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 -

1. List of publications

1.1 Publications directly related to the subject of the thesis

I. 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**

II. 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**

III. 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**

1.2. Other related publications

IV. Tombácz D, **Torma G**, Gulyás G, Moldován N, Snyder M, Boldogkői Z. Meta-analytic approach for transcriptome profiling of herpes simplex virus type 1. Sci Data. 2020 Jul 9;7(1):223. doi: 10.1038/s41597-020-0558-8. **IF: 6,444**

V. Moldován N, **Torma G**, Gulyás G, Hornyák Á, Zádori Z, Jefferson VA, Csabai Z, Boldogkői M, Tombácz D, Meyer F, Boldogkői Z. Time-course profiling of bovine alphaherpesvirus 1.1 transcriptome using multiplatform sequencing. Sci Rep. 2020 Nov 24;10(1):20496. doi: 10.1038/s41598-020-77520-1. **IF: 4,38**

VI. Tombácz D, Moldován N, **Torma G**, Nagy T, Hornyák Á, Csabai Z, Gulyás G, Boldogkői M, Jefferson VA, Zádori Z, Meyer F, Boldogkői Z. Dynamic Transcriptome Sequencing of Bovine Alphaherpesvirus Type 1 and Host Cells Carried Out by a Multi-Technique Approach. Front Genet. 2021 Apr 7;12:619056. doi: 10.3389/fgene.2021.619056. **IF: 4,772**

VII. Maróti Z, Tombácz D, Prazsák I, Moldován N, Csabai Z, **Torma G**, Balázs Z, Kalmár T, Dénes B, Snyder M, Boldogkői Z. Time-course transcriptome analysis of host cell response to poxvirus infection using a dual long-read sequencing approach. BMC Res Notes. 2021 Jun 24;14(1):239. doi: 10.1186/s13104-021-05657-x. **IF: 0**

VIII. Maróti Z, Tombácz D, Moldován N, **Torma G**, Jefferson VA, Csabai Z, Gulyás G, Dörmő Á, Boldogkői M, Kalmár T, Meyer F, Boldogkői Z. Time course profiling of host cell response to herpesvirus infection using nanopore and synthetic long-read transcriptome sequencing. Sci Rep. 2021 Jul 9;11(1):14219. doi: 10.1038/s41598-021-93142-7. **IF: 4,997**

IX. Kakuk B, Tombácz D, Balázs Z, Moldován N, Csabai Z, **Torma G**, Megyeri K, Snyder M, Boldogkői Z. Combined nanopore and single-molecule real-time sequencing survey of human betaherpesvirus 5 transcriptome. Sci Rep. 2021 Jul 14;11(1):14487. doi: 10.1038/s41598-021-93593-y. **IF: 4,997**

X. Tombácz D, Prazsák I, **Torma G**, Csabai Z, Balázs Z, Moldován N, Dénes B, Snyder M, Boldogkői Z. Time-Course Transcriptome Profiling of a Poxvirus Using Long-Read Full-Length Assay. Pathogens. 2021 Jul 21;10(8):919. doi: 10.3390/pathogens10080919. **IF: 4,531**

XI. Kakuk B, Kiss AA, **Torma G**, Csabai Z, Prazsák I, Mizik M, Megyeri K, Tombácz D, Boldogkői Z. Nanopore Assay Reveals Cell-Type-Dependent Gene Expression of Vesicular Stomatitis Indiana Virus and Differential Host Cell Response. Pathogens. 2021 Sep 15;10(9):1196. doi: 10.3390/pathogens10091196. **IF: 4,531**

XII. Fülöp Á, **Torma G**, Moldován N, Szenthe K, Bánáti F, Almsarrhad IAA, Csabai Z, Tombácz D, Minárovits J, Boldogkői Z. Integrative profiling of Epstein-Barr virus transcriptome using a multiplatform approach. Virol J. 2022 Jan 6;19(1):7. doi: 10.1186/s12985-021-01734-6. **IF: 4,8**

XIII. Torma G, Tombácz D, Moldován N, Fülöp Á, Prazsák I, Csabai Z, Snyder M, Boldogkői Z. Dual isoform sequencing reveals complex transcriptomic and epitranscriptomic landscapes of a prototype baculovirus. Sci Rep. 2022 Jan 25;12(1):1291. doi: 10.1038/s41598-022-05457-8. **IF: 4,6**

XIV. Tombácz D, Kakuk B, **Torma G**, Csabai Z, Gulyás G, Tamás V, Zádori Z, Jefferson VA, Meyer F, Boldogkői Z. In-Depth Temporal Transcriptome Profiling of an Alphaherpesvirus Using Nanopore Sequencing. Viruses. 2022 Jun 13;14(6):1289. doi: 10.3390/v14061289. **IF: 4,7**

XV. Prazsák I, Csabai Z, **Torma G**, Papp H, Földes F, Kemenesi G, Jakab F, Gulyás G, Fülöp Á, Megyeri K, Dénes B, Boldogkői Z, Tombácz D. Transcriptome dataset of six human pathogen RNA viruses generated by nanopore sequencing. Data Brief. 2022 Jun 18;43:108386. doi: 10.1016/j.dib.2022.108386. **IF: 1,2**

XVI. Kakuk B, Dörmő Á, Csabai Z, Kemenesi G, Holoubek J, Růžek D, Prazsák I, Dani VÉ, Dénes B, **Torma G**, Jakab F, Tóth GE, Földes FV, Zana B, Lanszki Z, Harangozó Á, Fülöp Á, Gulyás G, Mizik M, Kiss AA, Tombácz D, Boldogkői Z. In-depth Temporal Transcriptome Profiling of Monkeypox and Host Cells using Nanopore Sequencing. Sci Data. 2023 May 9;10(1):262. doi: 10.1038/s41597-023-02149-4. **IF: 9,8**

XVII. Tombácz D, **Torma G**, Gulyás G, Fülöp Á, Dörmő Á, Prazsák I, Csabai Z, Mizik M, Hornyák Á, Zádori Z, Kakuk B, Boldogkői Z. Hybrid sequencing discloses unique aspects of the transcriptomic architecture in equid alphaherpesvirus 1. Heliyon. 2023 Jun 28;9(7):e17716. doi: 10.1016/j.heliyon.2023.e17716. **IF: 4**

XVIII. Torma G, Tombácz D, Csabai Z, Almsarrhad IAA, Nagy GÁ, Kakuk B, Gulyás G, Spires LM, Gupta I, Fülöp Á, Dörmő Á, Prazsák I, Mizik M, Dani VÉ, Csányi V, Harangozó Á, Zádori Z, Toth Z, Boldogkői Z. Identification of herpesvirus transcripts from genomic regions around the replication origins. Sci Rep. 2023 Sep 29;13(1):16395. doi: 10.1038/s41598-023-43344-y. **IF: 4,6**

Cumulative IF: 83,3

2. Table of contents

Abbrevations

ASFV: African Swine Fever Virus

AcMNPV: Autographa californica nucleopolyhedrovirus

asRNA: antisense RNA

AST: Antisense Transcripts

BoHV-1: Bovine alphaherpesvirus-1

Cage: Cap Analysis of Gene Expression

cDNA: copy DNA

CTO: Close to replication origin

dRNA: direct RNA

EBV: Epstein-Barr virus

EHV-1: Equine herpesvirus-1

EPM: Early promoter motif

E: Early

HCMV: Human betaherpesvirus 5

HSV-1: Herpes simplex virus type 1

I: Intermediate

Inr: Initiatior

IE: Immediate-early

IRS: Inverted repeat sequence

KSHV: Kaposi’s sarcoma-associated herpesvirus

L: Late

LAT: Latency-associated transcripts

LLT: Large latency transcript

LRS: Long Read Sequencing

lncRNA: long non-coding RNA

mRNA: messenger RNA

MGF: Multigene family gene

ncRNA: non-coding RNA

NGS: Next Generation of Sequencing

ONT: Oxford Nanopore Technologies (Ltd)

Ori: origin of replication

uORF: upstream ORF

PAMs: Porcine alveolar macrophage

PRV: Pseudorabies virus

PTO: proximal to origin (of replication)

RT: Reverse Transcriptase

raRNA: replication-associated RNAs

SRS: Short Read Sequencing

sncRNA: small non-coding RNA

TIN: Transcription Interference Network

TRS: Terminated repeat sequence

TRIN: Transcriptional Replication Interference Networks

TS: Template Switching

TSS: Transcription start site

TES: Transcription end site

UL: long unique

US: short unique

UTR: Untranslated Region (of a mRNA)

VSV: Vesicular Stomatitis Indiana

3. Introduction

The transcriptome is defined as all the RNA transcribed from an organism's genome. This includes both protein-coding and non-coding RNAs (ncRNAs), as well as transcripts with splice sites, as well alternative initiator and terminator sites[1]. Understanding them is important for clarifying the molecular causes of diseases, examining the functional elements of the genome, and comprehending the stages of individual development[2]. The main goal of transcriptomics is to map and quantify the characteristics of these molecules at different stages of development.

In recent decades, numerous methods have been developed to determine the sequence of DNA and RNA molecules in a biological sample. Initially, these investigations were carried out using Northern blots and real-time PCRs [3,4]. However, the major drawback of these approaches is that they are methods suitable for characterizing a single transcript. The breakthrough occurred in 1977 when Fredrick Sanger developed the chain-termination method, and an automated version was introduced in 1986[5]. In the 1990s, The Human Genome Project was launched with the goal of determining the complete sequence of the human genome, sparking a significant demand for high-throughput technologies[6]. This led to the emergence of three generations of sequencing technologies[7]. Next-generation sequencing platforms (NGS) appeared, offering a key advantage over classical Sanger sequencing in that they do not require bacterial cloning and electrophoretic separation. The techniques provide high throughput and can offer information on the entire genome. Two important innovations in this context are Short Read Sequencing (SRS) and Long Read Sequencing (LRS) technologies[8]. The high throughput provided by SRS has accelerated the understanding of the genomes of various organisms[9,10]. Current genome programs are based on SRS approaches. However, the read size of this technology is limited, which lead to the adoption of LRS in the field of transcriptomics, such as PacBio and Oxford Nanopore Technologies (ONT) MinION[11]. Although these platforms provide lower coverage, they enable the detection of RNA molecules of various lengths from end to end[12]. This information contributes to identifying transcript isoforms of individual genes, including various splice patterns, readthrough transcripts, 5' and 3' untrunslated region (UTR) variants, and polycistronic messenger RNA (mRNA) molecules transcribing multiple genes. These molecules are translated into proteins, as well as ncRNAs that influence gene expression and protein synthesis.

Due to their small, compact genomes, viruses are ideal organisms for transcriptome analyses [12]. As a result, in recent years, the number of known transcripts within them has multiplied, and a new possible genetic regulation based on long transcriptional overlaps has been revealed.

Understanding the function of these complex mRNA and non-coding RNA molecules within a virus-infected cell can contribute to advancing our understanding of their operations.

3.1 First-Generation Sequencing Technologies

The very first sequencing approaches were based on chemical cleavage or degradation of molecules[5]. Walter Gilbert and Allan Maxam developed a technique based on chemical degradation to sequence the DNA of a Phix174 bacteriophage[13]. In this experiment, DNA was labeled with radioactive 32 phosphate at their 5' end, then the bases were removed from the purines and pyrimidines with different chemical treatments (hydrazine and dimethyl sulfate). Subsequently, the phosphodiester bond was cleaved with piperidine, resulting in fragments of different sizes that were later separated by gel electrophoresis. In parallel, Fredrick Sanger developed a technique based on chain-termination. In this the reaction, radioactive or fluorescently labeled dideoxynucleotides are used, which stops the synthesis of the DNA molecule, resulting in smaller DNA fragments, which will be separated by capillary electrophoresis at the end of the reaction[5]. The automated commercial example of this was the Applied Biosystems ABI 370. The appearance of these first-generation sequencing techniques led to the appearance of second- and third-generation sequencers.

3.2 Second-Generation Sequencing Technologies

After the first-generation platforms, the second-generation technologies appeared, the first of these technologies was Roche's 454 in 2005[7,11]. It was a bioluminescence method based on synthesis. In the reaction, dNTPs were cyclically added, and the released pyrophosphate upon incorporation was detected. The Roche 454 had a significant advantage in long reads (~1 kb), but was somewhat limited by its low coverage. In addition, developments also began using other technologies, resulting in the release of the Ion PGM platform in 2010[14]. The principle of its operation is that when the polymerase incorporates a nucleotide into the DNA, a proton is released, which causes a pH change and this change in pH detected. PGM does not require fluorescence and optical detection, which makes it a platform that can be operated at a lower cost.

Illumina appeared on the market in 2007 and is currently the most widely used and defining platform in the field of genomics[9,10]. Its operation is based on Sequencing By Synthesis (SBS), which means that the molecule to be sequenced hybridizes to the oligonucleotides on the flow cell, which are amplified several times by bridge amplification to form clusters[11].

Fluorescently labeled dNTPs are used during sequencing, and their fluorescence is detected by a CCD camera upon incorporation. The great advantage of Illumina sequencers is the high number of reads ~30-100 million[15].

3.3. Third-Generation Sequencing technologies

Due to the short sequence reads, second generation technologies are not suitable for detecting long RNA molecules and their isoforms[16]. Currently, two major sequencing platforms, ONT minION and PacBio, are the most widely used for the detection of long RNA molecules. In the case of the Illumina platform, there is also the option to define endpoints, with the Cap Analysis of Gene Expression (CAGE) technique for 5' ends and poly(A) seq for 3' ends[17,18]. However, these alternatives, due to their short sequence lengths, cannot determine how the starting and ending points of an RNA molecule are combined[16].

