

MULTIPLATFORM ANALYSIS OF HERPESVIRUS TRANSCRIPTOMES

Ph.D. THESIS

CSABAI ZSOLT

Orvosi Biológiai Intézet

Interdiszciplináris Orvostudományok Doktori Iskola

Általános Orvostudományi Kar

Szegedi Tudományegyetem

Témavezetők: Dr. Tombác Dóra és prof Dr Boldogkői Zsolt

Szeged

2017

Publications directly related to the subject of thesis

Csabai Zsolt*, Takács Irma, Michael Snyder, Boldogkői Zsolt, Tombác Dóra

Evaluation of the impact of ul54 gene-deletion on the global transcription and DNA replication of pseudorabies virus

ARCHIVES OF VIROLOGY pp. 1-16. (2017)

IF: 2,058

Póka Nándor*, **Csabai Zsolt***, Pásti Emese, Tombác Dóra, Boldogkői Zsolt

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)

IF: 1,431

Tombác D*, **Csabai Z***, Oláh P, Havelda Z, Sharon D, Snyder M, Boldogkői Z

Characterization of novel transcripts in pseudorabies virus

VIRUSES-BASEL 7:(5) pp. 2727-2744. (2015)

IF: 3,042

Balázs Zsolt, Tombác Dóra, Szűcs Attila, **Csabai Zsolt**, Megyeri Klára, Alexey N Petrov, Michael

Snyder, Boldogkői Zsolt

Long-Read Sequencing of Human Cytomegalovirus Transcriptome Reveals RNA Isoforms Carrying

Distinct Coding Potentials

SCIENTIFIC REPORTS 7: Paper 10.1038/s41598-017-16262-z. 9 p. (2017)

IF: 4,259

Tombácz Dóra, **Csabai Zsolt**, Szűcs Attila, Balázs Zsolt, Moldován Norbert, Donald Sharon, Michael Snyder, Boldogkői Zsolt

Long-Read Isoform Sequencing Reveals a Hidden Complexity of the Transcriptional Landscape of Herpes Simplex Virus Type 1

FRONTIERS IN MICROBIOLOGY (2017)

IF: 4,076

Olah P, Tombacz D, Poka N, **Csabai Z**, Prazsak I, Boldogkoi Z
Characterization of pseudorabies virus transcriptome by Illumina sequencing.

BMC MICROBIOLOGY 15: Paper 130. 9 p. (2015)

IF: 2,581

Tombacz D, Balazs Z, **Csabai Z**, Moldovan N, Szucs A, Sharon D, Snyder M, Boldogkoi Z:

Characterization of the Dynamic Transcriptome of a Herpesvirus with Long-read Single Molecule Real

Time Sequencing.

SCIENTIFIC REPORTS accepted for publication, 2017. jan. 26.

IF: 4,259

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 11:(9) Paper e0162868. 29 p. (2016)

IF: 2,806

Publications indirectly related to the subject of thesis

Moldován Norbert, Balázs Zsolt, Tombácz Dóra, **Csabai Zsolt**, Szűcs Attila, Michael Snyder, Boldogkői Zsolt

Multi-platform Analysis Reveals a Complex Transcriptome Architecture of a Circovirus

VIRUS RESEARCH 237: pp. 37-46. (2017)

IF:2,628

Tombácz Dora, Moldovan Norbert, Balazs Zsolt, **Csabai Zsolt**, Michael Snyder, Boldogkői Zsolt:

Genetic Adaptation of porcine Circovirus Type 1 to Cultured Porcine kidney Cells Revealed by Single-molecule Long-read Sequencing Technology.

GENOME ANNOUNCEMENTS , Accepted, scheduled publication date: 2017. febr 2.

IF: 0

Szűcs Attila, Moldován Norbert, Tombácz Dóra, **Csabai Zsolt**, Michael Snyder, Boldogkői Zsolt

Long-Read Sequencing Reveals a GC Pressure during the Evolution of Porcine Endogenous Retrovirus

GENOME ANNOUNCEMENTS 5:(40) Paper e01040-17. 2 p. (2017)

IF: 0

Tombacz D, Sharon D, Olah P, **Csabai Z**, Snyder M, Boldogkoi Z

Strain Kaplan of Pseudorabies Virus Genome Sequenced by PacBio Single-Molecule Real-Time Sequencing Technology.

