

Tézisfüzet

**A Varicella-Zoster Vírus és a Vaccinia Vírus Transzkriptom  
analízise hosszú-leolvasást adó szekvenálással**

Prazsák István M.Sc.



Szegedi Tudományegyetem

Általános Orvosi Kar

Orvosi Biológia Intézet

Interdiszciplináris Doktori Iskola

Témavezetők: Dr Tombáczi Dóra és Prof. Dr. Boldogkői Zsolt

Szeged

2021

## A tézishez közvetlenül kapcsolódó publikációk listája

Tombác, Dóra; Prazsák, István; Maróti, Zoltán; Moldován, Norbert; Csabai, Zsolt; Balázs, Zsolt; Dénes, Béla; Kalmár, Tibor; Snyder, Michael; Boldogkői, Zsolt: Long-read Assay Sheds New Light on the Transcriptome Complexity of a Viral Pathogen and on Virus-Host Interaction– SCIENTIFIC REPORTS 10: 13822 (2020)

IF: 3,998 Q1 (2019)

Boldogkői, Zsolt; Balázs, Zsolt; Moldován, Norbert; Prazsák, István; Tombác, Dóra: Novel Classes of Replication-associated Transcripts Discovered in Viruses RNA BIOLOGY 16: 2 pp. 166-175., 10 p. (2019)

IF: 5,216 Q1 (2018)

Tombác, Dóra, István Prazsák, Attila Szűcs, B Dénes, Michael Snyder, and Zsolt Boldogkői. 2018. “Dynamic Transcriptome Profiling Dataset of Vaccinia Virus Obtained from Long-Read Sequencing Techniques.” GIGASCIENCE.

IF: 7,267 D1 (2017)

Tombác, Dóra, István Prazsák, Norbert Moldován, Attila Szűcs, and Zsolt Boldogkői. 2018. “Lytic Transcriptome Dataset of Varicella-Zoster Virus Generated by Long-Read Sequencing.” FRONTIERS IN GENETICS 9.

IF: 4,151 Q1 (2017)

Prazsák, István; Moldován, Norbert; Balázs, Zsolt; Tombác, Dóra; Megyeri, Klára; Szűcs, Attila; Csabai, Zsolt; Boldogkői, Zsolt Long-read sequencing uncovers a complex transcriptome topology in Varicella-zoster virus BMC GENOMICS 19 Paper: 873 (2018)

IF: 3.73 Q1 (2017)

Prazsák, I, D Tombacz, A Szucs, B Denes, M Snyder, and Z Boldogkoi. 2018. “Full Genome Sequence of the Western Reserve Strain of Vaccinia Virus Determined by Third-Generation Sequencing.” GENOME ANNOUNCEMENTS 6 (11).

új név: Microbiology Resource Announcements: IF:0 Q3

## Rövidítések

|  |  |
|--|--|
| cDNA: copy DNA   | ONT: Oxford Nanopore Technologies (Ltd)          |
| CTO: close to replication origin   | ORF: open reading frame                          |
| CV-1: African green monkey ( <i>Cercopithecus aethiops</i> ) immortalized kidney cell line | Ori: origin of replication                       |
| cxRNA: complex RNA   | OriS: origin of replication at US genomic region |
| lncRNA: long non-coding RNA  | p.i.: post infection (timepoint)                 |
| LoRTIA: Long-read RNA-Seq Transcript Isoform Annotator (toolkit)                           | ppc: possible protein coding                     |
| LRS: long read sequencing  | PacBio: Pacific Biosciences (Ltd)                |
| MOI: multiplicity of infection   | PAS: polyA signal                                |
| MRC-5: human ( <i>Homo sapiens</i> ) immortalized lung fibroblast cell line                | ppc: possible protein coding                     |
| ncRNA: non-coding RNA  | ROI: Reads Of Inserts                            |
| NGS: next generation of sequencing   | SRS: short read sequencing                       |
| nroRNA: near to replication origin RNA   | TES:transcription end site                       |
| NTO: near the transcription origin   | TIS: translation initiation site                 |
|  | TSS: transcription start site                    |
|  | VACV: Vaccinia virus                             |
|  | VZV: Varicella zoster virus                      |
|  | WR: Western Reserve (strain of VACV)             |

## Bevezető

A vírusok életciklusának és a gazdasejtre gyakorolt hatásuknak megismerésében alapvető fontosságú megismerni, hogyan expresszálódik a virális genom. A jelenleg elterjedt szekvenáló módszerek számos limitációja akadályozza egyes transzkript izoformák kimutatását. Ezen problémát hosszú leolvasást (readet) adó szekvenálási módszerrel (LRS) küszöbölhetjük ki. A rövid leolvasást adó szekvenálási technikák (SRS), mint amilyen az Illumina (Goodwin, McPherson, and McCombie 2016), a BGI (Jeon et al. 2019), vagy az IonTorrent (Rothberg et al. 2011), tipikusan 75 és 300 bp hosszú leolvasást eredményeznek. Ami azt jelenti, hogy komplett virális mRNS-eket nem lehet egyetlen leolvasással vizsgálni, ezáltal az átfedő, bonyolult szerkezetű transzkriptek rejtve maradnak szemünk elől. Ezzel ellentétben LRS technikával rutinszerűen lehet 10kbp-nál is hosszabb leolvasásokat kapni (Pollard et al. 2018). Az SRS technika egyik nagy hátránya, hogy közvetlen RNS szekvenciát nem lehet vele leolvasni, így az RNS-módosulásokat sem lehet közvetlenül vizsgálni (Harel et al. 2019). A hosszú leolvasást adó szekvenálásoknak két elterjedt fajtája érhető el: a Pacific Biosciences (PacBio) valós idejű egy molekula szekvenálása (SMRT) és az Oxford Nanopore Technologies (ONT) nanopórusos szekvenálási technikája.

Az emberi genom és a főbb modell organizmusok transzkriptom annotációja ugyan manapság már elérhető, de vírusok esetében a transzkriptom annotáció általában leegyszerűsített, ezért az alternatív transzkript izoformák, az átfedő transzkripciós egységek esetében a génexpresszió mértékének becslése nem pontos, vagy éppenséggel lehetetlen (Depledge, Mohr, and Wilson 2018).