PacBio is based on a nanosensor technology[19]. The template molecule to be sequenced is ligated to a hairpin-shaped adapter (SMRTbell), which creates a circular molecule. This molecule is loaded onto a SMRTcell, which has many thousands of picoliter holes (ZMVs=zero-mode waveguides)[20]. DNA polymerase is attached to the bottom of these ZMWs, which incorporates fluorescently labeled nucleotides during synthesis, which will be detected by an optical detector. The advantage of the technology is that one molecule can be sequenced several times, since the template is a circular molecule. Two platforms have also been released in recent years by PacBio, one is RSII which has 150,000 ZMW and the other is the latest Sequel which has 1,000,000 ZMW[13,21].

In the case of the ONT MinION, the DNA molecule to be sequenced passes through a nanopore[22]. These pores are alpha-hemolysins embedded in a lipid bilayer[23]. The pores are located in a flow cell, of which there are 512 in an MK1 flow cell[22]. During cDNA library preparation, a motor protein is ligated to the DNA, which helps the double-stranded molecule to unwind and translocate through the pore[7,13]. As the molecule passes through the pore, a characteristic voltage shift occurs. The duration and magnitude of this voltage shift, which is characteristic of a given DNA sequence, is recorded. The summary of the three currently most widespread NGS technologies is shown in Figure 1.

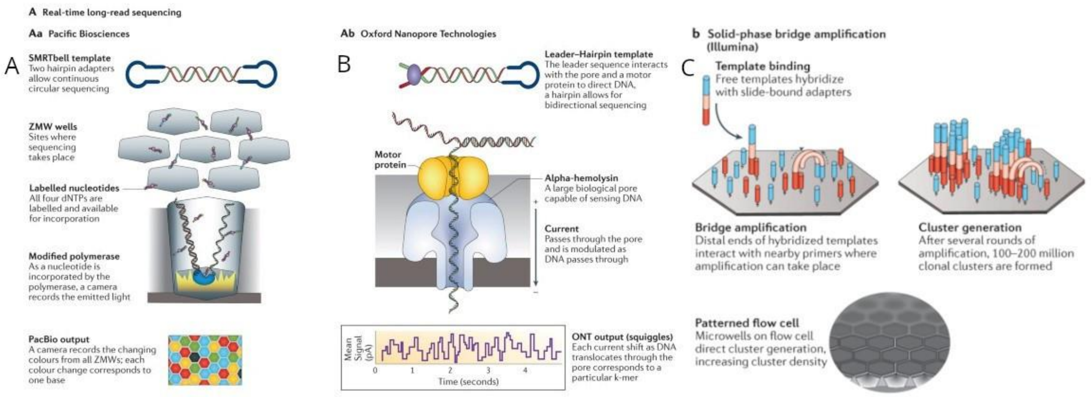

A Real-time long-read sequencing
Aa Pacific Biosciences
A
SMRTbell template
Two hairpin adapters allow continuous circular sequencing
ZMW wells
Sites where sequencing takes place
Labelled nucleotides
All four dNTPs are labelled and available for incorporation
Modified polymerase
As a nucleotide is incorporated by the polymerase, a camera records the emitted light
PacBio output
A camera records the changing colours from all ZMWs; each colour change corresponds to one base
Ab Oxford Nanopore Technologies
B
Leader–Hairpin template
The leader sequence interacts with the pore and a motor protein to direct DNA, a hairpin allows for bidirectional sequencing
Motor protein
Alpha-hemolysin
A large biological pore capable of sensing DNA
+
Current
Passes through the pore and is modulated as DNA passes through
–
Mean Signal (pA)
0
1
2
3
4
Time (seconds)
ONT output (squiggles)
Each current shift as DNA translocates through the pore corresponds to a particular k-mer
b Solid-phase bridge amplification (Illumina)
C
Template binding
Free templates hybridize with slide-bound adapters
Bridge amplification
Distal ends of hybridized templates interact with nearby primers where amplification can take place
Cluster generation
After several rounds of amplification, 100–200 million clonal clusters are formed
Patterned flow cell
Microwells on flow cell direct cluster generation, increasing cluster density

Figure 1. The three currently most important sequencing platforms in research [24]. A. PacBio B. ONT minION C. Illumina technical summary.

3.4 Direct RNA-Seq

During transcriptome sequencing, two significant problems may, leading to the identification of false isoforms[25]. One is the phenomenon of false priming, where the oligod(T) primer used for reverse transcription binds to an adenine-rich region, resulting in the creation of a truncated cDNA with an incorrect 3' end. Another significant issue is the phenomenon of template switching (TS), where the reverse transcriptase (RT) enzyme jumps to another template strand in a homology-dependent manner. This leads to the generation of transcripts containing false introns with non-canonical splice sites. The efficiency of TS is facilitated if the concentration of templates is high, homologous sequences are long or reverse transcription temperature is low[26,27].

To overcome these two errors, an effective alternative is provided by direct RNA (dRNA) sequencing developed based on nanopore technology, which sequences RNA directly[28]. This platform has several significant advantages, as it avoids errors from PCR amplification and reverse transcription. and it is also possible to detect 5 methyl cytosine and 6 methyl adenine modifications. Additionally, it allows the detection of modifications such as 5-methyl cytosine and 6-methyl adenine. One serious drawback of this technology is that it cannot precisely determine 15-30 nucleotides from the 5' ends of RNA[29,30]. This issue arises because the ratchet molecule releases RNA a few nucleotides (15-30) before the 5' end, causing the RNA end to quickly traverse the pore, producing a very weak signal that cannot be base-called. Another current major disadvantage of dRNA sequencing is its low throughput, as the sequencing speed is six times slower compared to cDNA-Seq[28].

3.5 The New Dimensions of Genetic Regulation in Viruses

Transcripts transcribed in viruses can create overlaps with each other[12,31]. Based on the positions of the genes, these overlaps can be convergent, divergent and parallel. The possible role of all these overlaps can be in gene regulation, that is, through the physical interaction of the transcription apparatuses, they can regulate and synchronize the kinetics of viral genes in space and time. This theory is called the transcription interference network (TIN).

Another interesting phenomenon can be produced by replication-associated RNAs (raRNAs) described near replication origins. Several such molecules have been detected by LRS in Herpesviruses previously[32,33]. These RNAs are assumed to regulate the initiation of replication and the orientation of the replication fork. The processes of DNA replication and transcription likely generate genome-wide interference, where these two processes tightly regulate each other at the genomic level. A particularly interesting aspect and evidence for the existence of this process may be that the genes surrounding Origin of replication (Ori) are regulatory genes involved in the initiation of replication and transcription. Moreover, it can not only play a role in these processes, but can also create DNA-RNA hybrids in the Ori regions of viruses. In Bk viruses, it has been observed that these raRNAs bind to sense and antisense DNA strands and prevent virus replication[34]. In another study published on the Epstein-Barr virus (EBV), it was demonstrated that an RNA called BHLF1, which partially overlaps with the origin of replication, can bind to DNA and thereby inhibit replication[35].

3.6 Analysis of viral transcriptomes using long-read sequencing

Our group has analyzed numerous human and animal viral transcriptomes, including: Human herpes viruses: Herpes simplex virus type 1 (HSV-1), Epstein-Barr virus (EBV), Human betaherpesvirus 5 (HCMV), Kaposi’s sarcoma-associated herpesvirus (KSHV)[30,36–38]. Animal herpes viruses: Pseudorabies virus (PRV), Bovine alphaherpesvirus-1 (BoHV-1), Equine herpesvirus-1 (EHV-1)[39–41]. Other animal pathogenic viruses: African Swine Fever Virus (ASFV), Autographa californica nucleopolyhedrovirus (AcMNPV), Vesicular Stomatitis Indiana (VSV), Vaccinia virus[42–45].

3.7 The African Swine Fever virus

African swine fever virus (ASFV) is a nucleo-cytoplasmic large DNA virus (NCLDV) belonging to the Asfarviridae family[46]. It infects pigs and wild boars, causing an acute fatal hemorrhagic disease. The vector of the virus spread is the soft ticks belonging to the Ornithrodos genus. It appeared in Kenya in the early 1900s and is still present in African countries[47 48]. Currently, 24 genotypes are known.[49] The genotypes are separated based on the C-terminal sequence of the B646L gene encoding the p72 capsid protein[50]. The I. and II. genotype causes world epidemics, which drastically reduced pig populations, causing enormous economic damage in the world[51].

ASFV has an icosahedral morphology with a diameter of 200 nm[52,53]. Its genome is linear double-stranded DNA with covalently closed terminal repeats at the ends. The A+T nucleotide ratio of the genome is 61-62%[54]. Their size varies between 170-190 kb and they may encode 151-167 genes [55]. The size of the genome and the number of its coding genes result from the change in the number of Multigene family genes encoded by the virus (MGF 100, 110, 300, 360, 505/530 and family p22), as well as the number of tandem repeats[54,56]. MGFs are localized to the ~40 kbp region of the left terminal and ~20 kbp region of the right terminal[57]. The genome structure of the virus is shown in Figure 2. Their replication takes place in the monocyte/macrophage cells of the host[58]. In addition, a BA71V strain adapted to and able to replicate in Vero cell lines was created for replication and in vitro laboratory studies[59,60]. The virion contains all the proteins necessary for replication, mRNA synthesis and post-transcriptional modifications, so viral replication occurs primarily in the cytoplasm, but also in the nucleus during the early phase of infection[55,61]. The virion contains an inner core, an outer and an inner lipid membrane, and an icosahedral capsid[62].

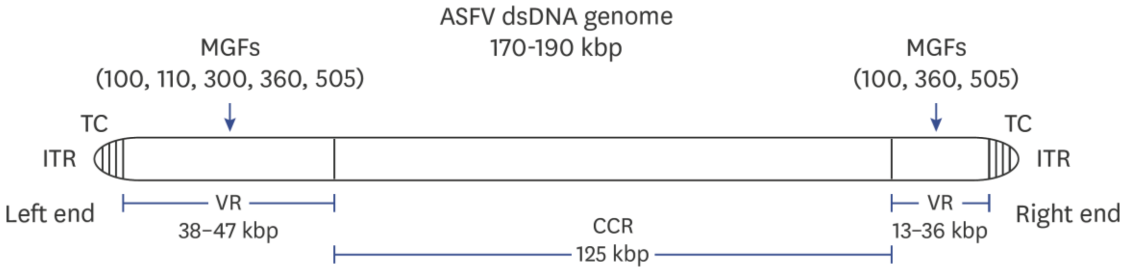


Figure 2. General structure of ASFV genome[63]. The ASFV genome contains inverted terminal repeats at the ends, a converged (CCR) and a variable region in the middle. The variable regions contain MGF genes, whose deletions and insertions cause differences in the number of ORFs and genome size between individual strains.

The virus is able to enter cells through primary macrophage receptors (CD163, CD45, MHCII)[64,65]. In addition, other mechanisms may play a role, so phagocytosis and macropinocytosis are also decisive[66–70]. After entering the cell, virus particles are transported to early endosomes, where they remain for 1-30 minutes, then into late endosomes, where they remain for 30-90 seconds[71,72]. During the transport through the endosomes, the pH of the medium gradually becomes acidic and at a pH below 5, the virus envelope is removed. After that, the inner envelope of the virus and the late endosomal membrane fuse, thereby releasing the viral DNA into the cytoplasm and allowing replication to occur. Replication will occur similarly to Pox viruses, as their genome structure is very similar and they contain an inverted and complementary hairpin loop at the ends of the genome[55]. A single-stranded nick is formed near these hairpin loops, which creates a free 3' OH group, which will serve as a primer for the synthesis of the DNA strand, which will be transcribed by DNA polymerase in the direction of the genomic ends. This continuous synthesis will lead to the formation of head-to-head concatemers, which will be fragmented into unit size by endonucleases and thus incorporated into the virion[55,73]. The synthesis of RNAs of the virus takes place in 4 stages: Immediate-early (IE), Early (E), Intermediate (I), Late (L)[74,75]. Pre-replicative genes have IE or E kinetics and are involved in replication, while postreplicative genes show I and L kinetics and encode structural proteins. Similar to poxviruses, ASFV is also characterized by temporal gene expression[76]. The transcription of early genes occurs 4-6 hours after infection[60,77]. Replication is initiated 6-8 hours after infection with the virus's own DNA polymerase, which is encoded on the G1211R gene. Expression of I and L genes occurs 8-16 hours after infection. Following transcription, mature virions are assembled and released. The complete life cycle of the virus is shown in Figure 3.

