Analysis of large DNA viruses by long-read RNA sequencing

Ph. D. Thesis

Gábor Torma

Department of Medical Biology

Doctoral School of Interdisciplinary Medicine

Faculty of Medicine

University of Szeged

Supervisor: Zsolt Boldogkői Prof. Dr., Ph. D, DSc

Dóra Tombácz Dr., Ph. D

Szeged

- 2024 –

Publications directly related to the subject of the thesis

(1) Olasz F, Tombácz D, **Torma G**, Csabai Z, Moldován N, Dörmő Á, Prazsák I,
Mészáros I, Magyar T, Tamás V, Zádori Z, Boldogkői Z. Short and Long-Read
Sequencing Survey of the Dynamic Transcriptomes of African Swine Fever Virus and the Host Cells. Front Genet. 2020 Jul 28;11:758. doi: 10.3389/fgene.2020.00758. **IF: 4,599**

(2) Torma G, Tombácz D, Csabai Z, Göbhardter D, Deim Z, Snyder M, Boldogkői Z. An Integrated Sequencing Approach for Updating the Pseudorabies Virus Transcriptome. Pathogens. 2021 Feb 20;10(2):242. doi: 10.3390/pathogens10020242. **IF: 4,531**

(3) Torma G, Tombácz D, Csabai Z, Moldován N, Mészáros I, Zádori Z, Boldogkői Z. Combined Short and Long-Read Sequencing Reveals a Complex Transcriptomic Architecture of African Swine Fever Virus. Viruses. 2021 Mar 30;13(4):579. doi:10.3390/v13040579. **IF: 5,818**

Introduction

All the RNA transcribed in the genome of an organism is called the transcriptome[1]. This includes protein-coding and non-coding RNAs (ncRNAs). Understanding them is important for a precise understanding of genetic regulations[2]. The main goal of transcriptomics is to study these RNA molecules at different stages of individual development.

The transcriptome studies were initially carried out with Northern blot and Real-time PCR, but these can provide information on a small part of the transcriptome[3,4]. The beginning of breakthroughs was the appearance of Sanger sequencers, as their use during the Human Genome Project awakened a huge demand for innovation in the development of high-throughput sequencing platforms[5,6]. This has led to the emergence of next-generation sequencing platforms, which have a major advantage over Sanger sequencers in that they do not require electrophoretic separation and bacterial cloning[7]. The very first representative of these sequencers was Roche's 454, whose biggest advantage was the long ~1 kb reads[7,8].

Parallel to it, the Ion PGM platform was also published, which does not require optical detection, as it detects pH changes that occur during the incorporation of nucleotides[9]. Subsequently, the Illumina platform appeared in 2007, which is currently the most widespread short read sequencer[10,11]. Its use led to a significant increase in our knowledge about the sequences of individual organisms. Current genome projects run on Illumina platforms, benefiting from their high throughput [12]. Its operation is based on the detection of the fluorescence of labeled nucleotides incorporated during synthesis[8]. However, due to the short sequence readings, this platform is not suitable for the detection of full-length RNA molecules[13]. Therefore, third-generation sequencers have recently appeared in transcriptome analysis, which are able to detect full-length RNA molecules end-to-end. This platform currently has two representatives, one is the minION devices developed by Oxford Nanopore, and the other is PacBio's RSII and Sequel devices. The two technologies are based on different sequencing principles. During PacBIO sequencing, the molecule to be sequenced is ligated to a hairpin-shaped adapter, which, loaded onto

an SMRTcell, will be copied by a surface-fixed polymerase with fluorescently labeled nucleotides, the fluorescence of which will be detected by an optical sensor[14]. In contrast, with ONT-minION, sequencing will take place on a nanopore, which will be formed by alpha-hemolysins embedded in a lipid bilayer[15,16]. In the case of this library preparation, a motor protein is ligated to the molecule to be sequenced, which helps the RNA molecule enter the pore[7,17]. The RNA molecule entering the nanopore will create a voltage shift characteristic of a specific sequence. Furthermore, using ONT-minION allows for direct RNA sequencing as well[18]. Its significant advantage is that during library preparation, there is no need for reverse transcription or PCR amplification. Consequently, only the native RNA molecule is sequenced. This effectively eliminates errors in transcriptome analysis, such as template switching and false priming during library preparation, which can lead to the detection of false intron and transcription end sites[19]. With the emergence of these two platforms, our image of the transcriptome has become much more complex over the past years. One interesting application is the analysis

of viral transcriptomes, where the use of these platforms has greatly increased the discovery of new transcript isoforms, non-coding RNAs, and intron variants, shedding light on new aspects of viral transcriptome regulation[20].