A dsDNS genommal rendelkező vírusokhoz tartozik a herpeszvírusok családja, ami három további aloszaládra osztható, úgymint  $\alpha$ - (pl. HSV-1, HSV-2, és VZV),  $\beta$ - (CMV, HHV6, és HHV7) és  $\gamma$ -herpeszvírusok (EBV és KSHV), amelyek különböznek a fertőzés helye és lefolyása alapján (De Pelsmaeker et al. 2018). A herpeszvírusok az egyik

legsikeresebb humán patogén kórokozók közé tartoznak, mivel az emberi populáció csaknem egésze fertőzött lehet legalább egyfajta herpeszvírussal.

A herpeszvírusok három nagy alcsaládjának megfelelő evolúciós stratégiát fejlesztett ki a VZV, azaz jellemző rá a látencia, ami az elsődleges fertőzés után a dorzális ganglionokban kialakuló élethosszig tartó nyugvó állapota a vírusnak(Levin et al. 2003). A VZV reaktivációjára hosszú periódus után kerülhet sor, ami a dermatómáknak megfelelő területeken fájdalmas kiütéssel járó övsömörként manifesztálódik (=herpes zoster)(Lungu et al. 1995).

A VZV 125 kbp hosszú, lineáris dsDNS genommal rendelkezik, amely egy hosszú (UL) és rövid (US) régióra osztható, amelyek fordított ismétlődésekkel határoltak(Davison and Scott 1986). A 71 annotált génből három, az orf62, orf63 és orf64 kétszer van meg a genomban a repeat régiók miatt. A VZV lítikus génexpressziós vizsgálatát microarray és SRS technikákkal már elvégezték(Baird et al. 2014; Hoover et al. 2006). Kis nem kódoló RNS-ek replikációban való szerepét szintén SRS technikával bizonyították (Bisht et al. 2020; Markus et al. 2017). A herpeszvírusok látenciája és látens állapotból történő reaktivációja különös érdeklődésre tart számot. Széles körben elfogadott az a nézet, hogy látencia asszociált transzkriptek szabályozzák a herpeszvírusok látens és lítikus ciklusa közötti váltást. VZV esetén ez a transzkript csak nemrégén vált ismertté, mivel SRS segítségével találták meg az orf61-el antiszensz pozíciójú többszörösen splice-olt RNS-t, amit emberi trigeminális neuronokból izoláltak (Depledge, Ouwendijk, et al. 2018).

A Vaccinia vírus (VACV) a Poxvírusok családjába tartozó vírus, rokona az emberi történelem egyik leghalálosabb vírusának tartott fekete himlőt okozó variola vírusnak. A VACV-val végzett aktív immunizálás a fekete himlő sikeres eradikálásához vezetett világszerte, azóta a VACV-t vírusvektorként használják rekombináns vakcinák kifejlesztésére génterápiában, rák immunterápiában és onkolitikus

vírus terápiában. A herpeszvírusokkal ellentétben a VACV a sejtplazmában, ún. citoplazmikus vírus gyárakban replikálódik, nagy genomja pedig alkalmas idegen gének befogadására (Chard et al. 2015; Dénes et al. 2014; Hung et al. 2007).

A VACV Western Reserve törzse (VACV-WR) lineáris, kétszálú DNS genomja 195kbp hosszú és fordított terminális ismétlődéseket tartalmaz (ITR), kovalensen zárt végekkel. A vírus 215 nyitott leolvasási keretet (ORF) hordoz. A korai VACV transzkriptom kutatások Northern-blot (Baldick and Moss 1993; Wittek and Moss 1982) és *in vitro* génexpressziós vizsgálatok voltak (Ahn et al. 1992; Ahn, Jones, and Moss 1990; Amegadzie, Ahn, and Moss 1991; Bajszár et al. 1983; Venkatesan, Gershowitz, and Moss 1982) ezúton egyedi gének expressziós mintázatát tudták megállapítani (Cooper and Moss 1979), míg a később alkalmazott microarray vizsgálatok már alkalmasak voltak a legtöbb ORF időbeli változását és a gazdasejt vírusfertőzésre adott transzkripciós válaszát is vizsgálni (Rubins et al. 2008). SOLiD és Illumina rövid read szekvenáló platformokon a teljes VACV genomra kiterjedő vizsgálatokkal határozták meg a korai RNS-ek 5' és 3' végeit (Yang et al. 2010, 2011). A VACV kései RNS-einek transzkripciós iniciáció (TSS) és termináció (TES) helyei gyakori aktivitást mutattak Yang és munkatársai kísérletei alapján (Yang et al. 2012), ami arra utal, hogy széleskörű transzkripciós túlírás jellemzi a VACV transzkripteket, különösen a fertőzés kései szakaszában (Yang et al. 2010). SRS technika alkalmazásával sikerült különböző promotóereket azonosítani, amelyek a korai és késői gének jellemző cisz-szabályozói, valamint riboszóma-profiling technikával számtalan transzkript translációs kezdő helyét is meghatározták (Yang et al. 2015). A VZV és a VACV transzkriptom szerkezetét cap-szelektált és nem cap-szelektált szekvenálási könyvtárak elemzésével vizsgáltam.

Számos, korábban le nem írt transzkript izoformát, rövid, beágyazott transzkripteket, valamint hosszú, több ORF-et átfedő transzkriptet és VZV esetén új splice variánsokat találtam, ezen kívül korábbi vizsgálatokból származó transzkript koordinátákat erősítettem meg.

## **Kutatási célok**

A VZV és VACV lítikus fertőzési stádiumában vizsgáltam a virális transzkriptomot. Az transzkriptom annotálás érdekében a következő célokat tűztük ki:

- a virális transzkriptok 5' és 3' végeinek bázispár pontossággal való meghatározása mind VZV, mind VACV esetén;
- a megkapott 5' és 3' végek összekötése transzkriptokká, adott génekhez tartozó izoformák meghatározása;
- mivel a nanopórusos szekvenálási technika lehetővé teszi teljes hosszúságú mRNS molekulák leírását, az átfedő, komplex struktúrájú intronok feltérképezésére is lehetőség nyílik, ezért új splice helyek meghatározása is célom volt;
- a nanopórusos szekvenálási technika RNS módosulások meghatározására is alkalmas, ezért kihasználtam ezt a lehetőséget, hogy RNS-módosításokat keressek a megkapott adatok között.