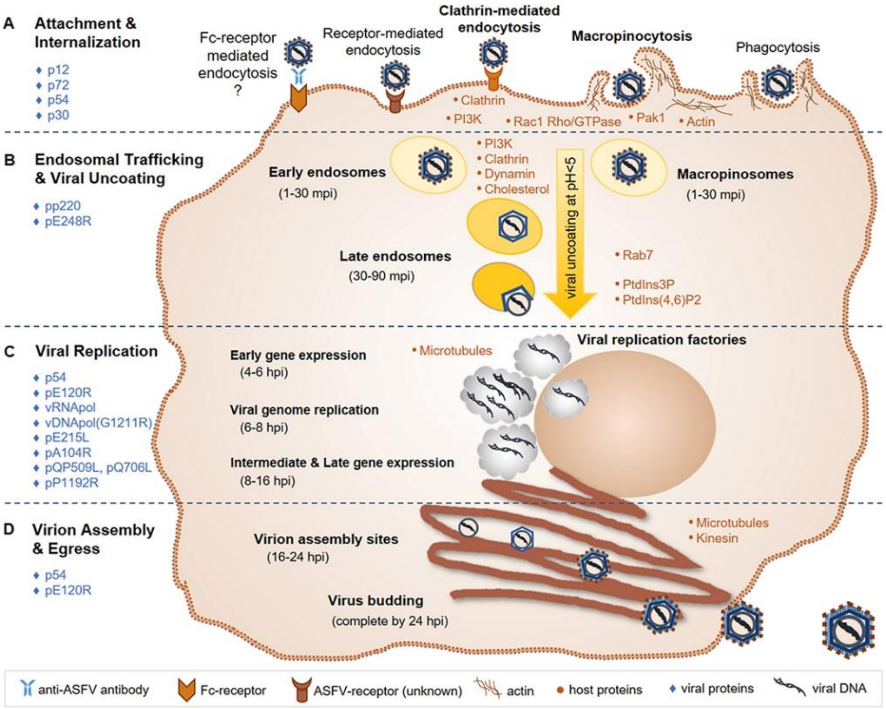

A Attachment & Internalization
♦ p12
♦ p72
♦ p54
♦ p30
Fc-receptor mediated endocytosis ?
Receptor-mediated endocytosis
Clathrin-mediated endocytosis
Macropinocytosis
Phagocytosis
• Clathrin
• PI3K
• Rac1 Rho/GTPase
• Pak1
• Actin
B Endosomal Trafficking & Viral Uncoating
♦ pp220
♦ pE248R
Early endosomes (1-30 mpi)
• PI3K
• Clathrin
• Dynamin
• Cholesterol
viral uncoating at pH<5
Macropinosomes (1-30 mpi)
Late endosomes (30-90 mpi)
• Rab7
• PtdIns3P
• PtdIns(4,6)P2
C Viral Replication
♦ p54
♦ pE120R
♦ vRNApol
♦ vDNApol(G1211R)
♦ pE215L
♦ pA104R
♦ pQP509L, pQ706L
♦ pP1192R
Early gene expression (4-6 hpi)
• Microtubules
Viral replication factories
Viral genome replication (6-8 hpi)
Intermediate & Late gene expression (8-16 hpi)
D Virion Assembly & Egress
♦ p54
♦ pE120R
Virion assembly sites (16-24 hpi)
• Microtubules
• Kinesin
Virus budding (complete by 24 hpi)
anti-ASFV antibody
Fc-receptor
ASFV-receptor (unknown)
actin
• host proteins
♦ viral proteins
viral DNA

Figure 3.The morphogenesis of ASFV from cell entry to the assembly of mature virions[60]. The virus can enter the cell in 5 ways: 1. Fc receptor mediated endocytosis, 2. Receptor mediated endocytosis 3. Clathrin mediated endocytosis 4 Macropinocytosis 5. Phagocytosis. After entering, they are transported to the early and late endosomes, where the increasing pH leads to the release of viral DNA. Then, through temporal gene expression, which is also characteristic of vaccinia viruses, the staged transcription of genes begins in 4 stages. Which means that the RNAs required for the expression of intermediate genes are expressed from the early genes, the RNAs required for the expression of late genes are expressed from the intermediate genes, while the RNAs required for the transcription of early genes are expressed from the late genes. Within 24 hours of virus infection, mature virions are released and reinfect.

The viral mRNAs have a 5' end cap and a 3' end poly(A)[78,79]. The poly(A) polymerase required for the synthesis of poly(A) is encoded by the C475 gene, while the enzyme that creates the cap is encoded by NP868R. The transcripts have UTR regions at their 5' and 3' ends, which are important in post-transcriptional regulation[74]. The promoters of early and late genes are characterized by different initiator (Inr) elements. In the case of the early ones, TA(+1)NA (where N can represent any nucleotide), while in the case of the late ones are characterized by the tetranucleotide sequence TA(+1)TA[74,80,81]. Early (EPM) and late promoter motifs (LPM) are located upstream of the initiator regions, which are AT nucleotide-rich and are highly conserved. EPM is located 9-10 nt, while LPM is located 4-6 nt from the transcription start sites. The transcription terminal sites are characterized by a Poly(U) of 6-7 nucleotides in length, for both early and late genes.

The virus transcriptome was previously analyzed by Cackett et al. They have used CAGE and poly(A) short read sequencing[80,80,81]. Studies on the host transcriptome were made to characterize virus-host interactions[58,82]. These studies were primarily carried out in the ASFV-GRG strains, where complete RNA-seq analysis of infected pig blood was performed, and in another study virus-infected macrophages were isolated for microarray analysis.

3.8 Pseudorabies virus

Pseudorabies virus (PRV) is a porcine Herpesvirus belonging to the Alphaherpesvirinae subfamily[83]. It first appeared in America in the early 1800s, and by the end of the 1980s it had spread throughout the world[84]. It causes a neurological disease in pigs and has a high mortality, which causes a significant problem for the pig industry[85,86]. It can also infect other animals, such as sheep, cats and mice[87,88]. It is found in the trigeminal ganglion of infected pigs[89]. Its virion consists of four elements: a central core, an icosahedral capsid, a tegument, and a surrounding lipid envelope containing glycoproteins[90].

It has a small, compact genome and does not infect humans, making it an ideal model organism for studying the pathogenicity of Herpes viruses[84]. It can also be used for tracking and mapping neurons[91–93]. PRV has also been used as a model organism to analyze the transcriptional regulation of genes, including TIN and the transcriptional replication interference networks (TRIN), which may represent a new aspect of gene regulation in viruses[31,32].

Its genome is double-stranded DNA, its size is approximately ~143 kb, the G+C content is relatively high at 74%, it contains at least 72 genes[90,94,95]. The genome can be divided into two segments, a long unique (UL) and a short unique (US) region[96]. The US region is bordered by Inverted and Terminated repeat sequences (IRS and TRS), the recombination of which can result in the formation of two types of isomers on the genome. In addition, two copies of the IE180 and US1 genes are found in the US region between IRS and TRS[84]. There are 3 origins of replication on the genome[97,98]. OriL is found in a single copy between the UL20 and UL21 genes in the UL region, while OriS is found in two copies in the Us region between the IE180 and US1 genes. All of them contain a GTTCGCAC motif, which is the binding site for the UL9 origin binding protein (OBP). The structure of the genome is shown in Figure 4. The majority of PRV genes are expressed in a polycistronic form, which is not common in eukaryotes[99]. PRV has 3 known splice sites, these appear on US1, UL15 and LLT transcripts[84]. For transcription, the host's transcription machinery will be used, but some transcription factors, such as IE180, US1 and EP0, will be expressed by the virus.

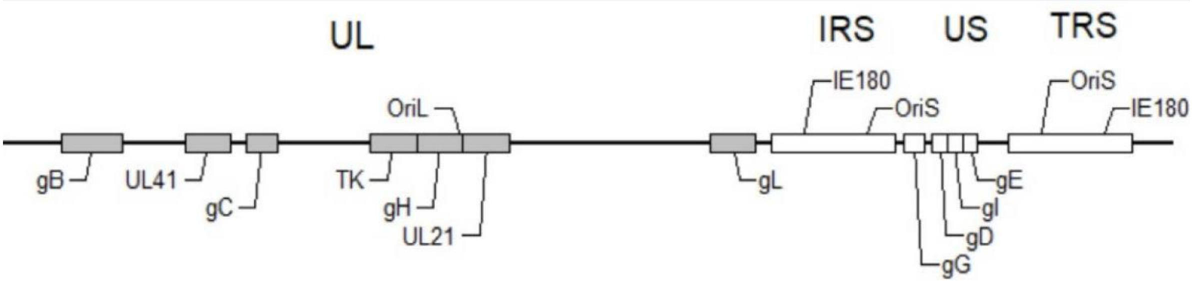


Figure 4.Structure of the PRV genome [90]. The genome can be divided into a Unique long (Ul) and a Unique short (Us) region. There are two repeat sequences in the Us region: IRS (Inverted repeat sequence) and TRS (Terminal repeat sequence). The Ul region contains one OriS, while the Us region contains two OriL origins of replication.

Glycoproteins play an important role in cell entry[84]. In cell attachment, glycoprotein C plays the initial role, binding to heparan sulfate. Subsequently, glycoprotein D, glycoprotein B, glycoprotein H, and glycoprotein L stabilize and facilitate virus-cell interaction and fusion. Following fusion, they associate with dynein to be transported from the cell periphery to the nucleus through microtubules. In the nucleus, PRV DNA is released, initiating virus replication and RNA transcription. The virus mRNAs will be expressed in a cascade manner in three steps: Immediate-Early (IE), Early (E), and Late (L) kinetic classes[100]. IE genes are the first to be expressed, producing transcriptional activators such as IE180, appearing within 40 minutes after infection and directly initiating the expression of PRV genes[84]. Subsequently, early gene expression begins, reaching its peak within 3-4 hours after infection. They are categorized into two groups: genes involved in replication (UL5, UL8, UL9/OBP, UL29/ssDNABP, UL30/DNAPol, UL42/Pap, and UL52) and those serving as transcription transactivators (EP0, US1, and UL54). Replication occurs through a rolling-circle mechanism[101]. Following early genes, late genes will be transcribed between 4-9 hours post-infection, encoding structural elements such as glycoproteins, tegument proteins, and the capsid[84,94]. The capsid proteins then move to the nucleus, where they assemble with the help of scaffold proteins UL26 and UL26.5. The mature capsid comprises 5 proteins: UL19/VP5, UL18/VP23, UL25, UL38, and UL35. Finally, the concatemer DNAs are fragmented into monomers and enter the newly synthesized mature capsid, in which the proteins UL32, UL33, UL15, and UL17 will play a role. The virus then enters the perinuclear space, where it loses its primary envelope and acquires the final envelope in the trans-Golgi, then leaves the cell in vesicles. The replication cycle of PRV is depicted in Figure 5.

After infection in the host, the viral genome remains present in the trigeminal ganglia[84,102]. These genomes are transcriptionally active and their reactivation enables further spread of the virus. The LAT (Latency-associated transcripts) gene is the only transcriptionally active region in latency. The IE180 protein can bind to LAT promoters, inhibiting LAT expression. LAT is located in an 11 kb region, covering the EP0 and IE180 genes with opposite polarities. RNAs of different sizes can be transcribed in this region: 0.95 kb, 1.0 kb, 2.0 kb, 8.0 kb, and 8.4 kb. The largest 8.4 kb is called large latency transcript (LLT). In order for the virus to remain present in the host, LAT will help neurons survive and will also have an anti-apoptotic effect.

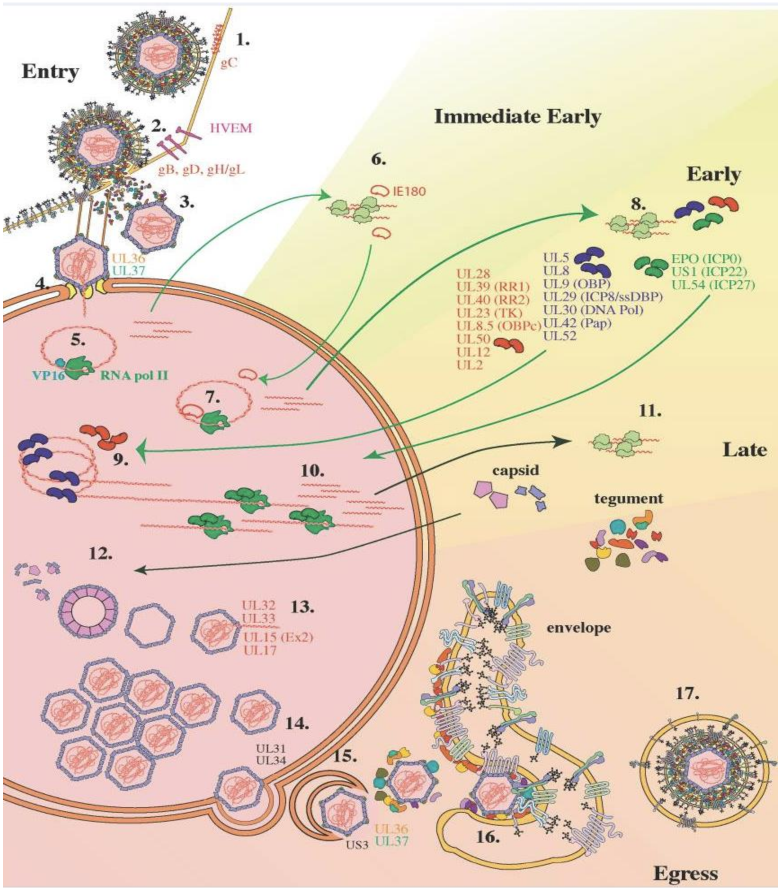


Figure 5.: PRV replication cycle [84]. 1. Viral glycoprotein C binds to the heparan sulfate of the cell membrane. 2. The binding of the virus to the cell membrane is stabilized by glycoprotein D, glycoprotein B, glycoprotein H and glycoprotein L. 3. The capsid and tegument proteins are released in the cell. 4. The capsid and tegument proteins enter the nucleus with the help of microtubules. 5.VP16 activates the transcription of the transactivator IE180. 6. The IE180 protein is synthesized in the cytoplasm and then transported back to the nucleus. 7. Transcription of early genes begins. 8. The RNAs synthesized from the early genes are transferred to the cytoplasm, where they are translated into proteins. 9. Early proteins are transported to the nucleus and 10. they will activate the transcription of late genes. 11. The RNAs of late genes enter the cytoplasm, where the structural proteins of the virus, such as the capsid and tegument proteins, are synthesized from them. 12. The capsid and tegument proteins migrate to the cell nucleus, where they are assembled with the help of scaffold proteins. 13. The mature capsid is formed, containing PRV DNA fragmented into monomers. 14. The virions leave the nucleus, then 15. lose their primary envelope and 16. enter the trans-Golgi apparatus, where they receive their final envelope. 17. The virus leaves the cell.

4. Aims

The aim of the work is to detect full-length RNAs and compile a transcriptomic atlas for PRV and ASFV.

