

**A PSEUDORABIES VÍRUS
TRANSZKRIPTOMIKAI ELEMZÉSE
NAGY ÁTERESZTŐKÉPESSÉGŰ
MÓDSZEREKKEL**

Ph.D. Tézis Összefoglaló

Oláh Péter Msc

**Orvosi Biológia Intézet
Interdiszciplináris Orvostudományok Doktori Iskola
Általános Orvostudományi Kar
Szegedi Tudományegyetem**

Témavezető: prof. Boldogkői Zsolt

Szeged

2017

Célkitűzések

1. A Pseudorabies vírus Kaplan törzsének *de novo* genomi DNS-szekvenálása, a korábbiakban rendelkezésre álló referencia- genomok pontosítása érdekében.
2. A Kaplan törzs első teljes transzkriptom-szekvenálása újgenerációs technológiával, több fertőzési időpontból gyűjtött vegyes mintában, potenciális új transzkriptek és izoformák feltérképezése.
3. Új transzkriptek részletes karakterizálása független módszerekkel.
4. A Transzkripció Interferencia Hálózatok hipotézisének vizsgálata a virális genomban előforduló átfedő, policisztronos génklaszterekben.

Megjegyzés: Jelen dolgozat főképp a kísérletek bioinformatikai kiértékelésére és a különböző virális szekvenciák kategorizálására fókuszál, mely a szerző legfőbb hozzájárulása a dolgozathoz felhasznált publikációkhoz. Ezek tartalmazzák a 3. sz. publikáció genomi DNS-szekvenálási eredményeit, az 1. sz. publikáció teljes transzkriptomra vonatkozó eredményeit, valamint a 2. sz. publikáció egyes eredményeit a CTO-S nem-kódoló RNS-ről.

RÖVID ÖSSZEFOGLALÁS

A Pseudorabies vírus (PRV) Kaplan törzsének DNS-szekvenciáját élenjáró, ún. „harmadik generációs” szekvenálási módszerrel határozta meg csoportunk annak érdekében, hogy az a későbbiekben transzkriptom- és epigenetikai vizsgálatok alapjául szolgálhasson. A virális transzkriptom feltérképezésére csoportunk első ízben alkalmazott nagy áteresztőképességű szekvenálási módszert, amelynek módosított változata (PA-Seq) segítségével a virális transzkriptok 3' végeinek pontos meghatározására is sor került. A várakozásoknak megfelelően a virális genom nagy része leíródott, egyes intergénikus repetitív szekvenciák, valamint az inverz repeat régiókban található bizonyos lokuszok kivételével. Az új felfedezések között szerepel egy hosszú nem-kódoló RNS az OriL replikációs origó közelében, valamint a 3' UTR-régiók és transzkripció átfedések genomi szintű feltérképezése. Számos génben volt megfigyelhető alternatív poliadeniláció, míg a korábban ismert splice-helyek sora egy újonnan felfedezett lokusszal bővült az *ep0* génben.

MÓDSZEREK

Fertőzés és vírus propagáció

A PRV Kaplan törzs fenntartása immortalizált PK-15 sertésvese sejtvonalon történt. A sejtek tenyésztése Dulbecco-féle tápoldatban, 5% foetalis borjú szérummal és 80 ug/ml gentamycinnel történt 37 °C-on 5% CO₂-tartalom mellett.