## Módszerek

### VZV szaporítás, RNS tisztítás és cDNS szekvenálás

Attenuált Varicella zoster vírus Oka törzset (Merck, Varivax) humán primer embrionális tüdő fibroblaszt sejtenyészeten (MRC5) tenyésztettünk 37°C-on 5 napon át, amíg a citopátiás hatás csaknem 100%-os nem lett. Teljes RNS-t izoláltunk ezután a Macherey-Nagel RNS izoláló kittel. A cDNS-könyvtár készítése az Oxford Nanopore Technologies Ligation Sequencing kit (ONT SQK-LSK108) alapján történt az 1D strand switching cDNS ligálási protokollnak megfelelően, majd a polyA(+) RNS frakciót reverz transzkriptáz és oligo(dT) primer segítségével írtuk át. Az 5'-végek meghatározására a fent említett ONT 1D strand switching cDNS ligálási protokollt kombináltuk az 5'-cap specifikus protokollal. A cDNS mintát teljes RNS-ből állítottuk elő TeloPrime teljes hosszúságú cDNS Amplification Kit (Lexogen) segítségével; Az RT-reakciót oligo(dT)-tartalmú primer alkalmazásával hajtottuk végre. Adaptereket ligáltunk az RNS-ek 5' végére duplaszál-specifikus ligázzal (Lexogen). A második cDNS szál írása a TeloPrime Kit Second-Strand Mix és Enzyme Mix segítségével történt a leírásnak megfelelően. Ezután a minták a TeloPrime Kit-nek megfelelően lettek amplifikálva. Az amplifikált PCR termékek end-repair után újabb adapterrel lettek ligálva, majd mágneses gyöngyön tisztítottuk őket (Agencourt AMPure XP, Beckman Coulter). Ezután a mintákat R9.4 SpotON Flow Cell-en szekvenáltuk, a leolvasott nyers adatok basecalling-ja pedig Albacore v2.1.10 szoftverrel történt. A polyA(+) és Cap-szekvenálás mellett célzott szekvenálást is végrehajtottunk a replikációs origó környéki RNS-ek validálása érdekében.

Az RNS hiper-szerkesztés másodlagos RNS struktúrára gyakorolt hatását „RNAstructure” szoftver segítségével végeztük (Reuter and Mathews 2010). Az inozin jelenlétét az adenozinok „*in silico*” guaninra cserélésével szimuláltuk, a readeket Geneious (Kearse et al. 2012) szoftverrel IGV-ben jelenítettük meg.



## **VACV szaporítás, RNS tisztítás**

Afrikai zöld majom vese fibroblaszt sejtek (CV-1) tenyészetét VACV-al fertőztük (10 MOI/sejt), majd 37°C-on inkubáltuk 1,2,3,4,6,8,12 és 16 órán át 5% CO<sub>2</sub> termosztátban. A fertőzést megállítottuk a megadott időpontokban és teljes RNS-t izoláltunk a Macherey-Nagel RNA Kit gyártói leírásának megfelelően a minta egy részéből. A polyA(+) mRNS-eket Oligotex mRNA Mini Kit (Qiagen) segítségével tisztítottuk ki a totál RNS mintából. A minták másik részéből RiboZero Magnetic Kit (Illumina) segítségével eltávolítottuk a rRNS-t a nem-polyA RNS-ek vizsgálata céljából.

## **VACV cDNS szekvenálása PacBio és ONT szekvenáló platformon**

A polyA(+) RNS frakciókból cDNS-t írtunk a SMARTer PCR cDNA Synthesis Kittel (Clontech) a PacBio 'Isoform Sequencing (Iso-Seq) és 'No Size Selection' protokoll leírásnak megfelelően. Ezek a minták az 1, 4, 8 és 12 p.i. időpontokból származtak és az RSII cDNS könyvtárkészítéshez össze lettek keverve, ill. az 1, 2, 3, 4, 6, és 8 p.i. minták egyesével lettek a Sequel platformon szekvenálva. Továbbá az rRNS-depletált RNS-minták keverékének (1, 4, 8 és 12 órás időpontokból) cDNS írása a módosított random hexamer primerekkel készült úgy, hogy ezt a primert használtuk a SMARTer Kit's oligo(d)T primere helyett. A SMRTbell mintákat előkészítettük a PacBio DNA Template Prep Kit 1.0 leírásának megfelelően és a PacBio MagBead Kit (v2)-tel tisztítottuk. A polyA(+) RNS frakciót használtuk a z ONT MinION platformon történő cDNS szekvenálásához. A különböző időpontokból származó RNS minták cDNS átírását az ONT 1D SQK-LSK108 protokoll szerint végeztük. A mintákat egyesével és keverten is megszekvenáltuk, a szekvenálási könyvtárak az ONT 1D ligációs kit és protokoll (SQK-LSK108) szerint lettek készítve a gyártó leírásának megfelelően.

## **Adatelemzés és illesztés**

A bioinformatikai számítások Ubuntu/Linux operációs rendszeren történtek python3 alatt. A cDNS szekvenálásból származó adatok egyrészt a VZV referencia genomra ([NC 001348.1](#)) és a gazdasejt genomra (*Homo sapiens* - GRCh37; [PRJNA31257](#)) lettek térképezve GMAP v2017-04-24(Wu and Watanabe 2005) programmal. A VACV ONT szekvenálási nyers adatok, valamint a VACV PacBio szekvenálásból származó ROI egyrészt a VACV általunk összeillesztett genomjára ([LT966077.1](#)) és a gazdasejt genomjára (*Chlorocebus sabaeus*; [GCA 000409795.2](#)) lettek illesztve minimap2 szoftverrel. Az 5' és 3' végek meghatározásához egy új algoritmust fejlesztett csoportunk, ami a később megjelenő transzkript annotáló programcsomag (LoRTIA: Long-read RNA-Seq Transcript Isoform Annotator) kifejlesztésének alapját adta. Az algoritmus és a számítás részletes jellemzése megtalálható a Prazsák et al. (2018)(Prazsák et al. 2018) és a <https://github.com/zsolt-balazs/LoRTIA> alatt.