Long-Read Sequencing has already been conducted for PRV[103,104]; however, in this project, we employed new sequencing chemistries compared to previous efforts. Specifically, we utilized direct RNA and direct cDNA sequencing to enhance the information obtained from the transcriptome. Additionally, in terms of bioinformatics methods, we replaced the old Albacore basecaller of ONT-minION with the newer and more advanced Guppy basecaller. Furthermore, we opted for the latest minimap2 instead of the GMAP mapping software to ensure the most precise analysis possible.

In the case of ASFV viruses, Cackett et al. have previously performed RNA-Seq using SRS Cage and poly(A) seq methods[80]; however, these are unable to annotate full-length transcripts and isoforms. We supplemented the previous results with ONT-minION PCR-amplified and non-amplified dcDNA and dRNA sequencing to reconstruct existing and newly discovered transcript endpoints into full-length RNAs. Cackett et al. reported experiments on the laboratory BA71V strain, but our experiments were conducted in the virus Porcine alveolar macrophage (PAMs) host cells.

Our investigations were aimed at the following:

1. Determination of the 5’ ends of mRNAs with base pair precision.
2. Determination of the 3’ ends of mRNAs with base pair precision.
3. Connecting annotated TSS and TES positions into transcripts.
4. Categorizing and determining the abundance of annotated transcript isoforms, polycistronic RNAs, ncRNAs, antisense RNAs, and 5’ truncated RNAs.
5. Detecting promoter elements (TATA box, CAAT box, GC box) and polyadenylation signals.

5. Materials and methods

5.1. PAM Cells, ASFV Viruses and Infection

Porcine alveolar macrophage (PAM) cells were freshly harvested from swine lungs according to the OIE Manual's instructions [105,106]. PAM cells were used for the propagation of the highly virulent ASFV_HU_2018 isolate of the African swine fever virus (Accession Number: MN715134.1). PAM cells were grown in RPMI 1640-containing L-glutamine (Lonza, Basel, Switzerland) medium supplemented with 10% fetal bovine serum (Euro Clone, Pero, Italy), 1% Na-pyruvate (Lonza), 1% non-essential amino acid solution (Lonza), and 1% antibiotic-antimycotic solution (Thermo Fisher Scientific, Waltham, MA, USA) at 37 °C in 5% CO2 in air gas phase. The infectious titer of serially diluted viral stock was calculated using an immunofluorescence (IF) assay as described earlier.[107] PAMs were cultivated in 6-well plates at a density of 3.3 × 105 cells and infected at a multiplicity of infection (MOI) of 10 plaque-forming units per cell at 4 h after cell seeding. The supernatant was replaced with a fresh medium after 1 h post-infection (hpi). Infected PAM cells were harvested at 4, 8, 12, and 20 hpi, whereas mock-infected cells were harvested at 20 hpi IF assay was used for monitoring the efficiency of infection in an infected control well fixed at 20 hpi.

5.2. Infection Efficiency

The length of the ASFV infection cycle length is approximately 18–22 h[108,109], yet virion production very often peaked around 72 hpi in PAM cells[110–112] since only a relatively low percentage of naïve PAM cells can be initially infected[113]. Since our intention was to characterize the dynamics of transcription, we harvested "first round" infected cells of two animals at 4, 8, 12 and 20 hpi. This sampling scheme allowed us to exclude a "second round" of infection in originally non-infected cells, which would lead to false conclusions about the kinetics of the viral transcripts. Indeed, infection efficiency remained at approximately 20% at 20 hpi despite the high viral titer (MOI = 10) applied for the infection.

5.3. PK-15 Cells and PRV Viruses

PK-15 porcine kidney epithelial cell line (ATCC® CCL-33™) was used for the propagation of strains Kaplan (PRV-Ka, Vanderbilt University, Nashville, TN, United States) and MdBio (PRV-MdBio, University of Szeged, Szeged, Hungary) of pseudorabies virus. Cells were cultivated in DMEM (Gibco/Thermo Fisher Scientific, Waltham, MA, United States), supplemented with 5% fetal bovine serum (Gibco/Thermo Fisher Scientific, Waltham, Massachusetts) and 80 μg of gentamycin per ml (Gibco/Thermo Fisher Scientific) at 37 °C in the presence of 5% CO_2. For the preparation of virus stock solution, cells were infected with 0.1 MOI. Viral infection was allowed to progress until complete cytopathic effect was observed. It was followed by three successive cycles of freezing and thawing of infected cells in order to release of viruses from the cells. In all of the previous experiments of which data were used here, PRV-Ka was grown in PK-15 cells using the same cultivation conditions. In this work, we exclusively used lytically infected PK-15 cells for the analyses. PRV-MdBio was used for the direct cDNA sequencing, while we used PRV-Ka for the rest of the experiments. PK-15 cells were infected with 10 MOI of PRV-Ka or 1 MOI of PRV-MdBio. Infected cells were incubated for 1 h at 37 °C, followed by removal of the virus suspension and washing the cells with phosphate-buffered saline. Following the addition of new culture medium, the cells were incubated for 1, 2, 4, 6, 8, or 12 h. After the incubation, the culture medium was removed, and the infected cells were frozen at −80 °C until further use. Except the non-amplified PacBio RSII sequencing [114], where we analyzed each time point separately, in the other experiments, we used mixed time point samples for the sequencing equal volumes from each sample was mixed.

5.4. RNA Purification

5.4.1 Extraction of Total RNA

We used the NucleoSpin® RNA (Macherey–Nagel, Düren, Germany) kit for isolation of the total RNA from the samples as was described in our previous publications[114]. Briefly, samples were incubated with a lysis buffer (supplied by the kit), then DNase I treatment was carried out. Purified RNA samples were eluted from the silica membrane in nuclease-free water. Samples were stored at −80 °C until further use. The total RNAs were used directly for the “amplified cDNA protocol” from ONT.

5.4.2. Purification of Polyadenylated RNAs

For the direct RNA (dRNA) and direct cDNA (dcDNA) sequencing approaches, the poly(A)+ fraction of the total RNAs were extracted. This process was carried out with Qiagen’s Oligotex mRNA mini kit using spin columns according to the kit’s manual.

5.4.3. Removal of the Ribosomal RNAs

The RiboMinus™ Eukaryote System v2 (Thermo Fisher Scientific) was used to obtain rRNA-free RNA samples, which is required by the applied Illumina library preparation approach. For explanation, the NEXTFLEX® rapid directional RNA-Seq Kit 2.0 was used for the preparation of strand-specific single-end or paired-end RNA libraries as recommended by the manufacturer. The following steps were followed: RNA fragmentation, first-strand synthesis, second-strand synthesis, adenylation, adapter ligation, and PCR amplification. The ribodepletion was carried out according to the kit's instructions.

5.5. Pacific Biosciences Isoform Sequencing Using the Sequel System

5.5.1. Synthesis of cDNAs

The cDNAs were generated from the Poly(A)+ RNA samples using the Clontech SMARTer PCR cDNA Synthesis Kit according to the PacBio Isoform Sequencing (Iso-Seq) protocol without size selection. The first-strand cDNAs were generated from the Poly(A)+ RNA with oligo(dT) primers [3′ SMART® CDS Primer II A (12 μM) part of the Clontech Kit]. They were incubated at 72 °C for 3 min with slow ramp to 42 °C at 0.1 °C/s and held at 42 °C for 2 min. 5× First-strand Buffer, DTT (100 mM), dNTP (10 mM), SMARTer II A Oligonucleotide (12 μM), RNase Inhibitor, and SMARTScribe Reverse Transcriptase (100 U) were mixed and heated to 42 °C for 1 min and then was measured into the RNA containing tube. This sample was incubated at 42 °C for 90 min, and finally the reaction was terminated at 70 °C for 10 min. These samples were amplified using KAPA HiFi Enzyme (Kapa Biosystems, Wilmington, MA, United States), according to the PacBio's recommendations (detaild in a previous publication: [50]. In summary, KAPA HiFi Fidelity Buffer (5×), KAPA dNTP Mix (10 mM), 5′ PCR Primer II A (12 μM) and KAPA HiFi Enzyme (1U/μL) were added to the first-strand cDNAs, and PCR reaction was carried out according to the following settings: Cycle the reaction with the following conditions (using a heated lid): The initial denaturation was at 95 °C for 2 min, and then 16 cycles at 98 °C for 20 s, 65 °C for 15 s and 72 °C for 4 min. The final extension was done at 72 °C for 5 min.

5.5.2. SMRTbell Template Preparation for PacBio Sequel Sequencing

About 500 ng from the amplified cDNA sample was used to prepare the SMRTbell library using the Clontech SMARTer PCR cDNA Synthesis Kit (Mountain View, CA, United States) based on the PacBio Isoform Sequencing (Iso-Seq) protocol, according to our earlier publication[114]. Briefly, the cDNA damages were repaired by the addition of DNA Damage Repair Buffer, NAD+ (1mM final concentration), ATP high (1 mM final concentration), dNTP (0.01 mM final concentration), and DNA Damage Repair Mix 37 °C for 20 min (both from the PacBio Template Prep Kit, PacBio, Menlo Park, CA, United States). This was followed by the repairing of the cDNA ends by using the End Repair Mix (PacBio Template Prep Kit). The adapters were ligated to the cDNA samples with ligase enzyme (0.75 unit/μL) and ATP low was also added (0.05 mM final concentration) at 25 °C for 15 min. Finally, the exonuclease treatment was carried out (with ExoIII and ExoVII enzymes from the Template Prep Kit at 37

°C for 1 h) in order to remove the incorrect SMRTbell templates (e.g., with free ends that did not receive an adapter, or contain nicks or other damage) from the library leaving only intact SMRTbell templates. AMPure® PB bead purification steps were performed after each of the enzymatic steps. The SMRTbell library was bound to the P6 DNA polymerase (Pacific Biosciences, Menlo Park, CA, United States) and annealed to v2 primers (PacBio, Menlo Park, CA, United States), then this library-polymerase complex was bound to MagBeads with MagBead Binding Kit (PacBio, Menlo Park, CA, United States). The total amount of the MagBead-bound complex was loaded onto the SMRT Cell. The MagBead One Cell Per Well protocol was used. One SMRT Cell was run on the Sequel platform.

5.5.3. PacBio RSII Long-Read Sequencing

The PacBio RSII instrument was used for sequencing the non-amplified [104] and amplified [103] cDNA libraries. In short, for the non-amplified method, the SuperScript Double-Stranded cDNA Synthesis Kit with SuperScript III reverse transcriptase and Anchored Oligo(dT)20 primers (both from Life Technologies, Carlsbad, CA, USA) were utilized to generate cDNAs from the poly(A)+ RNA samples. The PacBio's DNA Template Prep Kit 1.0 following the Pacific Biosciences' 2 kb Template Preparation and Sequencing protocol was followed as we described at the Sequel sequencing section, however P5 polymerase was used.

5.6. Oxford Nanopore Technologies Nanopore Sequencing Using the MinION Device

5.6.1. Direct RNA Sequencing

The Direct RNA sequencing (SQK-RNA002) protocol (Version: DRS_9080_v2_revM_14Aug2019) was used to obtain amplification-free transcriptomic data to remove RT and PCR biases, as well as to explore attributes of native RNA such as modified bases. Five hundred nanograms of Poly(A)+-tailed RNA was used. The library preparation was carried out according our previous publication[114] with the following modification: Agencourt RNAClean XP beads (Beckman Coulter, Brea, CA, United States) was used instead of the RNase OUT (Invitrogen)-treated Agencourt XP beads (Beckman Coulter, Brea, CA, USA).

5.6.2. Direct cDNA Sequencing

Non-amplified cDNA libraries were prepared from the poly(A)+ fraction of RNAs from the MdBio strain using the ONT's Direct cDNA Sequencing Kit (SQK-DCS109; DCS_9090_v109_revJ_14Aug2019, Oxford Nanopore Technologies, Oxford, United Kingdom) according to the manufacturer's protocol. In brief, the Maxima H Minus Reverse Transcriptase (Thermo Fisher Scientific, Waltham, MA, United States) with SSP and VN primers (supplied in the kit) were used for the synthesis of first cDNA strand from 100 ng of poly(A)+ RNA. Next, the potential RNA contamination was eliminated using RNase Cocktail Enzyme Mix (Thermo Fisher Scientific, Waltham, MA, United States). This step was followed by the second strand synthesis using LongAmp Taq Master Mix (New England Biolabs, Ipswich, MA, United States). Double-stranded cDNA ends were repaired using NEBNext End repair/dA-tailing Module (New England Biolabs, Ipswich, MA, United States), then the sequencing adapter ligation was carried out with the NEB Blunt/TA Ligase Master Mix (New England Biolabs).

5.6.3. Amplified cDNA Sequencing

ONT's ligation-based sequencing protocol (SQK-LSK109; Version: GDE_9063_v109_revU_14Aug2019). The ONT's LSK109 protocol was used for sequencing the Poly(A)-selected oligo(dT)-primed, rRNA-depleted random-primed, or TerminatorTM-handled oligo(dT)-primed samples. The usefulness of the application of TerminatorTM enzyme is that it enriches the capped full-length transcripts, because this enzyme processively digests the RNA molecules with 5′-monophosphate ends but not with 5′-triphosphate, 5′-cap or 5′-hydroxyl groups.

