	<u>19</u>	---	0	0
Kutatási adat	0	---	0	0
További tudományos művek4	<u>10</u>	---	<u>2</u>	<u>4</u>
Összes tudományos közlemény	<u>100</u>	---	<u>685</u>	<u>1225</u>
Hirsch index5	<u>21</u>	---	---	---
Oktatási művek	0	---	---	---
Felsőoktatási művek	0	---	---	---
Felsőoktatási tankönyv idegen nyelvű	---	0	0	0
Felsőoktatási tankönyv magyar nyelvű	---	0	0	0
Felsőoktatási tankönyv része idegen nyelven	---	0	0	0
Felsőoktatási tankönyv része magyar nyelven	---	0	0	0
Oktatási anyag	0	---	0	0
Óltalmi formák	0	---	0	0
Alkotás	0	---	0	0
Ismeretterjesztő művek	0	---	---	---
Folyóiratcikk	---	0	0	0
Könyvek	---	0	0	0
További ismeretterjesztő művek	---	0	0	0
Közérdekű vagy nem besorolt művek6	0	---	0	0
További közlemények7	0	---	0	0
Egyéb szerzőség8	0	---	0	0
Idézők szerkesztett művekre	---	---	0	0
Idézők disszertációban, egyéb típusban	---	---	<u>135</u>	<u>186</u>
Összes közlemény és összes idézőik	<u>100</u>	---	<u>820</u>	<u>1411</u>

14 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 atmoszféráért. 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 célom '*biológus leszek*' megvalósításában, ill. Totit és TBGt. Rengeteget segítettek az elmúlt évtizedekben. Köszönöm **A BARÁT**omnak a hosszú, évtizedes támogatását az élet minden kalandjában és kihívásában. Az ő bennem való hite is kellett ahhoz, hogy most itt álljak.

15 TÁMOGATÓK

Az értekezés alapját képező kutatásainkat, publikációink megjelenését és a munkámat támogatta

- 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
 - Magyari 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 Regulation [5P50HG00773502]; *Michael Snyder*)
- a Stanford Egyetem (Stanford University)
- a Weill Cornell Medicine