## **VZV transzkriptek annotációja és splice helyek meghatározása**

A mindkét végén polyA-t, vagy nem megfelelő szekevnáló adaptereket tartalmazó readeket nem vettük figyelembe, kivéve a komplex transzkriptek esetében, amelyeket egyesével ellenőriztünk le IGV-ben. Az 5' vagy 3' végek közötti 10 nt-nál nagyobb eltérések esetében tekintettük új hossz-izofornának a transzkripteket a következő jelölések mellett: L: *hosszabb* 5' UTR, S: *rövidebb* 5' UTR, AT: *alternatív 3' termináció*). Az ún. .5 variánsok azon rövid hossz-izofornák voltak, amelyek egy ismert ORF-en belül *in frame* helyezkedtek el, ezeket potenciális protein kódoló transzkriptekként írtuk el (ppc). Amennyiben több ppc volt jelen a leghosszabb 5' UTR-el rendelkező variáns lett .5 izofornának elnevezve, és a többi csökkenő sorrendben. Nem kódoló transzkripteknek azokat neveztük,

amelyek TES-e ORF-en belül végződött (így nem tartalmaztak STOP-kodont), vagy a TSS-e nem *in-frame* ORF-et eredményezett. Azokat a multigén transzkripteknek, amelyek legalább két gént tartalmaztak, és ezek ellentétes irányúak, komplex transzkriptnek (cxRNS) neveztük. Ha több TSS is előfordult az ORF előtt egy ilyen transzkriptnél, akkor az upstream legközelebbi TSS-t annotáltuk. A transzkriptek relatív gyakoriságát a térképezett összreadszámhoz viszonyítva adtuk meg.  $\pm 10$  nt lötyögést engedtünk meg a TSS-ek és TES-ek esetén egy adott transzkript izoforma végeinek annotálásánál. BLASTP algoritmus segítségével állapítottuk meg az új transzkript izoformák ORF-je által lehetségesen kódolt fehérjék homológiáját (a beállított küszöbérték 10 volt). A splice helyeket akkor fogadtuk el, ha az intron határokon a GT-AG konszenzus szekvenciák megvoltak legalább 10 read esetén, és az intron előfordulási gyakorisága legalább 1%-os volt.

## **RNS hiperszerkesztés vizsgálata**

Az RNS hiperszerkesztés másodlagos struktúrára gyakorolt hatását az 'RNAstructure Software'(Reuter and Mathews 2010) csomaggal elemeztük. Az inozin jelenlétét az adeninek guaninra cserélésével szimuláltuk, és a readeket a Geneious(Kearse et al. 2012) és IGV programmal vizualizáltuk.

## **A VZV transzkriptom analízis eredményei**

A cap-szelektált mintában 614192, a polyA-szelektált mintában 66455 és a targetált mintában 328118 readet kaptunk. A gazdasejt és a vírus össz-readszáma 10338565 volt. A nem-cap-szelektált minta esetén kaptuk a leghosszabb átlag read-hosszat, ami 1349 bp volt. Összesen 1124 5'-véget és 255 3'-véget detektáltunk ebben a mintában és 1428 5'-véget és 279 3'-véget a cap-szelektált mintában. A readek végeinek genomra vetített pozíciójának vizsgálatát a LoRTIA transzkript

annotálásra kifejlesztett programunkkal határoztuk meg, megkapva így a transzkriptek TSS és TES pozícióját. Az annotálás után az összes detektált 5' vég 10,86%-a bizonyult TSS-nek, míg a 3' végek 32,95%-a TES-nek. További 49 db 5', illetve 16 db 3' véget kizártunk a cap-szelektált és 21 db 5' és 16 3' véget a nem-cap-szelektált adatok közül, mert felmerült a *false-priming*-nak nevezett jelenség gyanúja. A szakirodalomból korábban detektált TSS-ek közül mindet, azaz 18-at, míg a TES-ek közül 9-et igazoltunk vissza. Az ONT MinION cDNS-szekvenálás cap-és nem-cap-szelektált adatsoraiban közös TSS-ek száma 124 a TES-ek száma 71 volt.

## Új kódoló és nem-kódoló transzkriptek, új splice helyek

A génekbe ágyazott, 5' csonkolt transzkriptek azonosítása viszonylag egyszerű LRS-technikával. Vizsgálataink során 25 ilyen transzkriptet találtunk, amelyek közül 18 kanonikus PAS-t is tartalmaztak, és potenciálisan fehérjét is kódolnak. Összesen 23 új nem-kódoló RNS-t találtunk, amelyek közül öt ún. kis, nem-kódoló RNS (200bp-nál kisebb) volt, míg a többi hosszú, nem-kódoló RNS volt (200bp-nál hosszabbak definíció szerint). Egy intergenikus és hét antiszensz ncRNS-t is kimutattunk.

VZV-ben eddig öt gén esetén volt ismert splicing. Megerősítettük kilenc korábban már leírt splice hely pozícióját és további tizenkét splice helyet azonosítottunk, amelyek splice-donor és -akceptor konszenzus szekveniája GT ill. AG volt. Ezen kívül az ORF42/45 gén által kódolt két új splice variációt találtunk, valamint az ORF63 gén transzkriptjének további 10 új (!) változatát detektáltuk, amelyek a VLTly transzkripttel átfedőnek bizonyultak. *In silico* módszerekkel 44 potenciális uORF-et észleltünk 81 VZV-transzkript 5'-UTR-jében.

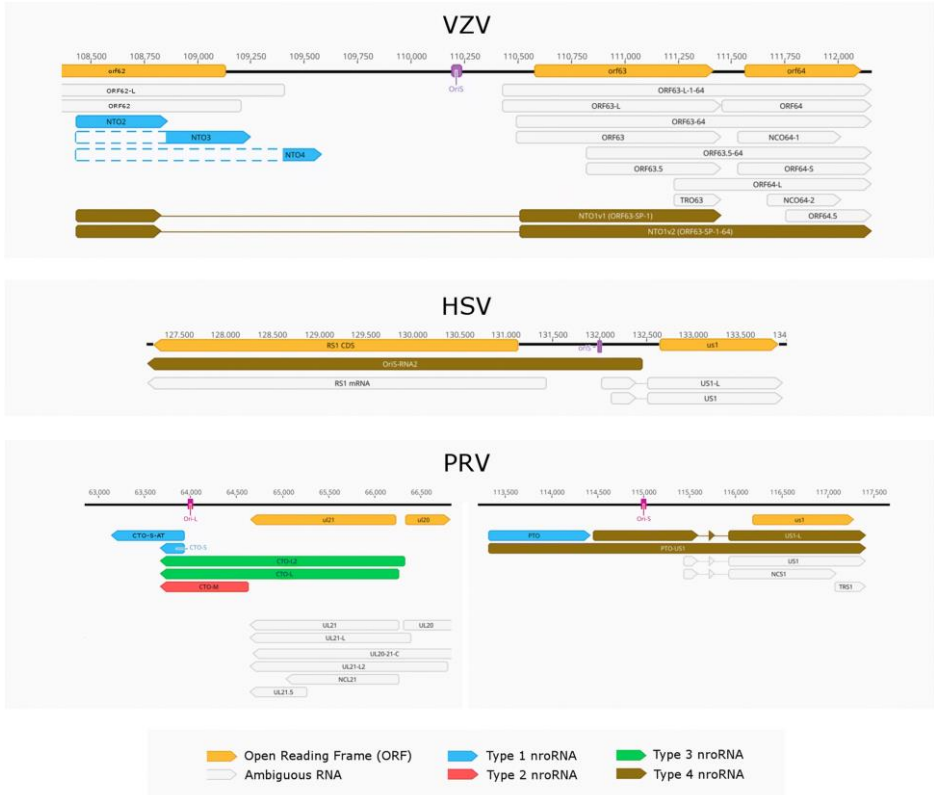
## nroRNS-ek, komplex transzkriptek, és RNS-szerkesztés