The generation of cDNA was conducted according to our previous publications[103,114] using oligo(dT) or random primers. The DNA repair was carried out according to the SQK-LSK109 protocol. Briefly, the NEBNext FFPE DNA Repair Mix and NEBNext Ultra II End repair/dA-tailing Module reagents (all from New England Biolabs, Ipswich, MA, United States) were mixed with the samples, then the mixtures were incubated at 20 °C for 5 min and at 65 °C for 5 min. This step was followed by the adapter ligation: The NEBNext Quick T4 DNA Ligase (New England Biolabs), the Ligation Buffer, and Adapter Mix (both from ONT's Kit, ONT, Oxford, United Kingdom) were mixed with the cDNA samples and incubated for 10 min at room temperature. Samples were purified using the AMPure XP magnetic beads (Beckman Coulter, Brea, CA, United States) after each enzymatic step. 1D Strand switching cDNA by ligation method (Version: SSE_9011_v108_revS_18Oct2016) and the ONT Ligation Sequencing Kit 1D (SQK-LSK108) (ONT, Oxford, United Kingdom) This protocol was used to analyze the random primed cDNA libraries. In short, ribodepleted RNA fraction was used to generate cDNA samples, first it was mixed with dNTPs (10 mM, Thermo Scientific) and random primers (ordered from IDT DNA) and then the mixtures were incubated at 65 °C for 5 min. After this step, the DTT and buffer form the SuperScipt IV Reverse Transcriptase kit (Life Technologies), RNase OUT enzyme (Life Technologies), and strand-switching oligo with three

O-methyl-guanine RNA bases (PCR_Sw_mod_3G; Bio Basic, Canada) were added and the mixtures were heated to 42 °C for 2 min. SuperScript IV Reverse Transcriptase enzyme (200 unit) was mixed into the samples. The generation of the first cDNA strand was conducted at 50 °C for 10 min, then the strand switching step was carried out at 42 °C for 10 min. For the inactivation of the enzymes, the samples were heated to 80 °C for 10min. Samples were amplified using the KAPA HiFi DNA Polymerase (Kapa Biosystems, Wilmington, MA, USA) and Ligation Sequencing Kit Primer Mix (provided by the 1D Kit, ONT, Oxford, UK). The NEBNext End repair/dA-tailing Module (New England Biolabs, Ipswich, MA, USA) was applied to repair cDNA ends, while NEB Blunt/TA Ligase Master Mix (New England Biolabs, Ipswich, MA, USA) was used to ligate the adapters (supplied by the kit, New England Biolabs, Ipswich, MA, USA).

5.6.4. Amplified cDNA-Sequencing Using MinION Device

Total RNA samples were sequenced using the ONT MinION device and the cDNA-PCR Barcoding protocol (SQK-PCS109 and SQK-PBK004). RT was carried out as described in the dcDNA protocol (above). The samples were amplified using the LongAmp Taq master mix. Low-input barcode primers (ONT's SQK-PBK004 kit) were added to the samples with the aim of multiplexing the samples on the Flow Cells. After the PCR reaction, cDNAs were treated with an exonuclease, and they were finally cleaned with AMPure XP beads, as it was explained in the dcDNA paragraph.

5.6.5. ONT Long-Read Sequencing

1D Strand switching cDNA by ligation method (Version: SSE_9011_v108_revS_18Oct2016, ONT, Oxford, USA) and the ONT Ligation Sequencing Kit 1D (SQK-LSK108, ONT, Oxford, USA) were used for sequencing the Poly(A)+ transcriptome of the virus[103], as we described at the 'Amplified cDNA sequencing' section's random-primed paragraph.

Cap-selection: The Lexogen's TeloPrime Full-Length cDNA Amplification Kit was used to generate cDNAs (and therefore sequencing libraries) only from the capped RNAs. This method works with specific double-stranded adapters. It is required by the second strand synthesis and the adapters only ligates to the cDNAs if the inverted Gs of the cap structure are present. The cDNAs were generated from total RNA samples using the TeloPrime Kit's buffer, reverse transcription primer, and enzyme at 46 °C for 50 min. After this step, the adapter was ligated to the cDNA in the hybrid by base-pairing of the 5′ C to the cap structure of the RNA, using a double-strand specific ligase (from the Kit, Lexogen, Vienna, Austria). The ligation was performed out at 25 °C, overnight. The second strand synthesis of the cDNA samples was carried out with the Second Strand, the PCR forward primer, and the enzyme mix (both from the Kit, Lexogen, Vienna, Austria) according to the following program: 1 cycle of 90 s at 95.8 °C, 60 s at 62 °C, 5 min at 72 °C. Finally, the double-stranded cDNAs were amplified by PCR. TeloPrime PCR Mix, PCR Forward Primer, PCR Reverse Primer, and Enzyme Mix were added. Sixteen PCR cycles of thermocycling with the following program was carried out: 1 cycle of 95.8 °C for 30 s, 50 °C for 45 secs, 72 °C for 20 min, then 15 cycles of 95.8 °C for 30 s, 62 °C for 30 s, 72 °C for 20 min, and a final extension at 72 °C for 20 min. Library generation was carried out with the ONT Ligation Sequencing Kit 1D (SQK-LSK108, ONT, Oxford, UK) as we described above, at the "1D Strand switching cDNA by ligation method" and earlier[103].

5.7. Measurement of Nucleic Acid Quality and Quantity

5.7.1. RNA

The Qubit RNA BR Assay Kit (Invitrogen, Carlsbad, CA, United States) was used for the total RNA measurement, while the Qubit RNA HS Assay Kit (Invitrogen, Carlsbad, CA, USA) was applied to check the quantity of the poly(A)+ and rRNA-depleted RNA fractions. The final concentrations of the RNA samples were determined by Qubit® 4.

5.8. Pre-Processing and Data Analysis

The processing of the MinION raw data was conducted with the Guppy basecaller v. 3.6.1. with--qscore_filtering. Reads with a Q-score greater than 7 were aligned to the viral genome (NCBI nucleotide accession PRV: KJ717942.1[115] and ASFV: MN715134.1 using the Minimap2 mapper[116]. The PacBio dataset was also mapped with Minimap2 to the KJ717942.1 reference genome. Transcription reads were visualized using Genious 11.1.5 software (**www.geneious.com**, (**accessed on 1 December 2020**)). The upset plot was visualized by the UpSetR program[117].

Our in-house scripts (SeqTools, Department of Medical Biology, University of Szeged, Szeged, Hungary) were used to generate the descriptive quality statistics of reads (ReadStatistics, Department of Medical Biology, University of Szeged, Szeged, Hungary) and to analyze promoters (MotifFinder, Department of Medical Biology, University of Szeged, Szeged, Hungary), are available on GitHub:

https://github.com/moldovannorbert/seqtools (accessed on 4 June 2020).

In this study, the LoRTIA (**https://github.com/zsolt-balazs/LoRTIA** (accessed on 20 August 2019)) software package (v.0.9.9, Department of Medical Biology, University of

Szeged, Szeged, Hungary) was used for the detection and annotation of transcripts and transcript isoforms, as described earlier. Briefly, the sequencing adapters and the homopolymer A sequences were checked by the LoRTIA toolkit for the identification of TSS and TES, respectively. For the elimination of spurious TSSs and TESs (which can be caused by RNA degradation, incomplete reverse transcription and PCR, or template switching), the putative start and end sites were tested against the Poisson distribution (using Bonferroni correction). Putative introns were accepted applying the following criteria: (1) They had one of the three most frequent splice consensus sequences (GT/AG, GC/AG, AT/AC) and (2) their abundance exceeded 1‰ compared to the local coverage.

The accepted putative TSSs and TESs were considered as existing if they were identified by the LoRTIA in the datasets obtained by at least two different techniques. Potential introns were accepted as real if they were present in both dRNA-Seq and at least one of the cDNA-Seq datasets, and if they were shorter than 10 Kbps. The accepted TSSs, TESs, and introns were then assembled into putative transcripts using the Transcript_Annotator software of the LoRTIA suite. Very long unique or low-abundance reads that could not be detected using LoRTIA were annotated manually. These reads were also accepted as putative transcript isoforms if they were longer than any other overlapping RNA molecule. In some of these cases, the exact TSSs were not annotated. Finally, a read was considered as a transcript if it was present in at least three different samples. Transcript annotation was followed by isoform categorization according to the following principles: the most abundant transcript containing a single ORF was termed as the canonical monocistronic transcript, whereas isoforms with longer or shorter 5′-UTRs or 3′-UTRs regions than the canonical transcripts were termed TSS or TES isoforms (variants), respectively. Similarly, transcripts with alternative splicing were named splice isoforms. Transcripts with 5′-truncated in-frame ORF (open reading frame) were termed as putative mRNAs. Transcripts containing multiple tandem non-overlapping ORFs were designated polycistronic, whereas those containing at least one ORF with opposite orientation were called complex transcripts. Transcripts with no ORFs or ORFs shorter than 30 nts were considered non-coding.

6. Results - Analysis of African Swine Fever virus (ASFV) transcriptome

6.1 Sequencing and mapping statistics

In our analysis, we utilized the ONT MinION LRS and Illumina MiSeq SRS platforms to characterize the viral transcriptome. For ONT MinION sequencing, oligo(d)T primers were employed, while random primers were used in the reverse transcription process during Illumina experiments to synthesize the first strand of cDNA. Three technical approaches were applied during ONT MinION sequencing, which were as follows: 1; PCR-amplified cDNA, 2; PCR amplification-free direct cDNA sequencing (dcDNA), 3; native direct RNA sequencing. In Illumina sequencing, a total of 69,068 reads were generated with an average coverage of 50.13 on the ASFV genome. For PCR-amplified ONT cDNA-Seq, 126,763 reads were generated with an average length of 458 bps. ONT direct cDNA sequencing resulted in 8,587 RNA reads with an average length of 568 bps. ONT dRNA sequencing produced 4,361 viral reads with an average length of 637 bps. The detailed figures of the readings can be seen in Table 1 and Figure 6.

The transcripts annotated and filtered by the LoRTIA software package had an average size of 900.45 bps.

Sam-ples	Time point	Library prep ap-proach	Read count	Mapped read count	Mean mapped read length
A	4h	amplified cDNA	1,130,020	18.517	454.5
A	8h	amplified cDNA	548.671	9.245	499.9
A	12h	amplified cDNA	1,230,079	12.779	489.9
A	20h	amplified cDNA	2,759,917	33.189	405.9
B	4h	amplified cDNA	1,562,973	17.668	462.3
B	8h	amplified cDNA	978.909	13.346	488.5
B	12h	amplified cDNA	1,861,829	13.141	461.6
B	20h	amplified cDNA	1,567,961	8.878	400.3
A & B	mixed	dcDNA	3,591,427	8.587	568.2
A & B	mixed	dRNA	1,891,036	4.361	636.9

Table 1. Mapping, Sequencing and read length statistics of ASFV virus.

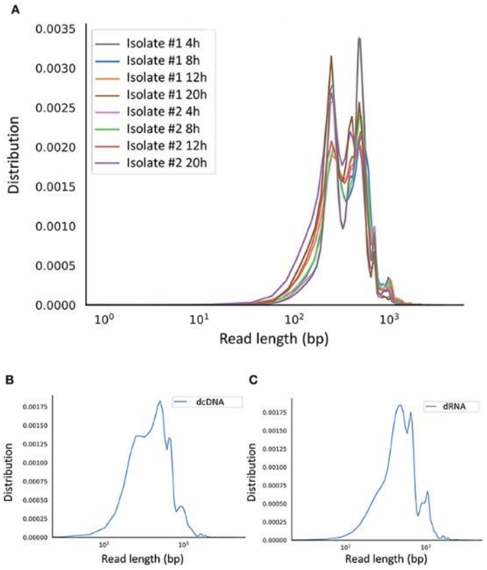


FIGURE 6. Aligned read length distribution. Line chart presentation of the average of aligned read lengths obtained via Nanopore sequencing. (A) Amplified cDNA sequencing at various time points. (B) Direct cDNA sequencing and (C) direct RNA sequencing using samples from multiple time points after the infection.

6.2 Annotating the viral transcriptome

This dataset has been published previously[106]. False priming and template switching artifacts resulting from reverse transcription and PCR amplification were filtered out using the LoRTIA software package [118]. In our study, we applied stringent conditions, accepting and annotating only those transcript ends that were present in at least two independent biological replicates. For ncRNAs and 5' truncated RNAs, in addition to the two biological replicates, the transcript ends had to be validated by dcDNA and dRNA experiments. For some genes with low coverage, milder criteria were applied for transcript annotation.

During our investigation, annotations were made for 132 LoRTIA and 70 non-LoRTIA transcription start sites (TSSs), as well as 137 LoRTIA and 83 non-LoRTIA transcription end sites (TESs). Among these, 98 TSSs and 57 TESs were validated by experiments conducted by Cackett and colleagues using CAGE and poly(A)-Seq.[74] In total, 311 viral mRNA molecules were identified in our experiments with LRS (Figure 8). Out of these mRNAs, the TSS of 273 molecules was precisely determined. However, in the case of 38 molecules, we were not able to determine the exact TSS positions due to 15-30 nt missing from the 5' of the dRNA. The length of 5' UTRs was 69.01 bp for TSSs, while the length of 3' UTRs was 369.31 bp for TESs.

6.3 Promoter motifs. poly(A) signal,

117 TATA boxes were identified with an average distance of 54.45 bps from the TSSs. Additionally, 3 GC boxes were found with an average distance of 9 bps from the TSSs, and 9 CAAT boxes with an average distance of 112.41 bps from the TSSs. Figure 7A shows the frequency of TSSs initiator sequences, and Figure 7C depicts the frequency of individual TATA motifs. Typically, TA+(1)NA sequences are the most common among TSS initiators, and the TA(+1)TA, characteristic of late transcripts, is also present, appearing in 20 TESs.

We determined 94 poly(A) signal with an average distance of 31,86 bps from the TESs. In the case of transcription end points, a 6 nt long poly(U) sequence was observed. Figure 7B shows the sequence environment of TESs, and Figure 7D depicts the frequency of poly(A) signal sequences.