African Swine Fever (ASFV) is a nucleocytoplasmic virus (NCLDV) belonging to the family Asfarviridae[21]. It infects pigs and wild boars, causing an acute, fatal hemorrhagic disease. Its replication will take place in the monocyte/macrophage cells of the host[22]. It has an icosahedral morphology, its diameter is ~200 nm[23,24] The genome is linear double-stranded DNA,with a size ranging from 170 to 190 kb. All the proteins necessary for replication and mRNA synthesis will be present in the virion, so its replication will occur in the cytoplasm[25,26]. The structure of its genome is characterized by the fact that it contains a conserved and a variable region in the middle and an inverted terminal repeat at the two edges, where the replication of the virus will take place, similar to Poxviruses[25,27]. The viral mRNAs have a 5' cap and a 3' poly(A) tail[28,29]. The synthesis of its RNAs is characterized by the temporal

gene expression typical of Poxviruses[30,31]. The synthesis of its RNAs can be classified into four kinetic classes: immediate-early (IE), early (E), intermediate (I) and late (L)[32,33]. The immediate early and early genes will be responsible for coding proteins involved in replication, while the intermediate and late genes will code for the structural proteins of the virus. The mRNAs of the virus were previously examined by Cackett et al. with CAGE and poly(A) short read sequencing[34].

The Pseudorabies virus (PRV) is a swine herpesvirus belonging to the family Alphaherpesviridae[35]. It causes neurological diseases in pigs[36,37]. Its genome is a linear double-stranded DNA, approximately ~143 kb in length[38–40]. Structurally, the virus DNA consists of two parts, the UL (long unique) and US (short unique) regions [41]. The US region is bounded by an inverted and a terminal repeat, whose recombination can result in two different isomers of the genome. Additionally, the IE180 and US1 genes are found in duplicate in this region[42]. The UL region contains a replication origin (OriL), while the US region contains two replication origins (OriS)[43,44]. Most of

the PRV genes will be expressed in a polycistronic form [45–47]. For transcription, the host transcriptional apparatus will be utilized, but the virus will express some transcription factors, such as IE180, US1, and EP0, for example[42]. Its genes can be classified into three kinetic classes: immediate-early, early, and late[48]. During infection, the immediate-early genes are expressed first, which include transcriptional activators such as IE180[42]. Subsequently, the early genes are expressed, playing a role in replication. Replication occurs via a rolling circle mechanism[49]. The expression of early genes is followed by the late genes, which encode structural proteins.[39,42]. After infection, the virus will persist in the trigeminal ganglion of the host[42,50]. In this latency, the LAT (Latency-associated transcripts) gene is the only one that is transcriptionally active. Its expression is inhibited when the IE180 protein binds. To allow the virus to persist in the host cell, LAT expression supports the survival of neurons and has anti-apoptotic effects.

Aims

In my thesis, I present the transcriptome analysis of African Swine Fever and Pseudorabies viruses. For these investigations, we utilized sequencing platforms that provide long reads, such as PacBIO and ONT-minION devices. We set the following goals for the analysis:

We set the following goals for the analysis:

1. Determination of the transcription start sites and transcription end points of viral mRNAs for both viruses with base pair accuracy.
2. Connecting annotated transcription start sites and transcription end points to form full-length transcripts.
3. Categorizing annotated transcripts structurally, such as identifying individual isoforms, non-coding RNAs, antisense, or hypothetical transcripts.
4. Identifying promoter elements and polyadenylation signals for the 5' and 3' ends of the RNAs.

Materials and methods

The African Swine Fever virus was cultured on primary alveolar macrophage (PAMs) cells, while the Pseduorabies virus was cultured on porcine kidney-15 (PK-15) cells. Infections were stopped at different time points, and RNA was isolated. Poly(A) selection was performed on the total RNA of both viruses, in the case of PRV, ribodepletion and cap selection were also performed on a part of the samples. From the poly(A)-selected samples, PCR-amplified cDNA, non-amplified dcDNA, and native dRNA sequencing libraries for ONT-minION were prepared. For PRV, PacBio RSII and Sequel sequencing libraries were also created. Subsequently, the viral transcriptome was sequenced on these LRS platforms. The data generated by the sequencers were base-called using the SMRTlink program package for PacBio and the Guppy software for ONT. Reads were mapped to the reference genome using the Minimap2 program. Following this, a Python-based transcript annotating package called LoRTIA was used for quantifying and characterizing the transcripts in the

resulting files. The IGV and Genious software tools were employed for visualization of the reads and annotated transcripts.