Kutatócsoportunk korábban már kimutatta, hogy herpeszvírusok a replikációs origó környékén transzkripteket expresszálnak, amelyeket nroRNS-eknek nevezhetünk - *near the replication origin RNA* (Boldogkői et al. 2019; Tombáczi et al. 2015). A VZV OriS régiója környékén 9 nroRNS-t azonosítottunk, mint pl. amilyen az NTO1v1, NTO1v3 és NTO1v5, amelyek az ORF63 transzkript hosszú, fúziós TSS variánsai. Az nroRNS-ek más herpeszvírusokban is megtalálhatók, ezekre példát az 1. ábra mutat be.

33 eddig még ismeretlen, hosszú, policisztronos RNS-t, és ezek között négy complex transzkriptet találtunk a VZV transzkriptomban. Továbbá alacsony abundanciájú complex transzkripteket találtunk a VLT régiójában, amelyek közül hét egészen az ORF63 és ORF64 RNS-ekkel koterminálisak voltak. Ezek az RNS-ek számos itt található kódoló szekvenciával ellentétes irányban átfedőek és a VLTly izoformáinak megfelelően, vagy ahhoz hasonló mintázatot követve splice-olódnak.

A replikációs origó környéki transzkriptek egyike - az általunk leírt NTO3 - nagy gyakorisággal mutatott A-G szubsztitúciót, ami érdekes módon nem volt jelen az ezzel a transzkriptekkel átfedő más readeken. Az ebben a régióban lévő szubsztitúciók 58%-a A-G volt, ami szignifikánsan nagyobb az átfedő readeken talált 12,98%-os szubsztitúciós rátával ( $p < 0.0001$ , Fischer-próba), így az itt található összes adenin 22,07%-a RNS-szerkesztettnek tűnt. Ezt a feltételezésünket *in silico* másodlagos szerkezet analízissel erősítettük meg, ami alapján elmondható, hogy a nem szerkesztett RNS izoforma szabadenergia állapota magasabb, mint az editálté (-143.2 kcal/mol vs -169.4 kcal/mol), ami arra utal, hogy a hyper-editing jelensége termodinamikai stabilitást kölcsönözhet az NTO3 transzkript másodlagos szerkezetének. Ezen kívül az editált pozíciók mellett

található konszenzus szekvenciák megléte is megerősíti azt a feltételezésünket, hogy az NTO3 nem-kódoló RNS poszt-transzkripcionális hiperszerkesztés szubsztrátja lehet.



1. *Ábra. Három alfa-herpesz vírus replikációs origó környéki RNS-típusai (nroRNS). Az nroRNS-eket négy különböző típusba sorolhatjuk a replikációs origó és a kódolási kapacitás szerinti helyzetük szerint. Az 1. típusú nroRNS-ek nem kódoló RNS-ek (ncRNS), amelyek nem fedik át az Ori-t (kék nyíl). A 2. típusú nroRNS-ek szintén ncRNS-ek, amelyek átfedik az Ori-t (piros nyíl). A 3. típusú nroRNS-ek alternatív TES-sel és hosszú 3'UTR-el rendelkező, Ori-t átfedő transzkriptek (zöld nyíl), míg a 4. típusú nroRNS-ek alternatív TSS-sel rendelkező mRNS-ek, amik az Ori-t átfedő hosszú 5'UTR-el rendelkeznek (barna nyíl).*

## A VACV transzkriptom elemzés eredményei

A nagy read-szám és magas lefedettség miatt a kapott cDNS-ekből rekonstruálni tudtuk a VACV genomot. Ehhez először a vírus readeket a VACV ún. Western Reserve (WR) törzsének referencia genomjához illesztettük (GenBank azonosító [NC\\_006998](#)). Az átlagos, egy nukleotidra eső read-szám 289 ill. 43 volt a PacBio és az ONT szekvenálások esetén. A referencia genom minden egyes pozícióján tapasztaltunk lefedettséget, különösen a késői időpontokban, mivel 8h infekciós idő után kb. 80%, 12h után a teljes VACV genom transzkripcionálisan aktív volt. A genom rekonstrukcióhoz a PacBio szekvenálási adatokat használtuk fel, az így kapott VACV WR törzs genomi szekvenciája 194888 bp hosszúnak adódott, az átlagos GC-tartalom pedig 33,3%-nak. Összehasonlítva az általunk kapott genomot a referencia genommal, összesen 163 pontmutációt találtunk, amelyek nagy része a genom F régiójában halmozódott. A szekvenciát a LT966077 azonosító alatt elérni a Génab

A VACV transzkriptom vizsgálatához a PacBio szekvenálás Iso-Seq módszerét használtuk PacBio RSII és Sequel platformon; az ONT-tól pedig a MinION szekvenáló készüléket. Összesen 25 egymástól független, különböző szekvenálási könyvtárat használtunk fel a vizsgálatokhoz, amelyek kiindulási RNS-ei több fertőzési időpontból származtak (1,2,3,4,6,8,12 és 16 órával a fertőzés után). Összesen 1115000 ROI-t (reads of insert) kaptunk a PacBio szekvenálás során, míg 535000 vírus readet a nanopórus szekvenálással. Az átlag read hossz a PacBio szekvenálások esetén  $1129\text{bp} \pm 126,7\text{bp}$  (SD) és  $619\text{bp} \pm 148,2\text{bp}$  (SD) volt az ONT szekvenálások esetén. A leghosszabb illesztett readek (a 6000bp-nál hosszabbak) az ONT MinION 1D

cDNS, valamint a 8 órás PacBio Sequel szekvenálási könyvtárakból származtak.