Virális DNS izolálás

A PK-15 sejt monolayerek fertőzése 10 pfu/sejt MOI-val (Multiplicity Of Infection) történt, a teljes sejtállomány-pusztulás bekövetkeztéig. A tápoldat összegyűjtését centrifugálás követte 4,000x fordulaton, 10 percen keresztül Sorvall GS-3 rotorral. A víruspartikulumok elválasztása 30%-os szukróz grádiensen történt, 24,000x-es fordulaton 1 órán keresztüli ultracentrifugálással Sorvall AH-628 rotorral. A reszuszenzió Tris-EDTA oldatban történt. A kapszidok lizálása 100 ug/ml Proteinase-K és 0.5% nátrium-dodecyl szulfát (SDS) hozzáadásával, 37 °C-on 1 órás inkubációval történt, amelyet fenol-kloroform extrakció követett.

totálRNS extrakció

A gazda és virális RNS-állomány izolálása a fertőzés egyes időpontjaiból egyesével történt, NucleoSpin RNA II izoláló kit használatával (Macherey-Nagel GmbH and Co. KG). A tápoldat centrifugálását és a sejtek kaotróp ionos oldat általi lízisét követően

a nukleinsavak izolálása szilikaoszlopon történt. A DNS-tartalom eltávolítására RNáz-mentes DNáz oldat hozzáadásával került sor. Végül az RNS-frakció elúciója RNáz-mentes vízben történt. Az esetleges fennmaradó DNS-szennyezés degradálására minden mintán Turbo DNáz kezelést (Ambion Inc.) használtunk. Az RNS-koncentráció meghatározása spektrofotométerrel (BioPhotometer Plus, Eppendorf) és Qbit fluorométerrel (Thermo Fisher Scientific) történt. Az izolált RNS-mintákat további felhasználásig -80°C -on tároltuk.

PacBio RS II gDNS előkészítés és szekvenálás

A SMRTbell szekvenálási könyvtárak előkészítése a 6-kb és 20-kb fragmensméretű könyvtárak standard preparációs előírásait követte. A szekvenancia-leolvasás 5 db single-molecule real-time (SMRT) cell-ben történt, a P5 jelű DNS-polimeráz és C5-ös jelű reagenskészlet felhasználásával, melynek eredményeként 78,111 sikeresen leolvasott DNS-fragmenst kaptunk. Ezen mennyiségű fragmensből igen magas (1,200x-os) lefedettséggel volt rekonstruálható a PRV genom. A könyvtárkészítés és szekvenálás a Stanfordi Egyetem Genetika Intézetében történt.

Illumina cDNS könyvtárkészítés és szekvenálás

A szálspecifikus, 2x100bp-os paired end totálRNS könyvtárak elkészítéséhez az Illumina-kompatibilis ScriptSeq v2 RNA-Seq könyvtárkészítő kitet (Epicenter) használtuk. A polyA(+)-RNS szekvenáláshoz single-end, 50bp-os könyvtár készült, melyben az amplifikációs lépés oligo(VN)T-10 egyedi primerek segítségével történt, melyben a két nukleotidos „horgony”-szekvenciát követi a

polyA-specifikus oligo(dT) szakasz. A horgonyzott primerek jobb hatásfokot biztosítanak azért, hogy az amplifikált fragmenseken a szekvenálás csakis a polyA-farok kezdeti szakaszától indulhat 5' irányba, ezzel biztosítva a specificitást. A könyvtárkészítés, szekvenálás és elsődleges elemzés a Debreceni Egyetem Klinikai Genomikai Központjában és Bioinformatikai Központjában történt Illumina HiScanSQ platformon.

Alternatív splicing RT-qPCR detektálása

A virális genomban előforduló splicing-események validálása két primerkészlettel történt, melyek 19 ill. 23 nukleotid hosszúságúak, az adott splice-helytől hozzávetőlegesen 100-100 bp-ra lévő felismerőhelyekkel. A reverz transzkripcióhoz 5 µl -es oldatok készültek 0.02 µg totálRNS tartalommal, 2 pmol génspecifikus primerrel, 0.25 µl dNTP mix-el, 1 µl 5× First-Strand Bufferrel, 0.25 µl (50 unit/µl) SuperScript III Reverz Transzkriptázzal (Invitrogen) és 1 unit RNAsin-nel (Applied Biosystems Inc.). A reverz transzkriptáz oldatok inkubációja 55 °C-on történt 1 órán keresztül. A reakció leállítása 15 percig 70°C-on történt. A No-RT kontroll kísérletek a SuperScript III enzim hiányában lettek elvégezve annak érdekében, hogy az esetleges virális DNS-kontamináció konvencionális PCR-rel kimutatható legyen. RT-qPCR reakciókhoz csak a DNS-mentes minták kerültek felhasználásra. Mintánként három párhuzamos mérés került rögzítésre Rotor-Gene 6000 (Corbett Life Science) készüléken. A reakciók összetétele: 7 µl 10x-esen hígított cDNS, 10 µl Absolute qPCR SYBR Green Mix (Thermo Fisher Scientific), 1.5 µl forward és 1.5 µl reverz primer