Results

In the case of PRV, we identified 465 transcription start sites (TSS) and 57 transcription end sites (TES), from which we annotated a total of 619 transcripts. Among these, 166 long and 24 short 5' UTR isoforms, as well as 22 3' UTR variants, were determined. Additionally, we identified 209 5' truncated transcripts, which are RNAs containing in-frame ORFs within genes. One significant discovery from the long-read sequencing was the identification of replication-associated transcripts. For OriL, we detected new length and alternative termination isoforms of CTO transcripts, as well as a new spliced isoform of PTO-US1 around the OriS region. Furthermore, we identified long non-coding transcript (>200bp) isoforms localized in the US region. Examples include new length and splice isoforms of the NOIR transcript, terminating co-terminally with AST and LLT transcripts, as well as antisense Azure transcripts to the

US3 gene. We also refined the previously annotated TSS of the US4 antisense transcript. Moreover, we identified a new length variant of the fORF15 transcript, originating from the UL15 intron region and lacking an in-frame ATG. Multicistronic transcripts that overlap multiple genes were also identified, with 87 being polycistronic and 24 being complex transcripts.

For ASFV, 202 TSS and 220 TES were identified, from which a total of 311 transcripts were annotated. This included 14 long and 2 short 5' UTR isoforms, along with 57 3' UTR isoforms. Sixteen 5' truncated transcripts were identified. Additionally, we found seven hypothetical genes containing small ORFs in the intergenic region of the virus. Multiple non-coding RNAs were identified, including 3' truncated RNAs, antisense transcripts, and replication-associated RNAs. Twenty-two 3' truncated RNAs were identified, containing the canonical ATGs of ORFs but lacking a STOP codon, indicating no functional ORFs. Seven antisense RNAs were also identified. Furthermore, in the inverted repeats at the genome ends where replication initiates, seven low-abundance

replication-associated RNAs were found. Until recently, it was assumed that ASFV expresses only monocistronic RNAs. However, long-read sequencing revealed the expression of multicistronic RNAs. Fifty-one polycistronic RNAs and twenty-two complex RNAs were annotated.

Discussion

In my thesis, I discuss the examination of the transcriptome of ASFV and PRV using sequencing platforms that provide long reads. Using these these approaches, we have identified numerous new transcript categories that were previously unknown.

As a result of our investigations, we have identified numerous new transcription start sites and transcription end points. Additionally, we have discovered new 5' and 3' UTR isoforms. The 5' UTRs may contain upstream open reading frames (ORFs) that can influence translation efficiency[51]. The 3' UTRs can contain sequences that can bind microRNAs, thereby influencing protein synthesis[52].

The Long-Read Sequencing (LRS) revealed a broad spectrum of 5' truncated hypothetical RNAs. These are RNAs that are transcribed from ORFs within a gene, meaning they share a canonical stop codon with the canonical gene but have their ATG downstream from the canonical ATG. Furthermore, in addition to identifying 5' truncated RNAs within genes, in the case of ASFV, we also found RNAs containing small open reading frames (ORFs) in intergenic regions. Among non-coding RNAs, we identified antisense transcripts that can be transcribed by independent promoters (LLT, AST, AZURE) or may arise from transcriptional overlap of convergent gene pairs or 5' UTR overlap in divergent gene pairs[53]. Numerous such antisense transcripts were identified in both PRV and ASFV. An example in PRV is Azure located in the US region, which is antisense to US3, and US4-AS, which is oriented opposite to the US4 gene. In the case of ASFV, we detected 7 such antisense molecules with opposing orientations.