## Új kódoló és nem kódoló transzkriptek

A „szabályos” genomi régiókban, ismert ORF-eket átfedő 135 monocisztronos kódoló transzkriptet találtunk a LoRTIA programmal, további 12 transzkriptet, amelyet a programcsomag nem detektált alacsony abundanciájuk miatt. A vizsgálatunkban összesen 292 különböző típusú policisztronos transzkript izoformáit találtuk: 43 bicisztronos, 137 tricisztronos, 92 tetracisztronos, 15 pentacisztronos és 5 hexacisztronos RNS-t. Új géneket is annotáltunk intergénikus régióban, valamint beágyazott géneket (5' csonkolt in-frame ORF-eket egy nagyobb ORF-en belül), továbbá rövid upstream ORF-eket (uORF, amelyek a fő ORF-et megelőzik)- ld. 2. Ábra. Két új, feltehetőleg kódoló gént találtunk új mRNS-ek annotálása közben; ezek új intergénikus pozícióban és a repeatben helyezkednek el (B25.5R és C19.5L), a C9.5L a C9L gén közelében. Vizsgálataink során 49 új beágyazott, potenciális gént azonosítottunk, amelyek 5'-végi csonkolt ORF-el rendelkeznek és rövidebbek, mint az *in silico* annotált kanonikus ORF. Csaknem az egész VACV genom mentén, és mindkét DNS-szálon transzkripciós aktivitást tudtunk kimutatni. Az mRNS-ek és az antiszensz RNS-ek (asRNS) expressziós szintje a genomi régióként változó volt, de általánosságban elmondható, hogy az asRNS-aktivitás relatíve alacsony volt, kivételt képezett ez alól az A18R gén asRNS-ei, amelyek nagy változatossággal több, mint 100 TES transzkript izoformát adtak.

A VACV gének többsége tandem orientációban helyezkedik el, csak néhány konvergens és divergensen elhelyezkedő génpár van jelen a VACV genomjában, ezért a komplex transzkriptek (cxRNS-ek) száma alacsony a herpeszvírusokban találhatókkal szemben (Tombácz et al.

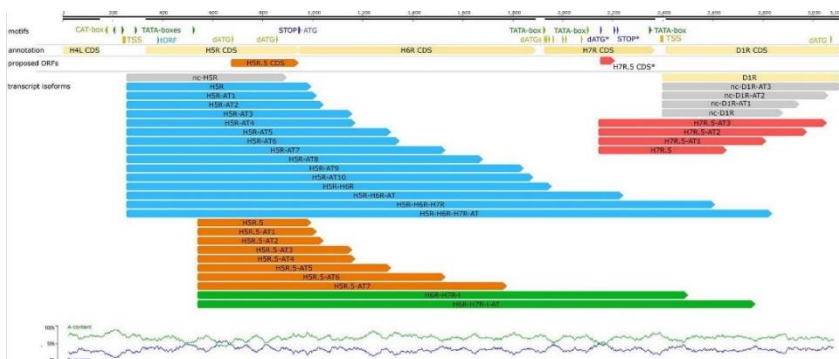


2016). Összesen 30 cxRNS molekulát találtunk, kilenc genom régióban.

## Átfedő transzkriptek

A VACV transzkriptommal foglalkozó korábbi tanulmányok több száz TSS-t és PAS-t azonosítottak már, de a transzkriptek 3'-végi pozícióját csak bizonytalanul tudták meghatározni, mivel a VACV transzkriptom bonyolult szerkezetű pl. a tandem orientációjú gének túlírása révén, ami a fertőzés késői időpontjában igen jellemző. Többféle szekvenálási megközelítést alkalmazva egy rendkívül komplex, sokféle átfedést tartalmazó transzkripciós hálózatot tártunk fel. Összesen 154 tandem, 32 divergens és 32 konvergens génpár található a VACV genomban;

a tandem gének közül 21 átfedő, a konvergens génpárok között 13 és a divergens génpárok között 5 transzkripttel. A szorosan egymás mellett elhelyezkedő tandem gének részleges és teljes transzkripciós átfedést is alkothatnak.



2. Ábra. A vizsgálatunk által azonosított, feltételezhetően fehérje kódoló beágyazott gének két példájának sematikus ábrázolása. Számos 5'-csonka transzkriptet lehet látni, amelyek rövid in-frame ORF-eket tartalmaznak (H5R.5 és H7R.5), az ábrán a H5R és H7R ORF régiók transzkriptjei mutatja.



## Diszkusszió és következtetések

A VZV és VACV transzkriptom vizsgálatát hosszú read szekvenálási technikákkal (ONT MinION és PacBio) hajtottunk végre annak érdekében, hogy átfogó, teljes genomra kiterjedő transzkript térképet készítsünk ezen vírusok transzkriptomáról lítikus fertőzés során. Capszelektált szekvenálási könyvtárat hoztunk létre a virális transzkriptek 5' és polyA-szelektáltat a transzkriptek 3' végeinek meghatározásához. A vizsgálataink által megállapíthatjuk, hogy mind a VZV, mind a VACV esetén a genom csaknem teljes egésze transzkripcionálisan aktív, főképp a kései virális életciklusban.

A hosszú leolvasást adó szekvenálási technikák nagy előnye, hogy complex transzkript struktúrákat is fel lehet velük tárni, ezért izoforma, TSS-, TES-variánsok, és intronok, ill. azok variációinak meghatározására kiválóan alkalmasak, hosszú, több gént átíró complex, vagy multigénes transzkriptek esetén, ill. beágyazott, átfedő transzkriptek felfedezésekor. A rövid read szekvenálási technikák nagy pontosságuk ellenére sem képesek visszaadni a hosszú és bonyolult genomi régiókban képződő teljes hosszúságú RNS-eket. Ezzel szemben a hosszú-read szekvenálási technikák megfelelő paraméterezés mellett hűen tükrözik a bonyolult, átfedő transzkripteket is. A VZV-kísérleteink során kimutatott transzkriptek részben már ismert transzkriptek izoformái vagy ismert gének új transzkriptjei voltak, valamint nem-kódoló transzkriptek és új, splice variánsok. VACV-kísérletek esetében a talált transzkriptek és izoformák pár kivételtől eltekintve mind újonnan annotált RNS-ek voltak. Mindkét vírus esetén elmondható, hogy kísérleteink nyomán igen nagy változatosságban találtunk új virális transzkripteket.