(10-10 μM). A futási paraméterek a következők: 15 perc 95 °C-on, 30 ciklus 94 °C-on 25 s-ig (denaturáció), 60 °C-on 25 s-ig (annealing), és 72 °C-on-ig 6 s (extenzió). A keletkezett termékek vizualizációja 12%-os poliakrilamid gélen történt GelRed festéssel, az előhívás pedig a ProteinSimple AlphaImager HV géldokumentációs készülékével.

Az adatok elérhetősége

A pseudorabies vírus Kaplan törzsének teljes genomszekvenciája és annotációi a KJ717942.1 GenBank kód alatt érhetők el. A kísérletek során generált nyers szekvenálási adatok fastq kiterjesztésben a European Nucleotide Archive-ban PRJEB9526 kód alatt kerültek archiválásra..

EREDMÉNYEK

A PRV Kaplan törzs genomterképe

A PRV egy széles körben kutatott herpeszvírus, azonban a Kaplan törzs genomszekvenciája csupán nem megfelelően annotált, valamint számos eltérést mutató változatokban volt elérhető. Továbbá, a jelenleg használt legjobban annotált és legszélesebb körben használt referenciagenom hat különböző törzs szekvenciáiból áll össze, így céljainknak nem felelt meg. Annak érdekében, hogy megismerjük a laboratóriumi törzsön belüli genetikai variabilitás mértékét, valamint jól használható referenciát készítsünk a későbbi transzkriptom- és epigenetikai vizsgálatokhoz, ún. „harmadik generációs” DNS-

szekvenálási eljárást alkalmaztunk. A Pacific Biosciences RS II szekvenáló által biztosított különösen hosszán leolvasott DNS-szakaszokból csaknem a teljes virális genomot lefedő kontigok voltak nyerhetőek, amelyek nagyban megkönnyítették a végleges genomai szekvencia összeállítását. A teljes genom 143,423 bázispár hosszúságú, 73,59%-os, kiemelkedően magas GC-tartalommal. Az egyéb PRV-törzsekkel való szekvencia-azonosság mértéke igen magas, 97-99% közötti, míg a törzsön belüli variabilitás alacsony mértékű. Rövid variábilis régiók főképp a nem-kódoló intergénikus szakaszokon fordulnak elő. A heterogenitás egy kiemelkedő forrása az *ul27* génben fekvő 12 bázispárnyi palindrom szekvencia, amely a virális DNS-kópiák mintegy 50%-ában hiányzik. További kísérletek során kizártuk a technikai hiba lehetőségét, valamint bebizonyosodott, hogy a repeat bizonyos mutáns Kaplan törzsekre jellemző. A fehérjekódoló szekvenciák predikciójára szintén sor került, míg a már ismert annotációk átemelése szekvencia-homológia alapján a GATU (Genome Annotation Transfer Utility) szoftverrel történt. A replikációs origók, repeatek és egyéb szekvenciamotívumok annotációja manuálisan történt. cDNS-szekvenálási kísérleteink nyomán a PRV genomban 11 fehérjekódoló gén új izoformáit, az *ep0* gén új hasítási helyét és a CTO (Close To OriL) hosszú nem-kódoló RNS-t annotáltuk. A PRV-ben korábban, az LLT látenciával