GENOME ANNOUNCEMENTS 2:(4) Paper e00628-14. 3 p. (2014)

IF: 0

Not related publications to the subject of thesis

Tombácz Dóra, Maróti Zoltán, Kalmár Tibor, **Csabai Zsolt**, Balázs Zsolt, Takahashi Shinichi, Palkovits Miklós, Snyder Michael, Boldogkői Zsolt

High-Coverage Whole-Exome Sequencing Identifies Candidate Genes for Suicide in Victims with Major Depressive Disorder

SCIENTIFIC REPORTS 7: Paper 7106. 11 p. (2017)

IF: 4,259

Szenasi T, Kenesi E, Nagy A, Molnar A, Balint BL, Zvara A, **Csabai Z**, Deak F, Boros Olah B, Mates L, Nagy L, Puskas LG, Kiss I

Hmgb1 can facilitate activation of the matrilin-1 gene promoter by Sox9 and L-Sox5/Sox6 in early steps of chondrogenesis.

BIOCHIMICA ET BIOPHYSICA ACTA-GENE REGULATORY MECHANISMS 1829:(10) pp. 1075-1091. (2013)

IF: 5,440

Cumulative impact factor: 36,839

Rövidítések listája

asRNA	antiszensz RNS
cDNA	komplementer DNS
E	korai
gE	glükoprotein E
gI	glükoprotein I
HCMV	Human citomegalovírus
HSV-1	Human herpesvírus 1
IE	azonnali korai
IsoSeq	Izoforma szekvenálás
Ka	Kaplan törzs
L	késői
lncRNA	hosszú nem kódoló RNS
ncRNA	nem kódoló RNS
ORF	leolvasási keret
PacBio	Pacific Biosciences
PA- Seq	poliadenilált szekvenálás
PK-15	sértésvese sejtvonala
PRV	Pseudorabies vírus
RT ² -PCR	reverz transzkripció kapcsolt kvantitatív PCR
SMRT	Egy molekulás valós idejű
TES	transzkripció végződés
TSS	transzkripció start hely
wt	vad típusú

Bevezetés

A Herpeszvírusok családjának több mint 100 tagja van, melyek közül 8 humán patogén (Herpesz szimplex vírus 1 és 2, Citomegalovírus, Varicella zoster, Epstein-Barr vírus, a Humán herpeszvírus 6 A és B variánsa, a Kaposi szarkómát okozó vírus, vagy a Humán herpeszvírus 7 és 8). A herpeszvírusok a duplaszálú (ds)DNS vírusok közé, az egyik legnagyobb vírustaxonómiai családba a *Herpesviridae*-be tartoznak. A világ népességének 90%-a fertőzött a fentiekben felsorolt vírusok valamelyikével. A herpeszvírusokra jellemző, hogy replikációjuk, transzkripciójuk, valamint kapszidjuk összeszerelődése is a gazdasejt sejtmagjában történik, majd annak Golgi apparátusát használják fel a sejtből való kijutásra. A herpeszvírusokat három nagy alcsaládba soroljuk: alphaherpeszvírusok, bétaherpeszvírusok, valamint a gammaherpeszvírusok. A herpeszvírusok génjeinek kifejeződése időben elkülöníthető: ezek lehetnek azonnali korai gének (IE), korai (E) és kései (L) gének. Az azonnali korai gének feladata elsősorban a többi gén kifejeződéséért felelősek, jellemzően transzaktivátor génekről beszélünk. A korai gének jellemzően a vírus DNS replikációjában, míg a kései gének a vírus strukturális elemeinek felépítésében játszanak szerepet.