Bár a natív RNS-szekvenálás jelenleg még nem teljesen mentes minden problémától (pl. a transzkriptek 5' vége nem teljes, magasabb leolvasási hibaszázalék), mégis egészen új távlatokat nyújt a transzkriptomikai kutatások jövőjében, és potenciálisan a transzkriptom kutatások új standard módszere lehet, amellyel RNS-ek

mind mennyiségi, mind szerkezeti változásait követni lehet, ahogyan ezt már demonstrálták humán és vírus mintákon (Braspenning et al. 2020; Soneson et al. 2019). Összegzésként elmondhatjuk, hogy az alkalmazott cDNS szekvenálási megközelítésünk új ablakot nyitott a herpeszvírusok transzkriptom kutatása területén, mivel jelenleg csak az általunk is használt hosszú-read technológia nyújt információt a teljes genomra kiterjedő, ritka és nehezen detektálható transzkripció eseményekről bázispár pontossággal.

Ezt a munkát az ÚNKP-20-4 -SZTE 140 Innovációs és Technológiai Minisztérium Új Nemzeti Kiválósági Programja támogatta a Nemzeti Kutatási, Fejlesztési és Innovációs Alap forrásából.



## Hivatkozások

- Ahn, B. Y., E. V. Jones, and B. Moss. 1990. "Identification of the Vaccinia Virus Gene Encoding an 18-Kilodalton Subunit of RNA Polymerase and Demonstration of a 5' Poly(A) Leader on Its Early Transcript." *Journal of Virology*.
- Ahn, B. Y., J. Rosel, N. B. Cole, and B. Moss. 1992. "Identification and Expression of Rpo19, a Vaccinia Virus Gene Encoding a 19-Kilodalton DNA-Dependent RNA Polymerase Subunit." *Journal of Virology*.
- Amegadzie, B. Y., B. Y. Ahn, and B. Moss. 1991. "Identification, Sequence, and Expression of the Gene Encoding a M(r) 35,000 Subunit of the Vaccinia Virus DNA-Dependent RNA Polymerase." *Journal of Biological Chemistry*.
- Baird, N. L., J. L. Bowlin, R. J. Cohrs, D. Gilden, and K. L. Jones. 2014. "Comparison of Varicella-Zoster Virus RNA Sequences in Human Neurons and Fibroblasts." *Journal of Virology* 88(10):5877–80.
- Bajszár, G., R. Wittek, J. P. Weir, and B. Moss. 1983. "Vaccinia Virus Thymidine Kinase and Neighboring Genes: MRNAs and Polypeptides of Wild-Type Virus and Putative Nonsense Mutants." *Journal of Virology*.
- Baldick, C. J. and B. Moss. 1993. "Characterization and Temporal Regulation of MRNAs Encoded by Vaccinia Virus Intermediate-Stage Genes." *Journal of Virology*.
- Bisht, Punam, Biswajit Das, Paul R. Kinchington, and Ronald S. Goldstein. 2020. "Varicella-Zoster Virus (VZV) Small Noncoding RNAs Antisense to the VZV Latency-Encoded Transcript VLT Enhance Viral Replication." *Journal of Virology*.
- Boldogkői, Zsolt, Zsolt Balázs, Norbert Moldován, István Prazsák, and Dóra Tombácz. 2019. "Novel Classes of Replication-Associated Transcripts Discovered in Viruses." *RNA Biology*.
- Chard, Louisa S., Eleni Maniati, Pengju Wang, Zhongxian Zhang, Dongling Gao, Jiwei Wang, Fengyu Cao, Jahangir Ahmed, Margueritte El Khouri, Jonathan Hughes, Shengdian Wang, Xiaozhu Li, Bela Denes, Istvan Fodor, Thorsten Hagemann, Nicholas R. Lemoine, and Yaohe Wang. 2015. "A Vaccinia Virus Armed with Interleukin-10 Is a Promising Therapeutic Agent for Treatment of Murine Pancreatic Cancer."

- Cooper, Jonathan A. and Bernard Moss. 1979. "In Vitro Translation of Immediate Early, Early, and Late Classes of RNA from Vaccinia Virus-Infected Cells." *Virology*.
- Davison, A. J. and J. E. Scott. 1986. "The Complete DNA Sequence of Varicella-Zoster Virus." *Journal of General Virology* 67(9):1759–1816.
- Dénes, Béla, Nadja Fodor, Andre Obenaus, and István Fodor. 2014. "Engineering Oncolytic Vaccinia Viruses for Non-Invasive Optical Imaging of Tumors." *The Open Biotechnology Journal*.
- Depledge, Daniel P., Ian Mohr, and Angus C. Wilson. 2018. "Going the Distance: Optimizing RNA-Seq Strategies for Transcriptomic Analysis of Complex Viral Genomes." *Journal of Virology*.
- Depledge, Daniel P., Werner J. D. Ouwendijk, Tomohiko Sadaoka, Shirley E. Braspenning, Yasuko Mori, Randall J. Cohrs, Georges M. G. M. Verjans, and Judith Breuer. 2018. "A Spliced Latency-Associated VZV Transcript Maps Antisense to the Viral Transactivator Gene 61." *Nature Communications* 9(1):1167.
- Goodwin, Sara, John D. McPherson, and W. Richard McCombie. 2016. "Coming of Age: Ten Years of next-Generation Sequencing Technologies." *Nature Reviews. Genetics* 17(6):333–51.
- Harel, Noam, Moran Meir, Uri Gophna, and Adi Stern. 2019. "Direct Sequencing of RNA with MinION Nanopore: Detecting Mutations Based on Associations." *Nucleic Acids Research*.
- Hoover, Susan E., Randall J. Cohrs, Zoila G. Rangel, Donald H. Gilden, Peter Munson, and Jeffrey I. Cohen. 2006. "Downregulation of Varicella-Zoster Virus (VZV) Immediate-Early ORF62 Transcription by VZV ORF63 Correlates with Virus Replication in Vitro and with Latency." *Journal of Virology* 80(7):3459–68.
- Hung, C. F., Y. C. Tsai, L. He, G. Coukos, I. Fodor, L. Qin, H. Levitsky, and T. C. Wu. 2007. "Vaccinia Virus Preferentially Infects and Controls Human and Murine Ovarian Tumors in Mice." *Gene Therapy*.
- Jeon, Sol A., Jong Lyul Park, Jong Hwan Kim, Jeong Hwan Kim, Yong Sung Kim, Jin Cheon Kim, and Seon Young Kim. 2019. "Comparison of the MGISEQ-2000 and Illumina Hiseq 4000 Sequencing Platforms for

RNA Sequencing.” *Genomics and Informatics*.