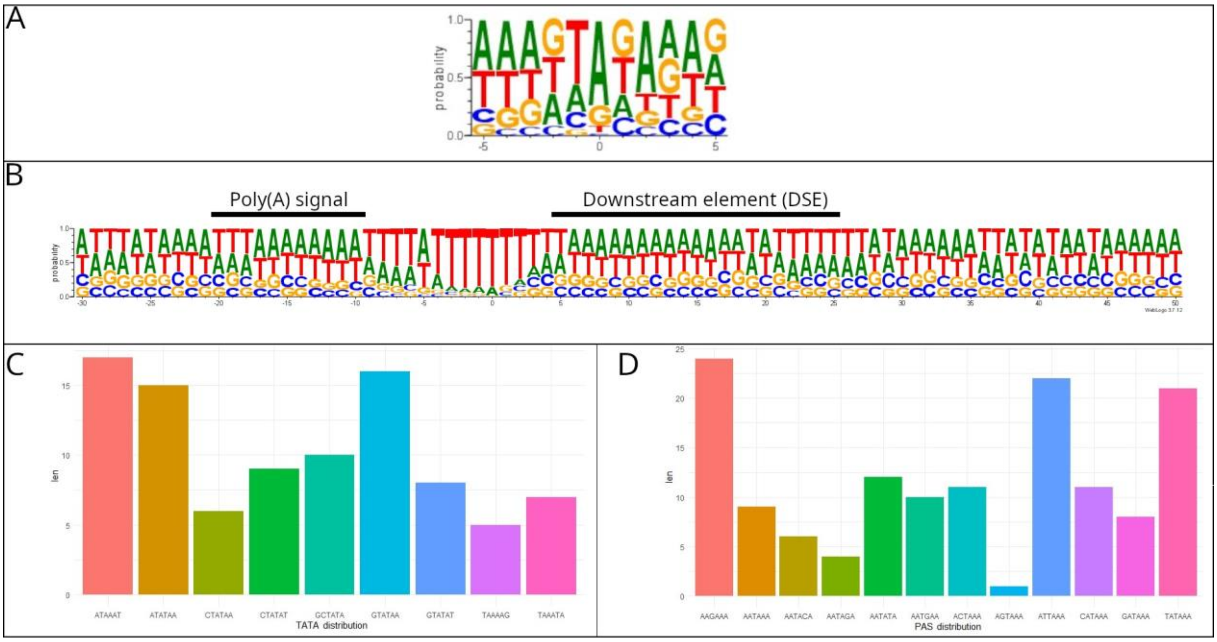


Figure 7. Initiators and polyadenylation signals of ASFV. (A) Genomic surrounding of TSSs within a ±5 bp interval. The first letter of TSSs (position 0) is enriched with A bases, while the − 1 position contains mainly T bases. (B) Sequence motifs of transcripts containing polyadenylation signals. (C) Distribution of TSS containing TATA promoter motifs. (D) Distribution of TESs containing poly(A) signal motifs.

6.4 Transcriptional start and end sites isoform

The TSS variants have the same ORF as canonical mRNAs; however, there are differences in the length of their 5' UTRs. The quantity of TSS isoforms showed less diversity in this analysis, which could be attributed to the lower coverage of LRS and the stringent criteria applied for the acceptance of TSSs and TESs. Using the LoRTIA software package, a total of 16 TSS isoforms were identified, of which 14 were longer and 2 were shorter compared to canonical mRNA molecules. From the 8 TSS isoforms reported by Cackett and colleagues in the Cage-Seq data, we could confirm one (MGF 300-2R); however, from their included CAGE-Seq non-primary dataset, we validated an additional 2 TSS variants (D345L, C147L)[74]. We detected a TATA box in the case of 12 TSS isoforms with an average distance of 51 nucleotides. Among 16 TESs, 5 of them possess a late promoter motif (MGF110-4L, MGF 300-2R, A151R, K205R, D345L).

The 3' UTR isoforms use the same TSSs as the canonical transcript but terminate at different points on these TESs. During this TES analysis, a total of 57 new isoforms were found among the 220 TESs. Out of them, 32 possess the poly(A) signal characteristic of eukaryotes. From the poly(A)-Seq data provided by Cackett and colleagues, we could validate 7 isoforms with our LRS dataset (ASFV G ACD 00,600-A224L-AT-L, A151R-AT-S2, I73R-AT-L4, I215L-AT-S, MGF 360-18R-DP71L-DP96R-AT-L)[74]. Throughout the analysis, intron-containing RNA molecules were not detected in the ASFV virus.

6.5 Upstream ORF-Containing mRNAs

Gene expression studies have recently revealed that in nearly half of mRNA molecules, upstream ORFs (uORFs) are present, which are translationally active and are located in the upstream direction from the main ORFs[119]. In our analysis, we detected 30 new uORFs, of which 6 overlapped with the main ORFs (Figure 9b). Additionally, in the case of 5' UTR isoforms, we identified 24 uORFs. The distance between the main ORFs and the upstream ORFs was 45.76 bp.

6.6 Canonic transcripts

In this section of transcript identification, we determined the canonical mRNA molecules for individual genes by aligning their TSS and TES points. Canonical transcripts were considered those RNA molecules that were the most abundant for the respective coding and non-coding genes. Canonical RNA molecules were found in 97 genes. These molecules are visible in Figure 8 in shades of gray. We did not categorize 5' truncated molecules into this group, as they did not contain the entire gene ORF but rather an internal ORF within themselves.

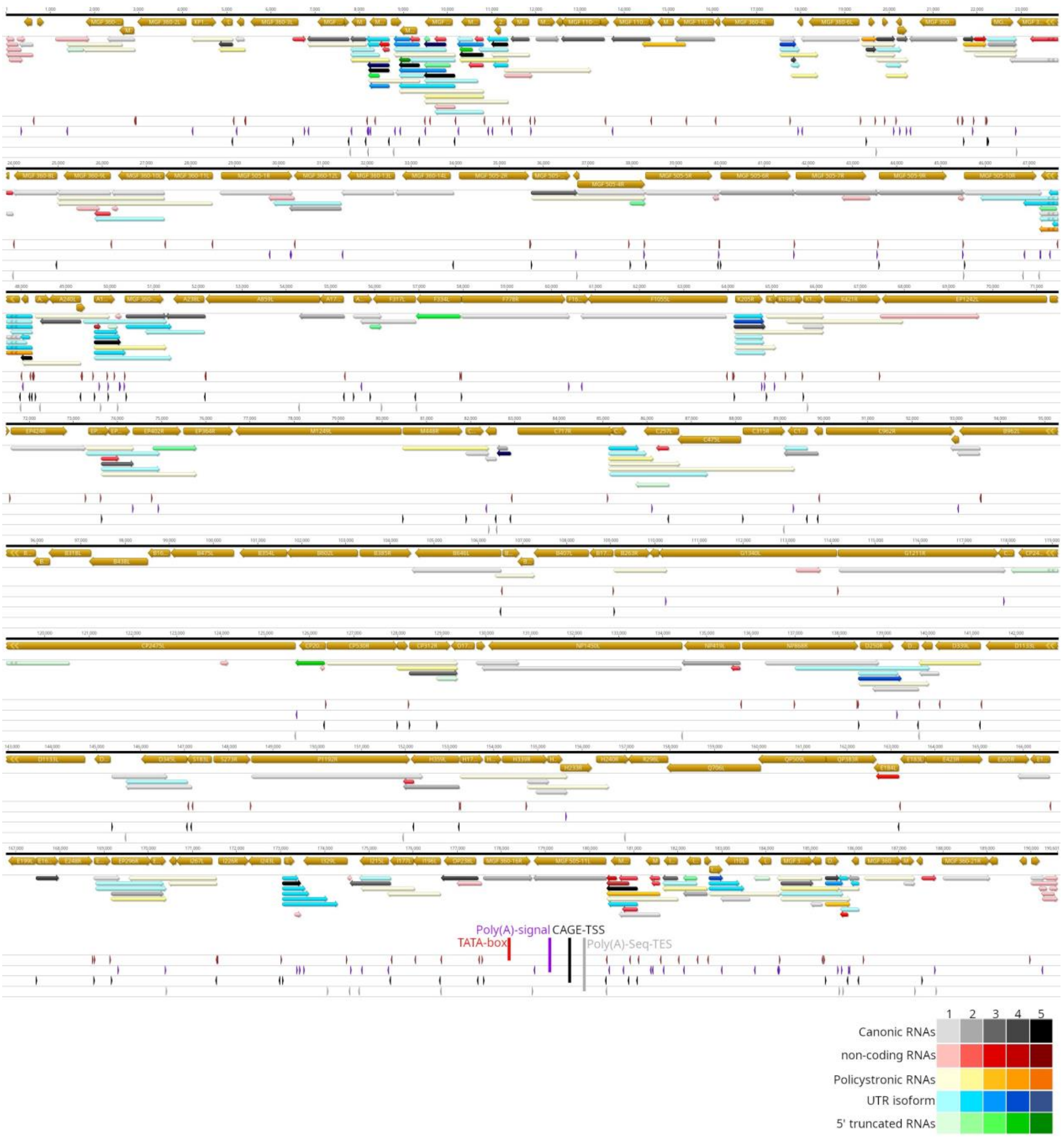


Figure 8. Atlas of the African swine fever virus transcriptome: We defined canonical RNAs as transcripts containing the same ORF within individual genes, marked with varying shades of gray to represent their abundance. Non-coding RNAs (ncRNAs) were defined as RNAs that are antisense or lack a functional ORF, localized in intra- or intergenic regions, and highlighted in red. Polycistronic RNAs were considered transcripts that transcribe multiple gene pairs, indicated with a yellow shade. Those RNAs with longer or shorter 5' or 3' ends compared to canonical RNAs were considered 5' and 3' UTR isoforms, marked with blue shades. RNAs containing truncated ORFs within a gene were classified as 5' truncated transcripts and labeled with a green shade. The arrows below the transcripts symbolize the following: red for TATA box, purple for Poly(A) signal, black for CAGE TSS[74], gray for poly(A)-

seq TES [74]. The five shades represent the abundance of transcripts: 1 for 1–9 reads, 2 for 10–49 reads, 3 for 50–199 reads, 4 for 200–999 reads, and 5 for >1000 reads.

6.7 Putative mRNAs

The 5' truncated mRNAs contain an in-frame ORF within the canonical ORF, with a separate ATG but a shared STOP codon. In our analysis, we identified 16 new 5' truncated transcripts (Table 2). Eight of them feature a TATA box, located at an average distance of 42.375 nucleotides from the TSS. In three cases, the TA(+1)TA late promoter motif characterizes the initiator sequences. The CAGE dataset from Cackett and colleagues confirmed the TSS of three such transcripts (A137R, CP312R, MGF 100-1L)[74]. Figure 9a illustrates an example of a 5' truncated transcript.

Novel intergenic transcripts with small ORFs. During our analysis, we identified 3 RNA molecules containing small ORFs (9-192 bp) detected by LoRTIA and 4 by non-LoRTIA methods, localized in the intergenic region of the virus. The CAGE analysis conducted by Cackett and colleagues validated a single such molecule, referred to as pNG6[74].

Name	TSS (+)/TES (−)	TSS (−)/TES (+)	Strand	Name	TSS (+)/TES (−)	TSS (−)/TES (+)	Strand
MGF 110-3L.3-AT-L	8183	8458	−	A224L.5	47,236	47,621	−
MGF 110-3L.3	8210	8458	−	A151R.3	49,961	50,173	+
MGF 110-4L.3-AT-L	8785	9146	−	A137R.4	55,895	56,146	+
MGF 110-4L.3	8902	9146	−	F334L.2	56,946	57,937	−
MGF 110-5L-6L.2-AT-L2	9468	9961	−	EP402R.2	74,801	75,785	+
MGF 110-5L-6L.1	9468	10,064	−	CP204L.2	125,685	126,344	−
MGF 110-5L-6L.6	9468	9585	−	CP312R.6	128,878	129,347	+
MGF 110-7L.1-AT-L	10,220	10,563	−	MGF 100-1L.2	180,386	180,891	−
MGF 110-7L.1	10,270	10,563	−	I8L.2	182,120	182,425	−
MGF 505-4R.13	37,956	38,295	+				

Table 2. List of the 5′ truncated transcripts.

6.6 Novel non-coding transcripts

We consider ncRNAs to be those RNAs located within or between protein-coding genes, which can be transcribed from both the negative or positive strands of the gene. In our analysis, we found 3 short ncRNAs (sncRNAs) and 42 long ncRNAs (lncRNAs - longer than 200 bp), including intragenic, intergenic, antisense (asRNAs), and replication origin-associated RNA molecules (raRNAs). Among these, 19 have a TATA box, with an average distance of 52.45 bps from TSS.

In our LRS analysis, we identified 3 putative sncRNAs localized in intergenic genomic regions (IG3-MGF 505-9R-MGF 505-10R, IG6-I73R-I329L, and IG7-I329L-I215L).

The 3’ truncated RNAs are regulated by promoters identical to canonical ORFs, using a common ATG with them, but they terminate before the canonical stop codons. Therefore, they either lack functional ORFs or encode short proteins (Table 3). In this virus, we identified 22 such 3’ truncated RNA molecules. To avoid false 3’ ends resulting from false priming, we filtered these 3’ ends using the LoRTIA false priming filtering algorithm[120].