Pseudorabies vírus (PRV)

Az alphaherpeszvírusok alcsaládjába tartozó álveszettséget okozó vírus. Széles gazdaspecifitással rendelkezik, de legnagyobb gazdasági károkat a fiatal sertések megfertőződése okozza, ugyanis ebben az

esetben az állatok elhullásához vezet. Nem humán patogén, széles körben használt modellorganizmus. Használták korábban a herpeszvírusok patogenitásának vizsgálatára [2,3], ezen felül neuronális jelölésre [4-6], valamint fluoreszcensen jelölve alkalmas az idegpályák jelölésére is [7]. A vírus genomja rendkívül kompakt, nagyrészt fehérje kódoló gének alkotják, rövid intergenikus régiók jellemzik.

A vírus transzkriptomának kinetikai elemzése a gének policisztonos szerveződésének okán rendkívül nehéz. Korábban számos technikát használva analizálták már a vírus transzkriptomát, mint például microarray technikával [8], Illuminával és reverz transzkripció-kapcsolt real-time PCR segítségével (RT²-PCR)[9]. Azonban a transzkripció izoformák (ide értve a splice variánsokat és a hossz variánsokat is) vizsgálata rendkívül bonyolult, vagy sok esetben lehetetlen.

A gének kinetikai elemzése tehát rendkívül bonyolult a vírus esetében ugyanis a legjellemzőbb a policisztronos elrendeződés, amikor a gének különböző transzkripció indulási helyet használnak (TSS) azonban a transzkripció terminációjuk azonos helyen van, tehát egyazon poliadenilációs szignált használnak. Ez az elrendeződés a következőképpen néz ki egy 4 elemből álló tandem géncsoport esetén, ahol az '1' a legkülső gént jelenti: 1-2-3-4; 2-3-4; 3-4; 4.

Herpesz szimplex vírus 1

A Herpesz szimplex vírus 1 egy humán patogén vírus, mely sok esetben látens formában van jelen a gazdaszervezetben, akár egy élethosszon át tartó tünetmentesség is jellemezheti. Súlyosabb esetekben azonban immunhiányos betegeknél agyhártyagyulladást is okozhatnak. A PRV-hez hasonlóan szintén az alphaherpeszvírusok közé tartozik. A WHO adatai szerint 3,7 milliárd ember érintett HSV-1 fertőzésben.

A korábbi annotációk alapján elmondhatjuk, hogy a HSV-1 genomja 89 fehérje kódoló gént tartalmaz, valamint 10 hosszú nem kódoló (lnc)RNS-t, valamint számos mikroRNS-t [11]. Több genomi régióról bizonyították, hogy az antiszensz szárlól lncRNS-ek íródnak le, ezek közül az első virális lncRNS a LAT (látenciában szerepet játszó transzkript) volt [12].

A HSV-1 transzkripció aktivitásának vizsgálatára számos technikát használtak, mint amilyen a microarray technika [13], Illumina szekvenálás [14], reverz transkripció - kapcsolt valós idejű PCR-t (qRT-PCR) [15], és a PacBio SMRT- szekvenálást. Ezen technikák használatával szinte teljes genomi transzkripció aktivitás volt kimutatható [14].

Citomegalovírus (HCMV)

A HCMV a bétaherpeszvírusok alcsaládjába tartozik. A HSV-1-hez hasonlóan humán patogén vírusról beszélünk mely rendkívül a veszélyes az újszülöttekre és az immunhiányos betegekre nézve. A

veleszületett HCMV fertőzés súlyos elváltozásokat okozhat, vagy akár halálos kimenetelű is lehet. A vírus genomjának korábbi analízise során 164 és 220 közé tették a fehérje kódoló gének számát. Egy korábbi riboszóma kötőhelyek meghatározásán alapuló vizsgálatban (ribosome profiling) 751 egyedi leolvasási keretet találtak [17].

A HCMV felépítését tekintve, hasonlóan más herpeszvírusokhoz, nagy komplexitásról beszélhetünk. Alternatív transzkripciósi indulási helyek [18] épp úgy megtalálhatóak benne, mint alternatív splicing helyek [19], vagy policisztronos transzkriptek [20]. A vírus nagyfokú kódoló potenciált mutat. A splicing jelensége rendkívül gyakorinak számít a HCMV esetén, mintegy 100 splice kapcsolatot írtak le korábban, melynek nagy része alternatív splicing [18;20].