- Kearse, Matthew, Richard Moir, Amy Wilson, Steven Stones-Havas, Matthew Cheung, Shane Sturrock, Simon Buxton, Alex Cooper, Sidney Markowitz, Chris Duran, Tobias Thierer, Bruce Ashton, Peter Meintjes, and Alexei Drummond. 2012. “Geneious Basic: An Integrated and Extendable Desktop Software Platform for the Organization and Analysis of Sequence Data.” *Bioinformatics (Oxford, England)* 28(12):1647–49.
- Levin, Myron J., Guang-Yun Cai, Michael D. Manchak, and Lewis I. Pizer. 2003. “Varicella-Zoster Virus DNA in Cells Isolated from Human Trigeminal Ganglia.” *Journal of Virology*.
- Lungu, Octavian, Paula W. Annunziato, Anne Gershon, Susan M. Staugaitis, Deborah Josefson, Philip Larussa, and Saul J. Silverstein. 1995. “Reactivated and Latent Varicella-Zoster Virus in Human Dorsal Root Ganglia.” *Proceedings of the National Academy of Sciences of the United States of America*.
- Markus, Amos, Linoy Golani, Nishant Kumar Ojha, Tatiana Borodiansky-Shteinberg, Paul R. Kinchington, and Ronald S. Goldstein. 2017. “Varicella-Zoster Virus Expresses Multiple Small Noncoding RNAs.” *Journal of Virology* 91(24).
- De Pelsmaecker, Steffi, Nicolas Romero, Massimo Vitale, and Herman W. Favoreel. 2018. “Herpesvirus Evasion of Natural Killer Cells.” *Journal of Virology*.
- Pollard, Martin O., Deepti Gurdasani, Alexander J. Mentzer, Tarryn Porter, and Manjinder S. Sandhu. 2018. “Long Reads: Their Purpose and Place.” *Human Molecular Genetics*.
- Prazsák, István, Norbert Moldován, Zsolt Balázs, Dóra Tombácz, Klára Megyeri, Attila Szűcs, Zsolt Csabai, and Zsolt Boldogkői. 2018. “Long-Read Sequencing Uncovers a Complex Transcriptome Topology in Varicella Zoster Virus.” *BMC Genomics* 19(1):873.
- Reuter, Jessica S. and David H. Mathews. 2010. “RNAstructure: Software for RNA Secondary Structure Prediction and Analysis.” *BMC Bioinformatics* 11(1):129.
- Rothberg, Jonathan M., Wolfgang Hinz, Todd M. Rearick, Jonathan Schultz, William Mileski, Mel Davey, John H. Leamon, Kim Johnson, Mark J.

Milgrew, Matthew Edwards, Jeremy Hoon, Jan F. Simons, David Marran, Jason W. Myers, John F. Davidson, Annika Branting, John R. Nobile, Bernard P. Puc, David Light, Travis A. Clark, Martin Huber, Jeffrey T. Branciforte, Isaac B. Stoner, Simon E. Cawley, Michael Lyons, Yutao Fu, Nils Homer, Marina Sedova, Xin Miao, Brian Reed, Jeffrey Sabina, Erika Feierstein, Michelle Schorn, Mohammad Alanjary, Eileen Dimalanta, Devin Dressman, Rachel Kasinskas, Tanya Sokolsky, Jacqueline A. Fidanza, Eugeni Namsaraev, Kevin J. McKernan, Alan Williams, G. Thomas Roth, and James Bustillo. 2011. “An Integrated Semiconductor Device Enabling Non-Optical Genome Sequencing.” *Nature*.

Rubins, Kathleen H., Lisa E. Hensley, George W. Bell, Chunlin Wang, Elliot J. Lefkowitz, Patrick O. Brown, and David A. Relman. 2008. “Comparative Analysis of Viral Gene Expression Programs during Poxvirus Infection: A Transcriptional Map of the Vaccinia and Monkeypox Genomes.” *PLoS ONE*.

Tombácz, Dóra, Zsolt Csabai, Péter Oláh, Zsolt Balázs, István Likó, Laura Zsigmond, Donald Sharon, Michael Snyder, and Zsolt Boldogkői. 2016. “Full-Length Isoform Sequencing Reveals Novel Transcripts and Substantial Transcriptional Overlaps in a Herpesvirus.” *PLoS ONE* 11(9).

Tombácz, Dóra, Zsolt Csabai, Péter Oláh, Zoltán Havelda, Donald Sharon, Michael Snyder, and Zsolt Boldogkői. 2015. “Characterization of Novel Transcripts in Pseudorabies Virus.” *Viruses* 7(5):2727–44.

Venkatesan, Sundararajan, Alan Gershowitz, and Bernard Moss. 1982. “Complete Nucleotide Sequences of Two Adjacent Early Vaccinia Virus Genes Located Within the Inverted Terminal Repetition.” *Journal of Virology*.

Wittek, R. and B. Moss. 1982. “Colinearity of RNAs with the Vaccinia Virus Genome: Anomalies with Two Complementary Early and Late RNAs Result from a Small Deletion or Rearrangement within the Inverted Terminal Repetition.” *Journal of Virology*.

Wu, Thomas D. and Colin K. Watanabe. 2005. “GMAP: A Genomic Mapping and Alignment Program for MRNA and EST Sequences.” *Bioinformatics (Oxford, England)* 21(9):1859–75.

Yang, Z., D. P. Bruno, C. A. Martens, S. F. Porcella, and B. Moss. 2011. “Genome-Wide Analysis of the 5’ and 3’ Ends of Vaccinia Virus Early



MRNAs Delineates Regulatory Sequences of Annotated and Anomalous Transcripts.” *Journal of Virology*.

Yang, Zhilong, Daniel P. Bruno, Craig A. Martens, Stephen F. Porcella, and Bernard Moss. 2010. “Simultaneous High-Resolution Analysis of Vaccinia Virus and Host Cell Transcriptomes by Deep RNA Sequencing.” *Proceedings of the National Academy of Sciences of the United States of America*.

Yang, Zhilong, Shuai Cao, Craig A. Martens, Stephen F. Porcella, Zhi Xie, Ming Ma, Ben Shen, and Bernard Moss. 2015. “Deciphering Poxvirus Gene Expression by RNA Sequencing and Ribosome Profiling.” *Journal of Virology*.

Yang, Zhilong, Craig A. Martens, Daniel P. Bruno, Stephen F. Porcella, and Bernard Moss. 2012. “Pervasive Initiation and 3'-End Formation of Poxvirus Postreplicative RNAs.” *Journal of Biological Chemistry*.