Another significant result of these platforms was that a new class of RNAs was identified with them, which

include transcripts transcribing replication origins and transcripts near replication origins[54]. They are called replication-associated RNAs (raRNAs). In the case of PRV, we annotated such molecules around all three replication origins. Our analysis revealed that the most abundantly expressed RNA is the CTO-S molecule located next to OriL in the PRV UL region. For CTO-S, we identified mRNA isoforms that overlap Ori, as well as alternative termination isoforms that overlap with the neighboring UL22 gene. The biological role of these RNAs is unclear, they probably have a role in the initiation of replication and in determining the direction of the replication fork, but they also have poly(A) sequences, which suggests that they can also function as mRNA molecules in the virus life cycle[54]. They may also have other functions, such as forming RNA-DNA hybrids[55]. We also identified such molecules around OriS in the US region of the virus. Among these, we identified the new splice variant of PTO-US1, which crosses the origin of replication and ends with US1 in a common co-terminal, and we also identified the splice and length isoforms of NOIR-1 transcripts, which are localized near OriS and

share common features with LLT/AST genes 3 ' their coterminal. A possible role of these overlaps may be that NOIR-1 somehow inhibits the expression of these latency genes. However, the proof of this hypothesis requires further studies. In the case of ASFV, we also identified 6 such raRNAs molecules, which are written in the inverted repeats located at the ends of the genome, from where replication starts. Among them, we identified non-coding transcripts, as well as the mRNA isoform starting from the protein-coding gene, which end co-terminally with the non-coding transcripts in the terminal repeats.

Multicistronic RNAs are molecules that overlap multiple genes, transcribing genes with the same polarity (polycistronic transcript) and genes with opposite polarity (complex transcript). In our analysis, we detected numerous such long RNA molecules in both PRV and ASFV viruses.

Based on our results, identified 5' and 3' UTR overlaps, as well as multicistronic transcripts, create a broad network of transcriptional overlaps on the viral genome. These overlaps can be divergent, convergent, or

parallel based on orientation. We hypothesize that their possible role could be in transcriptional interference, regulating the expression of neighboring genes through the transcriptional overlaps occurring on the viral genome[56].

References

1. K-H Liang. Bioinformatics for Biomedical Science and Clinical Applications - 1st Edition. https://shop.elsevier.com/books/bioinformatics-for-biomedical-science-and-clinical-applications/liang/978-1-907568-44-2 (2023).
2. Wang, Z., Gerstein, M. & Snyder, M. RNA-Seq: a revolutionary tool for transcriptomics. *Nat Rev Genet* **10**, 57–63 (2009).
3. Heather D. VanGuilder. Twenty-five years of quantitative PCR for gene expression analysis | BioTechniques. https://www.future-science.com/doi/10.2144/000112776 (2023).

4. Southern, E. The early days of blotting. *Methods Mol Biol* **1312**, 1–3 (2015).
5. Heather, J. M. & Chain, B. The sequence of sequencers: The history of sequencing DNA. *Genomics* **107**, 1–8 (2016).
6. Collins, F. S., Morgan, M. & Patrinos, A. The Human Genome Project: Lessons from Large-Scale Biology. *Science* **300**, 286–290 (2003).
7. Dijk, E. L. van, Jaszczyszyn, Y., Naquin, D. & Thermes, C. The Third Revolution in Sequencing Technology. *Trends in Genetics* **34**, 666–681 (2018).
8. Metzker, M. L. Sequencing technologies — the next generation. *Nat Rev Genet* **11**, 31–46 (2010).
9. Liu, L. *et al.* Comparison of next-generation sequencing systems. *J Biomed Biotechnol* **2012**, 251364 (2012).

10. Weimer, B. C. 100K Pathogen Genome Project. *Genome Announc* **5**, e00594-17 (2017).
11. Turnbull, C. *et al.* The 100 000 Genomes Project: bringing whole genome sequencing to the NHS. *BMJ* **361**, k1687 (2018).
12. van Dijk, E. L., Auger, H., Jaszczyszyn, Y. & Thermes, C. Ten years of next-generation sequencing technology. *Trends Genet* **30**, 418–426 (2014).
13. Byrne, A., Cole, C., Volden, R. & Vollmers, C. Realizing the potential of full-length transcriptome sequencing. *Philosophical Transactions of the Royal Society B: Biological Sciences* **374**, 20190097 (2019).
14. Eid, J. *et al.* Real-time DNA sequencing from single polymerase molecules. *Science* **323**, 133–138 (2009).

15. Lu, H., Giordano, F. & Ning, Z. Oxford Nanopore MinION Sequencing and Genome Assembly. *Genomics Proteomics Bioinformatics* **14**, 265–279 (2016).
16. Schneider, G. F. & Dekker, C. DNA sequencing with nanopores. *Nat Biotechnol* **30**, 326–328 (2012).
17. Satam, H. *et al.* Next-Generation Sequencing Technology: Current Trends and Advancements. *Biology* **12**, 997 (2023).
18. Garalde, D. R. *et al.* Highly parallel direct RNA sequencing on an array of nanopores. *Nat Methods* **15**, 201–206 (2018).
19. Cocquet, J., Chong, A., Zhang, G. & Veitia, R. A. Reverse transcriptase template switching and false alternative transcripts. *Genomics* **88**, 127–131 (2006).