Célkitűzések

Célul tűztük ki a Pseudorabies vírus teljes transzkriptomának újrameghatározását különböző szekvenálási platformok használatával. Ezen platformok az Illumina HiSeq és a PacBio RS II platform (PA-Seq és random primer-alapú RNA-Seq). A transzkriptom analízis a sertésvese sejtek (PK-15) megfertőzésével, és az innen nyert polyA+ RNS-ek vizsgálatával indult.

Terveink között szerepelt, hogy a HSV-1 globális transzkriptomát meghatározzuk a lítikus ciklus során. Ehhez a PacBio hosszú leolvasásokat biztosító szekvenálási technológiáját terveztük használni, mely PCR amplifikáción alapul.

A HCMV transzkriptom vizsgálata során szintén a különböző transzkript izoformák, új splice variánsok, új transzkriptek álltak a vizsgálataink középpontjában, hogy meghatározzuk a vírus genom komplex expressziójának struktúráját. A vizsgálatainkhoz a virális RNS-eket a fertőzött humán fibroblaszt sejtekből nyerjük, ahol szintén a vírus lítikus ciklusa során jelenlévő transzkripteket vizsgáljuk.

A PRV génjeinek funkcionális analízise során terveink között szerepelt, hogy készítsünk olyan gényiütött vírusokat, melyeknél megvizsgáljuk, hogy az adott gén kiütése a vírusból, hogyan hat a többi vizsgált gén transzkripció aktivitására. A vizsgálni kívánt gének az us4, az us7 és az us8. A transzkripció aktivitás vizsgálatához kvantitatív valós idejű PCR technológiát használunk.

Módszerek

A vírusok szaporításához a gazdaspecificitásnak megfelelő sejteket használtuk. Több különböző fertőzési időpontot vizsgáltunk az analízis során. A DNS és RNS izolálást követően a különböző szekvenálási technikáknak megfelelő könyvtárkészítési protokollokat végeztük el. Ezen használt platformok az Illumina HiSeq, PacBio SMRT- és PacBio Isoseq.

Az alkalmazott technikákkal tudjuk vizsgálni az átfedő transzkripteket, hiszen a hosszú leolvasású szekvenálás lehetővé teszi ezen átfedő régiók elkülönítését. A kapott új kódoló és nem kódoló géneknél a leggyakoribbakat Northern blott analízissel is azonosítottuk.

Ezen kívül a mutáns vírusok analízise során a fertőzött sejtekből izolált RNS-t szál-, és génspecifikus primerekkel cDNS-sé írtuk át, majd ezt követően valós idejű kvantitatív PCR segítségével vizsgáltuk.

Eredmények

A kapott eredményeink azt mutatják, hogy az alkalmazott technikák alkalmasak voltak a teljes hosszúságú RNS-ek meghatározására, mind az egy-molekulás, mind pedig az amplifikált mintákban. Egy teljesen új transzkriptomikai profilt mutattunk a PRV tekintetében, számos új transzkripciós variánst határoztunk meg. Bebizonyítottuk, hogy a vírus teljes szekvenciáján transzkripciós aktivitás jellemző a DNS mindkét szálát érintve, mind a kódoló, mind pedig a nem kódoló

szakaszokon. A teljes virális genom transzkripciós aktivitása és az egymással átfedő transzkriptek arra engednek következtetni, hogy a leíródo transzkriptek a leíródásuk során egy új transzkripciós szabályozási színteret hoznak létre.

A szekvenálási adatokból kiderült, hogy az átfedő RNS molekulák számos kombinációban képesek transzkripteket képezni, felülmúlva ezzel a korábbi elképzeléseinket a folyamatról. Azonban a valóságban leíródo antiszensz RNS-ek száma valószínűleg jóval meghaladja az általunk talált RNS-ek mennyiségét, ugyanis ha túl rövid vagy túl hosszú RNS-ek képződnek, akkor azokat az alkalmazott szekvenálási technikákkal elveszíthetjük, ugyanígy a poly-A-t nem tartalmazó transzkriptek jó részéről sem kapunk információt.

