# MULTIPLATFORM ANALYSIS OF HERPESVIRUS TRANSCRIPTOMES

**Ph.D. THESIS**

**ZSOLT CSABAI**

Department of Medical Biology

Doctoral School of Interdisciplinary Medicine

Faculty of Medicine

Supervisors: Dr. Dóra Tombácz and Prof Dr. Zsolt Boldogkői

Szeged

2017

# Publications directly related to the subject of thesis

**Csabai Zsolt***, Takács Irma, Michael Snyder, Boldogkői Zsolt, Tombácz 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, Tombácz 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

Tombácz D*, **Csabai Z***, Oláh P, Havelda Z, Sharon D, Snyder M, Boldogkoi Z
Characterization of novel transcripts in pseudorabies virus
**VIRUSES-BASEL** 7:(5) pp. 2727-2744. (2015)
IF: 3,042

Balázs Zsolt, Tombácz 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 Transciptome 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, Boldogkoi Zsolt:

Genetic Adaptation of porcine Circovirus Type 1 to Cultured Porcine kidney Cells Revealed be 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**

**List of abbreviations**

| | |
|---|---|
| asRNA | antisense RNA |
| cDNA | complementary DNA |
| E | early |
| gE | gylcoprotein E |
| gI | gylcoprotein I |
| HCMV | Human cytomegalovirus |
| HSV-1 | Human herpesvirus 1 |
| IE | immediate-early |
| IsoSeq | Isoform Sequencing |
| Ka | Kaplan strain |
| L | late |
| lncRNA | long non coding RNA |
| ncRNA | non coding RNA |
| ORF | open reading frame |
| PacBio | Pacific Biosciences |
| PA- Seq | polyadenylated sequencing |
| PK-15 | porcine kidney cell line |
| PRV | pseudorabies virus |
| $RT^2$-PCR | Reverse transcription linked real-time PCR |
| SMRT | Single Molecule Real-time |
| TES | transcription end site |
| TSS | transcription start site |
| UL | unique long |
| wt | wild type |

## Introduction

Herpesviridae is a large family of double stranded (ds)DNA viruses. Herpresviruses have more than 100 members, 8 of these are human pathogen (herpes simplex virus types 1 and 2, varicella-zoster virus, cytomegalovirus, Epstein-Barr virus, human herpesvirus 6 (variants A and B), human herpesvirus 7, and Kaposi's sarcoma virus or human herpesvirus 8.). Worldwide ~90 % of population have been infected with one of these viruses.

Herpesviruses are nuclear replicating viruses; transcription, genome replication and capsid assembly occur in the host cell nucleus. The intracellular trafficking of these viruses is connected to Golgi transport

Herpesviruses are divided into three groups: alphaherpesviruses, betaherpesviruses, gammaherpesviruses. Herpesvirus genes are expressed in a coordinated temporal cascade and grouped into three kinetic classes, immediate-early (IE), early (E) and late (L) [1]. The IE proteins are required for the transcription of both E and L genes. The E genes typically encode proteins that play a role in DNA replication, while the L genes specify the structural components of the virus.

## Pseudorabies virus

The pseudorabies virus (PRV), an alphaherpesvirus with a broad host range, causes fatal encephalitis in a wide variety of animals, with the exception of its natural reservoir, the adult pig. It is a commonly employed model organism in studies of the molecular pathogenesis of herpesviruses [2,3], for labeling neural circuits [4–6] and for the delivery of genetically-encoded fluorescent activity markers to the neurons [7]. The genomes of viruses are very compact, composed mainly of protein-coding genes and short intergenic regions.

Kinetic analysis of the PRV transcriptome confronts a serious problem in due to the polycistronic organization of the viral genes. Previous approaches to analysis of the herpesvirus transcriptome have used microarrays [8], Illumina sequencing, and real-time reverse transcription PCR ($RT^2$-PCR) analysis [9]. However, the identification of transcript isoforms, including splice and length variants, with these techniques is difficult or impossible.

Kinetic analysis of the herpesvirus transcriptome faces a significant challenge due to the overlapping nature of the viral genes. The typical architecture of polycistronic units is characterized by varying transcription start sites (TSSs) that are caused by the control of distinct promoters, and shared

transcription end sites (TESs). As an example, the following transcripts are produced from a tetracistronic unit: 1-2-3-4, 2-3-4, 3-4 and 4, where '1' represents the most upstream gene, while '4' is the most downstream gene within the given unit.

## Herpes simplex virus 1

Herpes simplex virus type 1 (HSV-1) is a human pathogenic alphaherpesvirus from the *Herpesviridae* family. Herpes is a lifelong infection, which often has mild or no symptoms. The most common symptoms of viral infection are cold sores. HSV-1 can cause acute encephalitis in immunocompromised patients. According to WHO's first global estimates, worldwide more than 3.7 billion people under the age of fifty are infected with HSV-1 [10]. According to earlier annotations, the HSV-1 DNA contains 89 protein-coding, 10 long non-coding (lnc)RNA genes and several micro RNAs [11]. Several genomic regions containing protein-coding genes also encode antisense lncRNAs from the complementary DNA strand. The HSV LAT was described as the first viral lncRNA [12].