Name	**TSS (+)/TES (−)**	**TSS (−)/TES (+)**	**Strand**	**Name**	**TSS (+)/TES (−)**	**TSS (−)/TES (+)**	**Strand**
nc-MGF 110-3L	8455	8680	−	nc-c275L	86,209	86,487	-
nc-MGF 110-3L-AT-S	8526	8680	−	nc-CP204L	126,247	126,344	-
nc-MGF 110-4L	9146	9368	−	nc-NP419L	135,550	135,755	-
nc-MGF 110-5L-6L	9705	9961	−	nc-H359L	151,958	152,195	-
nc-MGF 110-5L-6L-L	9705	10,162	−	nc-E184L	162,670	163,182	-
nc-MGF 110-7L-AT-L	10,481	10,814	−	nc-DP238L	176,989	177,547	-
nc-MGF 110-7L	10,619	10,814	−	nc-MGF 100-1L	180,386	180,613	-
nc-MGF 300-4L	23,203	23,976	−	nc-MGF 100-3L-MGF 100-1L-AT-L2	180,671	181,084	-
nc-MGF 360-9L	25,832	26,183	−	nc-MGF 100-3L-MGF 100-1L-AT-L3	180,743	181,084	-
nc-A151R	49,646	49,784	+	nc-MGF 100-3L	181,362	181,585	-
nc-EP152R-EP153R	73,637	74,032	+	nc-MGF 100-3L-AT-S	181,408	181,585	-

Table 3. List of 3′-truncated transcripts

Replication origin-associated RNAs (raRNAs) are molecules that overlap with or are located near the replication origin. Such transcripts have been identified in all herpesviruses, although their precise functions are often unclear[32,33]. These molecules can be variants of coding or non-coding mRNAswith a longer TSS or transcription end site TES[121]. In the case of ASFV, we identified 6 raRNA molecules during the late stages of infection (12-20 h). Their localization is within the terminal repeats at the ends of the genome, which likely contain the replication origin of the virus. (Figure 9C) The absence of raRNAs in the early stages of infection is likely due to the lower sequencing coverage of the viral transcriptome. Two out of the 6 raRNAs are in antisense orientation compared to the other four. DP60R and raRNA-1 share a common promoter, indicating that, unlike the others, they are mRNAs. However, due to the small coverage and the fact that these ends are not identified by LoRTIA, it cannot be excluded that the TSS of raRNA2-4 is in fact derived from the TSS of DP60R and raRNAs-1. Similarly, it is possible that raRNAs-5 and raRNAs-6 actually originate from a common TSS, but are different RNA molecules because they terminate at different points.

The regulation of antisense RNAs (asRNAs) molecules can involve distinct promoters, but they may also result from long transcriptional overlaps between neighboring and proximal convergent or divergent genes. In this study, we identified seven asRNA molecules (Figure 9d). Our Illumina analysis revealed the antisense expression of the whole genome. We hypothesize that one reason for the presence of these antisense molecules could be transcriptional read-throughs occurring between convergent genes and the overlap of long 5’ UTRs of divergent genes. However, these asRNA molecules can also be regulated by their own cis-regulatory sequences, as in the case of the example shown in Figure 9d.

6.8 Polycistronic and complex transcripts

Until now, it was widely assumed that the ASFV virus only transcribes monocistronic transcripts[74,75,122]. However, during our LRS study, it was proven that not only monocistronic molecules are produced, but we also revealed a broad polycistronism in the viral transcriptome of the ASFV genome. A total of 51 polycistronic molecules were identified, of which 41 were bicistronic, 8 were tricistronic, and 2 tetracistronic transcript molecules (Figure 9e).

In this study, we also discovered complex RNAs that transcribe several genes, and at least one of them also transcribes a gene of opposite polarity (e.g., →←→). A total of 22 complex RNA molecules were identified by LRS analysis (Figure 9f).

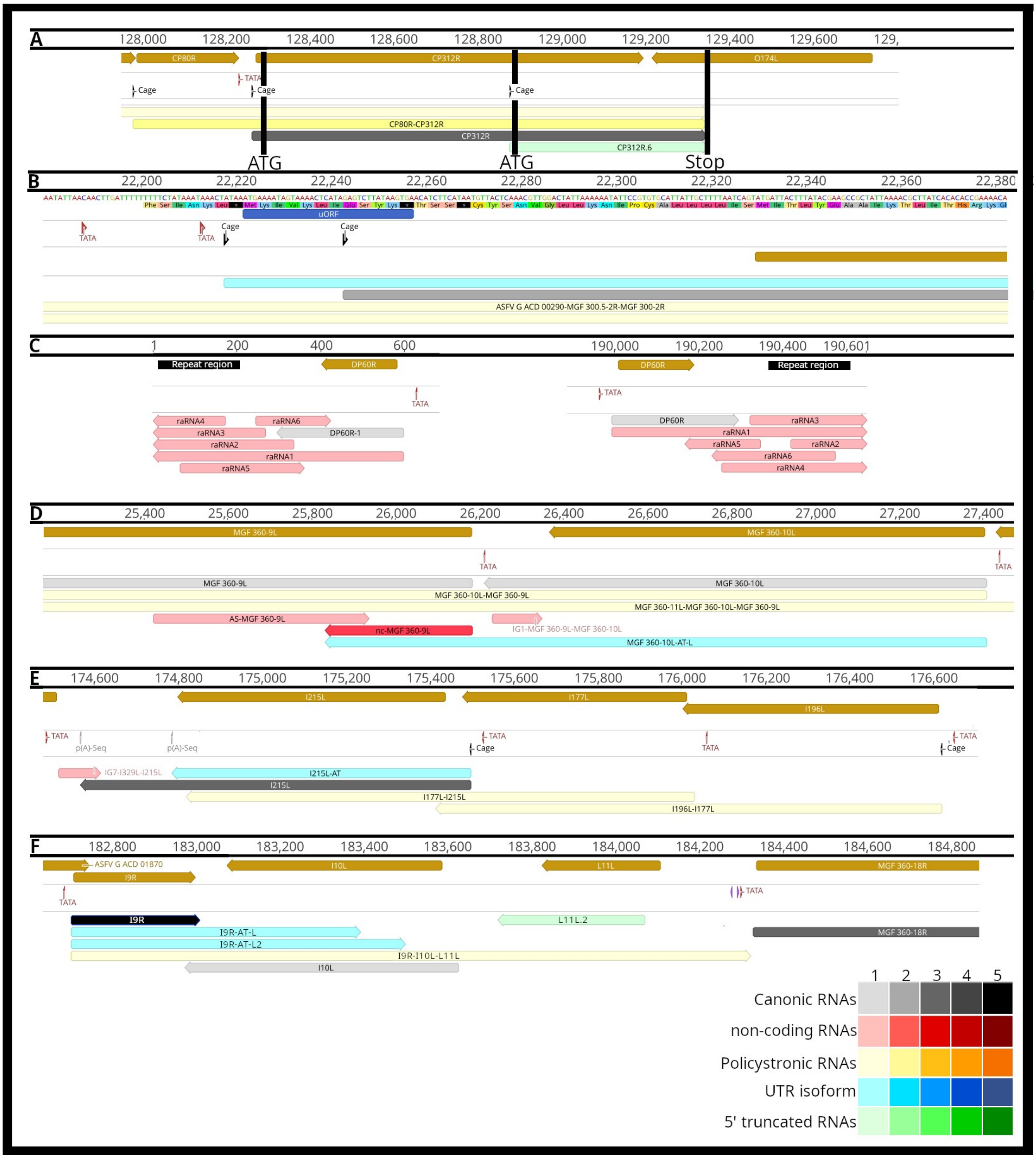


Figure 9. Examples for the ASFV transcript classes (**a**) 5′-truncated mRNAs. This part of the figure illustrates the putative nested mRNAs (CP312R.6; red arrow, embedded into the larger CP312R transcript; gray arrow) containing short 5′-truncated in-frame ORFs within the canonical ORFs (CP312R gene; yellow arrow). (**b**) Transcripts containing upstream ORFs MGF 300-2R transcript and its longer 5′-UTR isoform (MGF 300-2R-L) containing an uORF (light blue rectangle) is illustrated in this picture. The bracketed picture contains two full-length transcript isoforms (gray arrows), as well as the ORF of the gene encoding them and the uORF (blue arrows). (**c**) Replication origin-associated RNAs This figure illustrates the repeat region at the genomic termini (black rectangle) and the raRNAs (light red arrows),

which overlap or situated at the vicinity of the Oris, of which the precise location is unknown. (**d**) Antisense RNAs The red arrow indicates an asRNA (AS-MGF 360-9L) overlapping the MGF 360-9L gene (gray arrow) in opposite polarity. (**e**) Polycistronic RNAs I177L-I215L bicistronic transcript (light green arrow) comprising two tandemly oriented genes, of which the I215L is also expressed as a monocistronic RNA molecule (gray arrow). (**f**) Complex transcripts The I9R-I10L-L11L complex transcript (yellow) comprises three genes, of which I9R (encoding three TSS isoforms; blue arrows) stands in an opposite orientation compared to the other two genes (L11L.2 and I10L).

6.9 Transcriptional overlaps

In this work, a total of 540 parallel (→→), 19 divergent (←→), and 60 convergent (→←) transcriptional overlaps were identified. Between the transcriptional overlaps, we distinguished hard and soft overlaps. In the "hard" case, the canonical transcripts of the gene pairs create overlaps with each other, while in the "soft" case, only alternative terminations of the gene pairs create overlaps. The examples of the three types of transcriptional overlaps can be seen in the gene clusters of ASFV in the Figure 10. Transcriptional read-throughs between neighboring genes, which can occur in convergent and tandem orientations relative to each other, result in transcriptional overlaps. Tandem gene clusters usually result in shared 3' co-terminal transcripts. In these gene clusters, upstream genes may have their own transcriptional termination, meaning they can express monocistronic transcripts as well. The production of mono or multicistronic transcripts of the same gene is regulated by transcriptional termination sequences.

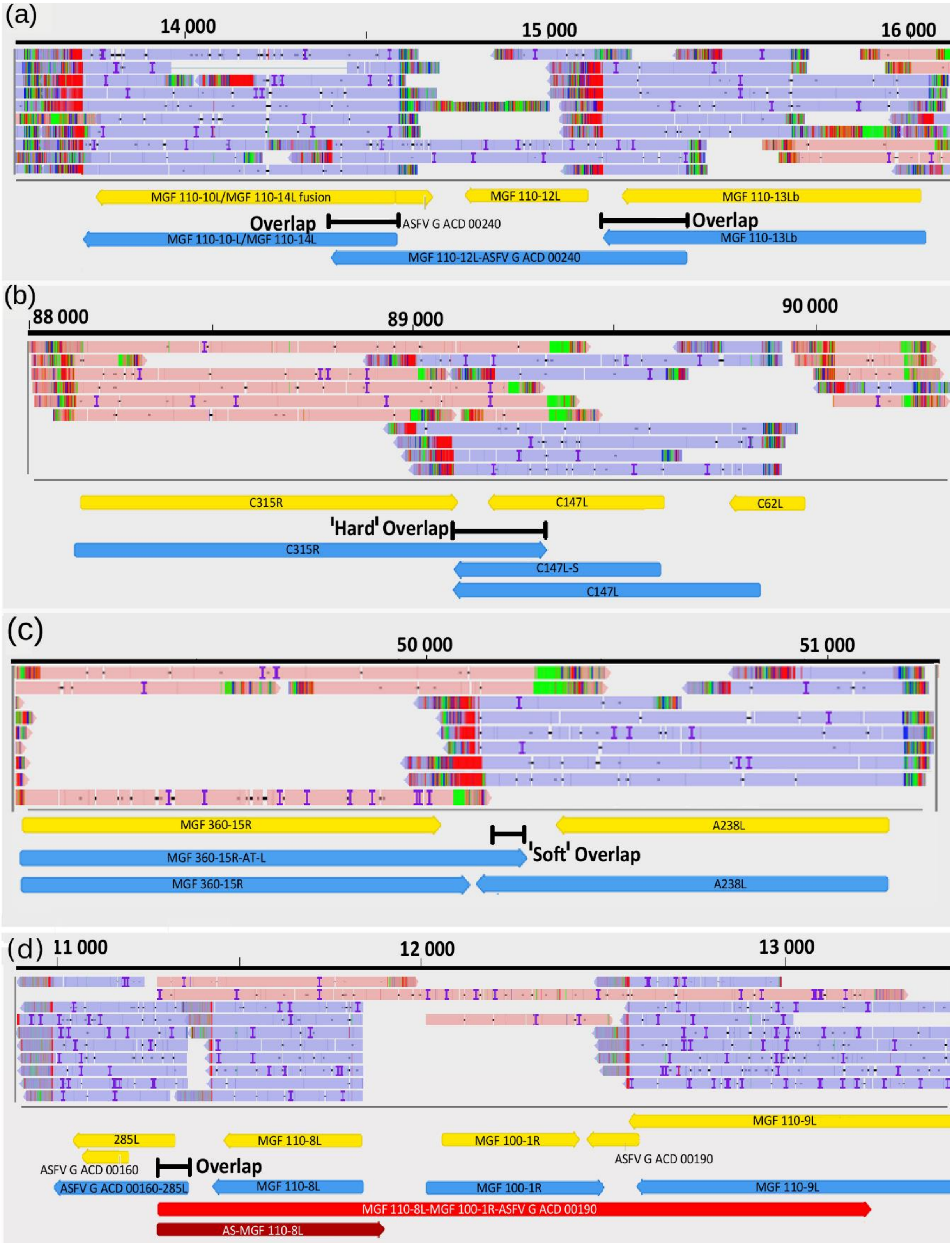


Figure 10. Examples for transcription overlaps of ASFV RNA molecules. (a) parallel overlaps, (b) convergent "hard" overlap (c) convergent "soft" overlap, (d) divergent overlap.

7. Results - Analysis of the Pseudorabies virus (PRV) transcriptome

7. 1 Sequencing and mapping statistics

In our analysis, we used two LRS platforms, which were the ONT-minION and the PacBio RSII and Sequel devices. For ONT MinION and PacBio sequencing, oligo(d)T and random primers were used to perform reverse transcription. Amplified cDNA libraries were prepared on both platforms, as well as PacBio Isoform and ONT-minION dRNA sequencing without PCR amplification.