Demonstráltuk, hogy a hosszú leolvasásokat biztosító szekvenálási technikák alkalmasak arra, hogy akár átfedő géneknél is transzkripciós dinamikát határozzunk meg. A technológia tehát jól használható más időben kontrollált kísérleteknél is a vizsgált gének kifejeződésének a mértékére. Ezen vizsgálatokat a *Pseudorabies* vírusnál végeztük el hat különböző fertőzési időpontban.

Mind a három vizsgált herpeszvírusnál megsokszoroztuk a már ismert transzkriptek számát. Ide sorolhatjuk a különböző transzkript izoformákat, az új- és az alternatív splice variánsokat, a komplex, úgynevezett policisztronos transzkripteket is, ezen felül meghatároztuk a TSS és TES helyeket is. Analízisünk során új nem kódoló és kódoló transzkripteket is felfedeztünk, mind a kódoló, mind pedig a nem kódoló régióban.

Létrehoztunk egy új deléciós mutánst a PRV esetében. Az ul54 mutáns kvantitatív PCR vizsgálata azt mutatja, hogy a gén felelős a korai génextpresszió normális kifejeződéséért, ugyanis a mutáns vírus esetében egy elhúzódó korai génkifejeződési fázis jellemző, valamint az ul54 gén szerepet játszik a virális DNS replikációjában, ezáltal hatással van a késői gének kifejeződésére is.

Másrészt megvizsgáltuk az us7-8 mutáns vírus génjeinek kifejeződését is a vad típushoz képest. Ezen gének a vírus glükoprotein génjeit kódolják. Azt vártuk, hogy egy lassabb fertőzési idő lesz jellemző a mutáns vírusra, azonban ez nem volt tapasztalható. Az us7-8 deletáns vírus a fertőzés első hat órájában leszabályozza a globális transzkripciót a kinetikai osztályoktól függetlenül, a hat óras fertőzés után viszont egy túlexpresszált állapotot tapasztaltunk a vizsgált gének tekintetében.

Referenciák

- [1] Harkness, J. M., Kader, M., & DeLuca, N. A. (2014). Transcription of the Herpes Simplex Virus 1 Genome during Productive and Quiescent Infection of Neuronal and Nonneuronal Cells. *Journal of Virology*, 88(12), 6847–6861. <http://doi.org/10.1128/JVI.00516-14>
- [2] Pomeranz, L. E., Reynolds, A. E., & Hengartner, C. J. (2005). Molecular Biology of Pseudorabies Virus: Impact on Neurovirology and Veterinary Medicine. *Microbiology and Molecular Biology Reviews*, 69(3), 462–500. <http://doi.org/10.1128/MMBR.69.3.462-500.2005>
- [3] Szpara, M. L., Kobiler, O., & Enquist, L. W. (2010). A Common Neuronal Response to Alpha herpesvirus Infection. *Journal of Neuroimmune Pharmacology*, 5(3), 418–427. <http://doi.org/10.1007/s11481-010-9212-0>
- [4] Strack, A. M. (1994). Pseudorabies virus as a transneuronal tract tracing tool: specificity and applications to the sympathetic nervous system. *Gene Therapy, 1 Suppl 1*, S11-4. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/8542383>
- [5] Card, J. P., & Enquist, L. W. (2014). Transneuronal Circuit Analysis with Pseudorabies Viruses. In *Current Protocols in Neuroscience* (Vol. 68, p. 1.5.1-1.5.39). Hoboken, NJ, USA: John Wiley & Sons, Inc. <http://doi.org/10.1002/0471142301.ns0105s68>
- [6] Boldogkői, Z., Sík, A., Dénes, A., Reichart, A., Toldi, J., Gerendai, I., ... Palkovits, M. (2004). Novel tracing paradigms--genetically engineered herpesviruses as tools for mapping functional circuits within the CNS: present status and future prospects. *Progress in Neurobiology*, 72(6), 417–45. <http://doi.org/10.1016/j.pneurobio.2004.03.010>
- [7] Boldogkoi, Z., Balint, K., Awatramani, G. B., Balya, D., Busskamp, V., Viney, T. J., ... Roska, B. (2009). Genetically timed, activity-sensor and rainbow transsynaptic viral tools. *Nature Methods*, 6(2), 127–30. <http://doi.org/10.1038/nmeth.1292>