20. Boldogkői, Z., Moldován, N., Balázs, Z., Snyder, M. & Tombácz, D. Long-Read Sequencing – A Powerful Tool in Viral Transcriptome Research. *Trends in Microbiology* **27**, 578–592 (2019).
21. Reis, A. L., Netherton, C. & Dixon, L. K. Unraveling the Armor of a Killer: Evasion of Host Defenses by African Swine Fever Virus. *J Virol* **91**, e02338-16 (2017).
22. Zhu, J. J. *et al.* Mechanisms of African swine fever virus pathogenesis and immune evasion inferred from gene expression changes in infected swine macrophages. *PLOS ONE* **14**, e0223955 (2019).
23. Andrés, G., Simón-Mateo, C. & Viñuela, E. Assembly of African swine fever virus: role of polyprotein pp220. *J Virol* **71**, 2331–2341 (1997).

24. Salas, M. L. AFRICAN SWINE FEVER VIRUS (ASFARVIRIDAE). in *Encyclopedia of Virology (Second Edition)* (eds. Granoff, A. & Webster, R. G.) 30–38 (Elsevier, 1999). doi:10.1006/rwvi.1999.0008.
25. Dixon, L. K., Chapman, D. A. G., Netherton, C. L. & Upton, C. African swine fever virus replication and genomics. *Virus Research* **173**, 3–14 (2013).
26. Alejo, A., Matamoros, T., Guerra, M. & Andrés, G. A Proteomic Atlas of the African Swine Fever Virus Particle. *J Virol* **92**, e01293-18 (2018).
27. Yoo, D., Kim, H., Lee, J. Y. & Yoo, H. S. African swine fever: Etiology, epidemiological status in Korea, and perspective on control. *Journal of Veterinary Science* **21**, (2020).
28. Du, X., Gao, Z.-Q., Geng, Z., Dong, Y.-H. & Zhang, H. Structure and Biochemical Characteristics of the

Methyltransferase Domain of RNA Capping Enzyme from African Swine Fever Virus. *J Virol* **95**, e02029-20 (2021).

29. Salas, M. L., Kuznar, J. & Viñuela, E. Polyadenylation, methylation, and capping of the RNA synthesized in vitro by African swine fever virus. *Virology* **113**, 484–491 (1981).
30. Gaudreault, N. N., Madden, D. W., Wilson, W. C., Trujillo, J. D. & Richt, J. A. African Swine Fever Virus: An Emerging DNA Arbovirus. *Front. Vet. Sci.* **7**, 215 (2020).
31. Wang, Y. *et al.* Structure of African Swine Fever Virus and Associated Molecular Mechanisms Underlying Infection and Immunosuppression: A Review. *Front. Immunol.* **12**, 715582 (2021).

32. Cackett, G., Portugal, R., Matelska, D., Dixon, L. & Werner, F. African Swine Fever Virus and Host Response: Transcriptome Profiling of the Georgia 2007/1 Strain and Porcine Macrophages. *Journal of Virology* **96**, (2022).
33. Rodríguez, J. M. & Salas, M. L. African swine fever virus transcription. *Virus Research* **173**, 15–28 (2013).
34. Cackett, G. *et al.* The African Swine Fever Virus Transcriptome. *J Virol* **94**, e00119-20 (2020).
35. Pellett, P. E. *et al.* Order - Herpesvirales. in 99 (Elsevier, 2012). doi:10.1016/B978-0-12-384684-6.00005-7.
36. Xiang, S. *et al.* Complete Genome Sequence of a Variant Pseudorabies Virus Strain Isolated in Central

China. *Genome Announcements* **4**, 10.1128/genomea.00149-16 (2016).