asszociált nem-kódoló RNS intronjából izolált mikroRNS-ek prekurzorait szintén annotáltuk, és így elsőként tettük elérhetővé a GenBank adatbázisában. Az informatív genomszekvencia a későbbiekben alapul szolgálhat a világszerte rendszeresen izolált fertőző törzsek jobb megértéséhez. Technikai megközelítésből kísérletünk az elsők között volt, amely egy vírus genomszekvenálásához a Pacific Biosciences SMRT (Single-Molecule Real-Time Sequencing) technológiát használta fel. A kezdeti tapasztalatok között szerepel a módszer érzékenysége a hosszú homopolimer guanin-szakaszok szekvenálásában, amely szisztematikusan felülreprezentálja ezek előfordulását a kapott konszenzusszekvenciában. Mivel az extrém magas GC-tartalmú PRV-genomban ezen hiba fokozottan jelentkezik, felfedése és kiszűrése is kézenfekvőbbé válik a változó lefedettség és szekvenciatartalom által. A hosszú leolvasásokat produkáló szekvenálási technológia mindazonáltal gazdaságosnak és jól alkalmazhatónak bizonyult a kompakt virális genom megismerésében.

cDNS-szekvenálás

A cDNS-szekvenálást Illumina nagy áteresztőképességű technológiával végeztük. A vírus minden polyA⁺ RNS-típusának átfogó megismerésére random hexamer primerekkel amplifikált totálRNS könyvtárat használtunk, míg a

pontosabb kvantitálás, transzkripciós átfedések és alternatív izoformák karakterizálására 3'-RNS-szekvenálást (PA-Seq) alkalmaztunk, amely horgonyzott primerek segítségével, nagy hatékonysággal működik. Ezen módszerekkel lehetővé vált a vírusra jellemző policisztronos génklaszterek, valamint a hozzájuk kapcsolódó Transzkripciós Interferencia Hálózatok hipotézisének vizsgálata.

Splice-helyek a PRV genomban

A random-hexamer primerekkel készült Illumina cDNS-könyvtárban került leírásra az *ep0* korai transzaktivátort kódoló gén alternatív splicingja. A rövid szakasz kivágásának funkcionális jelentősége a fehérje rendezetlen szerkezete miatt *in silico* módszerekkel nem prediktálható ill. modellezhető. Az aminosavszekvencia azonban arról tanúskodik, hogy a kisméretű *ep0* fehérje cink-ujj alegysége p marad, így elképzelhető, hogy a spliceolt izoforma is képes funkcióját ellátni.

A PRV genom további, korábban már leírt karakterisztikus hasítási helyeit szintén sikerült kimutatni, ill. azok szekvenciáit pontosítani, az *us1* és *ul15* génekben. További, a jelen kísérletekben csak kis lefedettséggel, nem szignifikánsan megjelenő alternatív splicing-eseményeket későbbi tanulmányaink igazoltak ill. pontosítottak.

A hagyományosan a látens életciklus kialakításában kulcsszerepet játszó látencia-asszociált nem-kódoló RNS-ek szintén kimutathatóak voltak, alacsony expresszióval, lítikus fertőzésből származó mintáinkban, bár csekély kifejeződésük nem tette lehetővé az LLT nem-kódoló RNS hasítási helyének azonosítását.

A PRV poliadenilációs térképe

PolyA-szekvenálás segítségével a virális génexpresszió és az egyéb PCR-alapú technikákkal gyakran nem észlelhető mRNS-ek (pl. az UL7-UL9 klaszter transzkriptjei) pontosan detektálhatók. A módszer további előnye a virális genomban sűrűn egymásba ágyazott génklaszterek összetételének felderítése az egyes mRNS-ek pontos 3'-terminusai által. Ennek ismeretében lehetőség nyílik a precíz virális génexpressziós szabályzás alaposabb megismerésére, valamint az egymásba ágyazott klaszterek kölcsönhatásainak modellezésére, a transzkripció interferencia hálózatok elmélete alapján. A konzervatív poliadenilációs szignálok használatának relatív gyakoriságát szintén megállapítottuk, amelyben az elvárásoknak megfelelően, nagyrészt az eukariota szervezetekben megfigyelhető szabályszerűségek érvényesültek, egyedi mintázatot mutattak azonban az egymásba ágyazott, policisztronos virális gének.