- [8] Stingley, S. W., Ramirez, J. J., Aguilar, S. A., Simmen, K., Sandri-Goldin, R. M., Ghazal, P., & Wagner, E. K. (2000). Global analysis of herpes simplex virus type 1 transcription using an oligonucleotide-based DNA microarray. *Journal of Virology*, 74(21), 9916–27. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/11024119>
- [9] Tombácz, D., Tóth, J. S., Petrovszki, P., & Boldogkoi, Z. (2009). Whole-genome analysis of pseudorabies virus gene expression by real-time quantitative RT-PCR assay. *BMC Genomics*, 10(1), 491. <http://doi.org/10.1186/1471-2164-10-491>
- [10] Looker KJ, Magaret AS, May MT, et al. Global and Regional Estimates of Prevalent and Incident Herpes Simplex Virus Type 1 Infections in 2012. DeLuca NA, ed. *PLoS One*. 2015;10(10):e0140765. doi:10.1371/journal.pone.0140765.
- [11] Du, T., Han, Z., Zhou, G., Zhou, G., and Roizman, B. (2015). Patterns of accumulation of miRNAs encoded by herpes simplex virus during productive infection, latency, and on reactivation. *Proc. Natl. Acad. Sci. U. S. A.* 112, E49-55. doi:10.1073/pnas.1422657112.
- [12] Stroop, W. G., Rock, D. L., and Fraser, N. W. (1984). Localization of herpes simplex virus in the trigeminal and olfactory systems of the mouse central nervous system during acute and latent infections by in situ hybridization. *Lab. Invest.* 51, 27–38. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/6330452> [Accessed February 18, 2016].
- [13] Stingley, S. W., Ramirez, J. J., Aguilar, S. A., Simmen, K., Sandri-Goldin, R. M., Ghazal, P., et al. (2000). Global analysis of herpes simplex virus type 1 transcription using an oligonucleotide-based DNA microarray. *J. Virol.* 74, 9916–27. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/11024119> [Accessed February 4, 2016].
- [14] Harkness, J. M., Kader, M., and DeLuca, N. A. (2014). Transcription of the herpes simplex virus 1 genome during productive and quiescent infection of neuronal and nonneuronal cells. *J. Virol.* 88, 6847–61. doi:10.1128/JVI.00516-14.
- [15] Tombácz, D., Tóth, J. S., Petrovszki, P., and Boldogkoi, Z. (2009).

Whole-genome analysis of pseudorabies virus gene expression by real-time quantitative RT-PCR assay. *BMC Genomics* 10, 491. doi:10.1186/1471-2164-10-491.

- [16] Rubin, R. H. Impact of Cytomegalovirus Infection on Organ Transplant Recipients. *Clin. Infect. Dis.* **12**, S754–S766 (1990)
- [17] Stern-Ginossar, N. *et al.* Decoding human cytomegalovirus. *Science* **338**, 1088–93 (2012).
- [18] Isomura, H. *et al.* Noncanonical TATA sequence in the UL44 late promoter of human cytomegalovirus is required for the accumulation of late viral transcripts. *J. Virol.* **82**, 1638–46 (2008).
- [19] Rawlinson, W. D. & Barrell, B. G. Spliced transcripts of human cytomegalovirus. *J. Virol.* **67**, 5502–13 (1993).
- [20] Ma, Y. *et al.* Human CMV transcripts: an overview. *Future Microbiol.* **7**, 577–593 (2012).