Various methods have already been used for the analysis of the herpesvirus transcriptome including microarrays [13], Illumina sequencing [14], multi-time-point real-time reverse transcription PCR (qRT-PCR) analysis [15], and PacBio SMRT sequencing. Next-generation sequencing platforms have only

been used for analysing the transcriptional activity along the viral genome [14].

## Cytomegalovirus

The Human Cytomegalovirus (HCMV) is a human pathogenic beta-herpesvirus that can cause life-threatening infections in new-born infants and immunocompromised patients. Congenital HCMV infections can lead to severe malformations or even death [16]. The genome of the HCMV is one of the largest in the *Herpesviridae* family, and its coding potential is not fully understood. The number of its protein coding sequences ranged from 164 to 220, while a recent study identified 751 individuals, translationally active open reading frames (ORFs) by ribosome profiling [17].

HCMV, similarly to other herpesviruses, has a complex transcriptional architecture; alternative transcription initiation [18], alternative splicing events [19], and polycistronic transcripts [20] all increase the coding potential of the viral genome. Splicing in herpesviruses is relatively rare [20], over 100 splice junctions have been described in HCMV [18;20] – many of which are alternatively spliced.

# AIMS

To re-evaluate the currently available knowledge concerning the structures of PRV transcripts by using Illumina HiSeq and PacBio RS II platform (PA-Seq and random primer-based RNA-Seq), which can identify all poly(A)$^+$ RNA molecules generated in cultured porcine kidney (PK-15) cells productively infected with the virus.

Using PacBio long-read sequencing technology for the characterization of the global lytic transcriptome of HSV-1. Application an amplified isoform sequencing (Iso-Seq) protocol that based on PCR amplification of the cDNAs prior to sequencing.

Our focus was to identify novel transcripts, transcript isoforms, novel splice junctions, and to determine the coding potential of these transcripts, in HCMV RNA population in human fibroblast cells during lytic infection.

Generate an *ul54*-KO virus and examine the effects of the mutation on the replication and global transcription of PRV by using quantitative real-time-PCR and reverse transcription (qRT)-PCR platforms.

Characterisation the dynamic transcriptome of *us7/us8*-deleted PRV in comparison with the wild-type (*wt*) virus, using a multi-time-point quantitative reverse transcriptase-based real-time PCR technique.

## Methods

Cells were infected at a multiplicity of infection, after the DNA and RNA sample preparation three different sequencing technique were used; Illumina HiSeq, Pacbio SMRT- and Pacbio Isoseq. These techniques allow for distinction between transcript isoforms, including length- and splice variants, as well as between overlapping polycistronicRNA molecules. The non-amplified Isoform Sequencing method was used to analyse the kinetic properties of the lytic PRV transcripts and to then classify them accordingly.

We investigated the effect of the deletion of the PRV us7 and us8 and ul54 genes on the genome-wide transcription using reverse transcription based real-time pcr technique.

## Results and conclusion

Our results revealed the feasibility of the deep sequencing of full-length RNA molecules from the transcriptome of a herpesvirus both at a single-molecule level and in amplified samples. Our investigations essentially redefine the transcriptome of the PRV. We demonstrated that herpesviruses exhibit considerably more genetic complexity than predicted from *in silico* ORF-based genome annotations and gel-based assays. Our investigations uncovered that essentially the entire PRV genome is transcriptionally active, including both DNA strands of the coding and intergenic sequences. Identification of a pervasive genome-wide overlapping pattern of PRV transcripts and of *ori*-overlapping RNA molecules raise the possibility for the potential existence of a genome-wide network exerting joint control on gene expression and replication.

Our investigations revealed an intricate meshwork of transcriptional read-throughs leading to overlapping RNA molecules. It turned out that herpesvirus genes are

transcribed in more combinations than it had been previously thought. The number of asRNAs and the complex transcripts of herpesviruses are likely to be underestimated, because most of them may have been undetected due to their non-polyadenylated nature or because they are too long to be identifiable with even a long-read platform.

We demonstrated the utility of long-read sequencing for the investigation of the dynamic transcriptome of a herpesvirus. We have established that this technique can also be applied in the study of processes exhibiting a definite, well-controlled time-course of transcription, such as during viral replication, embryogenesis, tissue regeneration. We have characterised the kinetic properties of several novel PRV transcripts

We described the generation of a mutant PRV strain with a deletion at the *ul54* locus and the transcriptional characterization of this virus in cultured cells using a real-time RT-PCR technique. We also analyzed the dynamics of viral DNA synthesis and correlated the obtained data on

replication with the transcription patterns of the viral genes. We obtained that the abrogation of *ul54* function leads to a differential effect on the various kinetic classes of the PRV genes. This effect may be direct at the early phase of gene expression, but later this mutation likely exerts its influence on global gene expression at least partly through the DNA replication, which is impeded compared to the *wt* virus.

We investigated the role of gE/gI protein, which may be unrelated with spreading by the analysis of the impact on the mutation on global transcriptome. Our results reveal that the deletion of the *us7* and *us8* genes of PRV leads to significant overall reduction of gene expressions in the first six hours post.infection. in every kinetic class of genes without bias toward any of them. However, later (8-24h pi) the genes are upregulated in the *mutant* virus compared to the *wt* virus. This faciliatory effect was much higher on the E and E/L genes compared to the L genes, which is indicated by the decrease of the relative contribution of L gene products to the global viral transcript in the null mutant.

# References

[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 Alphaherpesvirus 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] Boldogköi, 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).