37. Zheng, H.-H. *et al.* Seroprevalence investigation and genetic analysis of pseudorabies virus within pig populations in Henan province of China during 2018–2019. *Infection, Genetics and Evolution* **92**, 104835 (2021).
38. Zheng, H.-H., Fu, P.-F., Chen, H.-Y. & Wang, Z.-Y. Pseudorabies Virus: From Pathogenesis to Prevention Strategies. *Viruses* **14**, 1638 (2022).
39. Nauwynck, H., Glorieux, S., Favoreel, H. & Pensaert, M. Cell biological and molecular characteristics of pseudorabies virus infections in cell cultures and in pigs with emphasis on the respiratory tract. *Vet. Res.* **38**, 229–241 (2007).

40. Zhou, M. *et al.* Characterization of a moderately pathogenic pseudorabies virus variant isolated in China, 2014. *Infection, Genetics and Evolution* **68**, 161–171 (2019).

41. Klupp, B. G., Hengartner, C. J., Mettenleiter, T. C. & Enquist, L. W. Complete, Annotated Sequence of the Pseudorabies Virus Genome. *Journal of Virology* **78**, 424–440 (2004).

42. Pomeranz, L. E., Reynolds, A. E. & Hengartner, C. J. Molecular Biology of Pseudorabies Virus: Impact on Neurovirology and Veterinary Medicine. *Microbiol Mol Biol Rev* **69**, 462–500 (2005).

43. Fuchs, W., Ehrlich, C., Klupp, B. G. & Mettenleiter, T. C. Characterization of the replication origin (OriS) and adjoining parts of the inverted repeat sequences

of the pseudorabies virus genome. *Journal of General Virology* **81**, 1539–1543 (2000).

44. Klupp, B. G., Kern, H. & Mettenleiter, T. C. The virulence-determining genomic BamHI fragment 4 of pseudorabies virus contains genes corresponding to the UL15 (partial), UL18, UL19, UL20, and UL21 genes of herpes simplex virus and a putative origin of replication. *Virology* **191**, 900–908 (1992).
45. Mainguy, G., Koster, J., Woltering, J., Jansen, H. & Durston, A. Extensive Polycistronism and Antisense Transcription in the Mammalian Hox Clusters. *PLOS ONE* **2**, e356 (2007).
46. Tombácz, D. *et al.* Transcriptome-wide survey of pseudorabies virus using next- and third-generation sequencing platforms. *Sci Data* **5**, 180119 (2018).

47. Moldován, N. *et al.* Multi-Platform Sequencing Approach Reveals a Novel Transcriptome Profile in Pseudorabies Virus. *Frontiers in Microbiology* **8**, (2018).

48. Honess, R. W. & Roizman, B. Regulation of Herpesvirus Macromolecular Synthesis I. Cascade Regulation of the Synthesis of Three Groups of Viral Proteins. *Journal of Virology* **14**, 8–19 (1974).

49. Berthomme, H., Monahan, S. J., Parris, D. S., Jacquemont, B. & Epstein, A. L. Cloning, sequencing, and functional characterization of the two subunits of the pseudorabies virus DNA polymerase holoenzyme: evidence for specificity of interaction. *J Virol* **69**, 2811–2818 (1995).

50. Deng, J., Wu, Z., Liu, J., Ji, Q. & Ju, C. The Role of Latency-Associated Transcripts in the Latent

Infection of Pseudorabies Virus. *Viruses* **14**, 1379 (2022).

51. Kronstad, L. M., Brulois, K. F., Jung, J. U. & Glaunsinger, B. A. Reinitiation after translation of two upstream open reading frames (ORF) governs expression of the ORF35-37 Kaposi's sarcoma-associated herpesvirus polycistronic mRNA. *J Virol* **88**, 6512–6518 (2014).
52. Cackett, G., Sýkora, M. & Werner, F. Transcriptome view of a killer: African swine fever virus. *Biochemical Society Transactions* **48**, 1569–1581 (2020).
53. Torma, G. *et al.* An Integrated Sequencing Approach for Updating the Pseudorabies Virus Transcriptome. *Pathogens* **10**, 242 (2021).

54. Boldogkői, Z., Balázs, Z., Moldován, N., Prazsák, I. & Tombácz, D. Novel classes of replication-associated transcripts discovered in viruses. *RNA Biol* **16**, 166–175 (2019).

55. Rennekamp, A. J. & Lieberman, P. M. Initiation of Epstein-Barr virus lytic replication requires transcription and the formation of a stable RNA-DNA hybrid molecule at OriLyt. *J Virol* **85**, 2837–2850 (2011).

56. Boldogkoi, Z. Transcriptional interference networks coordinate the expression of functionally related genes clustered in the same genomic loci. *Frontiers in Genetics* **3**, (2012).