Új transzkriptek

Napjainkban számos új nem-kódoló RNS került leírásra olyan, klinikailag fontos patogénekben, mint a KSHV, HCMV vagy EBV. Ezek egyik közös tulajdonsága a lítikus fertőzés során mutatott igen magas expresszió, amely gyakran a vírus által leírt RNS-mennyiség 50%-át is meghaladja. Így a PRV lítikus transzkriptomjának vizsgálata során felfedezett új, 286 bp hosszúságú polyA(+) nem-kódoló RNS molekula magas expressziója kevésbé meglepő a rokon vírusokkal összehasonlítva. A CTO (Close to OriL) névre keresztelt transzkript a lítikus replikációs origóhoz közel található, a – DNS-szálon, két konvergensen orientált génklaszter között. Érdekessége a szabálytalan GC-összetétel, mely nem figyelhető meg a PRV-hez legközelebbi rokon herpeszvírusokban. A CTO funkciójának meghatározása további kísérleteket igényel, mivel direkt targetszekvenciája nem azonosítható a gazdaszervezetben; az OriL-hez való közelsége és expressziós kinetikája azonban alátámasztani látszik a felvetést, mely szerint a DNS-replikáció szabályzásában játszhat szerepet. Az egyéb karakterizált RNS-ek között szerepel az eddig csupán hipotetikus ORF 1.2, a SANC rövid 3'-antiszensz RNS, valamint számos 3'-variábilis izoforma.

Transzkripciós interferencia

A Transzkripciós Interferencia Hálózatok elmélete alapján az egymással különböző orientációkban átfedő gének, génklaszterek egymás transzkripciós mechanizmusára gyakorolnak hatást önszabályozó módon, ezzel a transzkripciós reguláció önálló szintjét képviselve. A PRV-ben mint modellorganizmusban számos olyan lokusz volt már korábban ismert, ill. került leírásra csoportunk tanulmányaiban, melyek a fenti elmélet vizsgálatára alkalmasak; köztük tandem- és divergensen orientált génklaszterek, alternatív 5' és 3' UTR-átfedések, intergénikus barrierként szolgáló repetitív szekvenciák.

Ezen eredmények, amellett hogy a témában korábban végzett, kiterjedt qPCR alapú génexpressziós tanulmányokhoz adnak többletinformációkat, a transzkripciós interferencia-jelenségek jövőbeni kutatásához is részletes genetikai térképet biztosítanak.

A doktori disszertáció témájához közvetlenül kapcsolódó publikációk listája:

- I. **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. **IF:2.581 (2015)**
- II. Tombácz, D., Csabai, Z., **Oláh, P.**, Havelda, Z., Sharon, D., Snyder, M., & Boldogkői, Z. (2015). Characterization of Novel Transcripts in Pseudorabies Virus. *Viruses*, 7(5), 2727–2744. <http://doi.org/10.3390/v7052727> **IF:3.437 (2015)**
- III. Tombácz D, Sharon D, **Oláh P**, Csabai Z, Snyder M, Boldogkői Z. Strain Kaplan of Pseudorabies Virus Genome Sequenced by PacBio Single-Molecule Real Time Sequencing Technology. *Genome Announcements*. 2014;2(4):e00628-14. doi:10.1128/genomeA.00628-14.

A doktori disszertáció témájához közvetetten kapcsolódó publikációk listája:

- IV. Tombácz D, Csabai Z, **Oláh P**, et al. 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. **IF: 3.54 (2015)**