During PacBio sequencing, a total of 210,573 reads were generated, with an average length of 1255 bps. In the case of ONT-minION sequencing, 5,067,913 reads were generated, with an average length of 589 bps. Detailed information about the reads can be found in Table 4.

Sequencing Techniques	Number of Host Cell Reads	Viral RNA Read Counts	Average Length of Viral Reads in bp
PacBio Sequel	117,079	13,292	1553
PacBio RS II amplified	462,202	116,905	1255
PacBio RS II non-amplified	176,919	52,012	1282
PacBio-random primers	112,081	28,364	932
MinION-amplified oligo(d)T	4,273,446	1,385,284	517
MinION-non-amplified oligo(d)T	4,907,412	3,451,129	909
MinION-random	5,144,609	231,500	341

Table 4. Mapping, Sequencing and read length statistics of ASFV virus.

7.2 Annotating the viral TSSs, TESs, promoter motif and poly(A) signals

Using the LoRTIA software package, we annotated and determined the TSSs and TES of the transcripts. Two different technical approaches were set as conditions for accepting endpoints. In total, we annotated 465 TSSs and 57 TESs, identifying a total of 619 transcripts. Among these 619 transcripts, 410 correspond to newly identified transcripts. Additionally, we have confirmed 55 canonical transcripts detected in the previous works[103]. Figure 12 illustrates the updated PRV transcriptome.

In promoter analysis, a total of 74 TATA boxes were found, with an average distance of 34,85 bps from the TSS. Additionally, 17 CAAT boxes were identified with an average distance of 113.294 bps from the TSS. Furthermore, 50 GC boxes were found, with an average distance of 34,88 bps from the TSS. Initiators showed a high proportion of G at the first two positions, and G/C at the third position (see Figure 11a). The high G content in initiators has been demonstrated in the HSV-1 VP5 promoter.

Our investigation revealed that 51 TESs possess a PAS signal, with an average distance of 26,81 nt from the TES. For TESs, a sequence environment typical of eukaryotes was observed, including a C/A cleavage signal and downstream element rich in U/G.

A total of 24 short and 166 long 5’ UTR TSS isoforms were determined, as well as 22 3’ UTR TES isoforms. Due to the strict criteria applied to annotations, the number of short isoforms is likely higher than accepted in this study. Some long 5’ UTR variants, containing 5’ truncated in-frame ORFs, may not necessarily represent real isoforms, as downstream genes are not translated if the ORFs are active. (See Figure 11b)

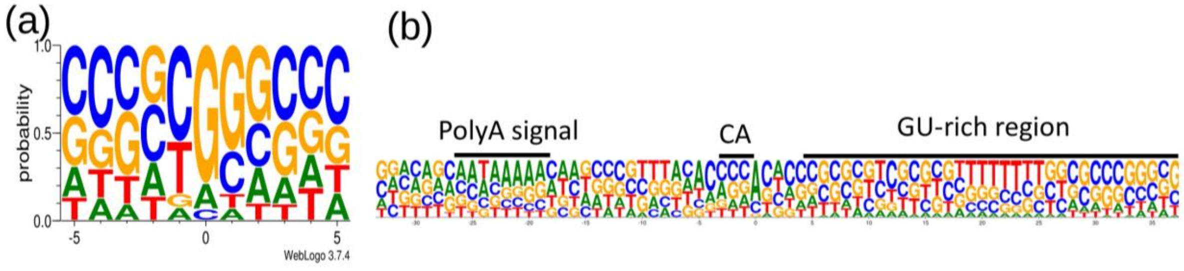


Figure 11. The base content at the 5′- and 3′-termini of the PRV transcripts: The x-axis represents the position of the nucleotides relative to the TSS (a) or TES (b), whereas the y-axis shows the frequency of the given nucleotide at a given position. (a) Most of the transcripts have GC-rich 5′ termini. The position “0” is the first nucleotide of the TSS. (b) Three primary sequence elements are frequent at the 3′-end of the RNAs: The hexameric polyadenylation signal (typically AAUAAA), the

cleavage site (most commonly a CA dinucleotide), and the downstream sequence element (typically U/UG rich). “U”-s are shown as “T”. The position “0” is the potential polyadenylation (PA) site. The logo shows the +/− 50 bp interval of the PA site. The image is generated using weblogo 3.0[123].

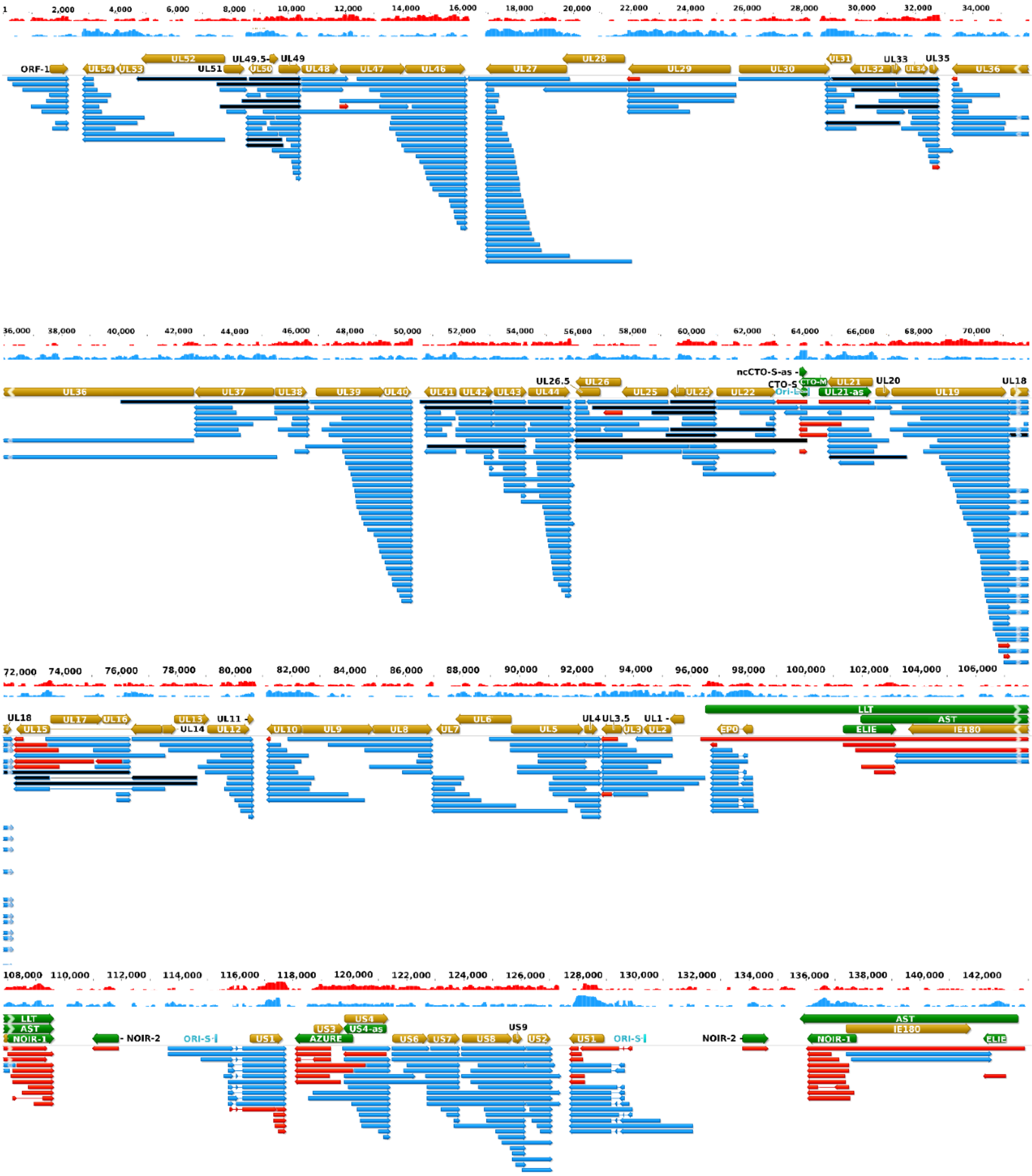


Figure 12. The updated pseudorabies virus (PRV) transcriptome. PRV transcriptome contains those transcripts that were identified by the integrated approach using novel and earlier short-and long-read sequencing datasets. The light brown arrows represent the open reading frames of genes; the green arrows show the non-coding RNA genes; the blue arrows are the mRNAs; and the red arrows illustrate the non-coding transcripts. Complex transcripts are also colored by black, although it cannot be excluded

that they function as mRNAs. These RNA molecules contain multiple genes of which at least one is oriented in an opposite direction relative to the others. Long latency Transcript (LLT) and Antisense Transcript (AST) were detected in latently infected neurons, and by real-time RT PCR in lytic infection (especially when cells were treated with cycloheximide, a protein synthesis inhibitor[124], but not in any of our sequencing experiments that all used lytically infected samples. The depth of viral read coverages generated by Illumina sequencing are represented by red (+ strand) and blue (−strand) colors. The coverage is plotted in a log scale using Geneious software.

7.3 Novel putative mRNAs

Truncated RNAs with common co-termini are a feature of alphaherpesviruses. In-frame ORFs shorter than the canonical gene and located downstream of its ATG. These molecules, containing a shorter ORF, can encode N-terminally truncated polypeptides. Such molecules have been identified in several alphaherpesviruses. Previously, such genes were described in PRV virus, such as UL36.5 within UL36 and UL26.5 within UL26. In this analysis, a total of 206 5' truncated RNAs were identified, 189 of which are new truncated RNAs. Many of them contain the same in-frame ORF, so they can be considered as 5' UTR isoforms. (See Figure 13.) For the accepted truncated RNAs in the analysis, we added to the existing strict conditions that they must also be present in the dRNA sequencing. Another interesting phenomenon was that in the case of these truncated RNAs, practically no ATGs were found in the other two reading frames outside of the in-frame.

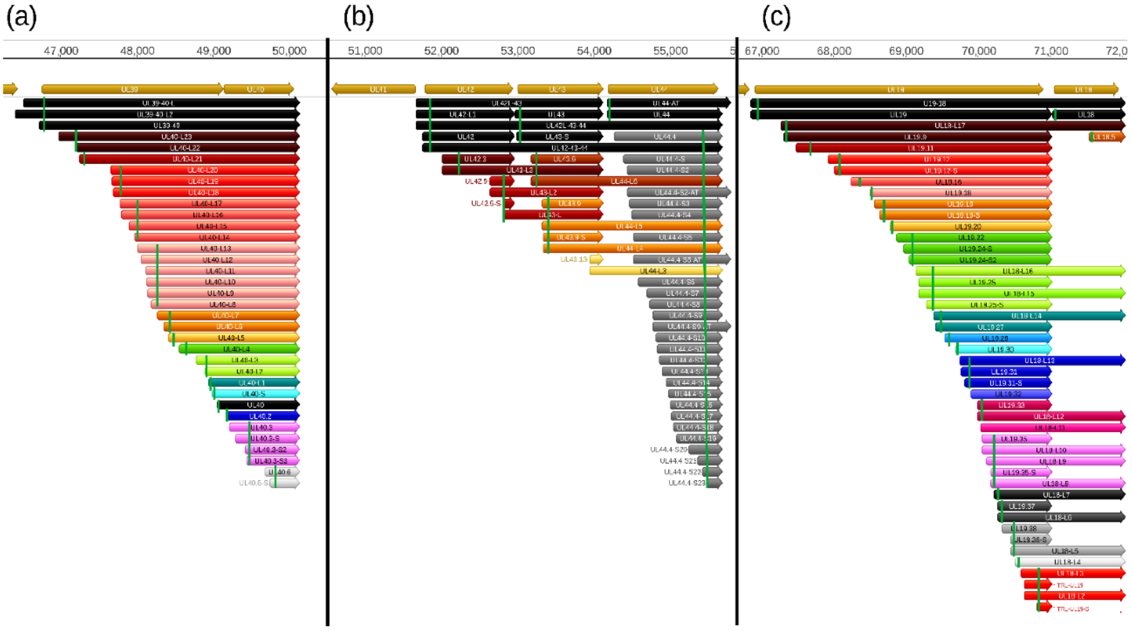


Figure 13. Truncated mRNAs. The following genomic regions are selected for the illustration of the truncated mRNAs: (**a**) ul39-ul40, (**b**) ul42-43-44, and (**c**) ul19-18. Arrows with the same color represent transcripts containing the same ORFs but distinct TSSs or TESs. The rectangular green lines indicate

the first in-frame ATGs within the transcripts. The light brown arrows represent the open reading frames of the PRV genes.

7.4 Replication-Origin Associated Transcripts

Replication-associated RNAs have been identified in alpha, beta, and gamma herpesviruses. Initially, in the PRV virus, CTO-S (Close to replication origin) was discovered, which is localized in the UL region near ORI-S, followed by the later identification of PTO (PTO: proximal to origin) overlapping ORI-L and US1 in the genome's US region. In the CTO region, four transcripts (CTO-S, CTO-M, CTO-L, and CTO-AT) had previously been identified. In addition, for CTO-M, a new longer isoform was found, and for CTO-S, a new alternative termination, CTO-S-AT2, was detected. (See Figure 14a)

Another interesting feature of the region is a 8135 bp-long RNA, CTO-S-cx, which uses the same promoter as CTO-S. It transcribes five genes, including Ul22, UL23, UL24, UL25, and Ul26. Regarding read counts, another interesting observation was made. CTO-S had a total of 570,653 reads, while the second most highly expressed gene, UL18, had 20,543 reads, and the third most expressed gene, UL16, had a total of 12,206 reads. This implies that in the PRV genome, the most abundantly expressed RNA molecule is CTO-S, whose function is currently unclear. In addition to these, we also detected a new spliced version of the PTO-US1 RNA that completely overlaps the OriL region. (See Figure 14b.)

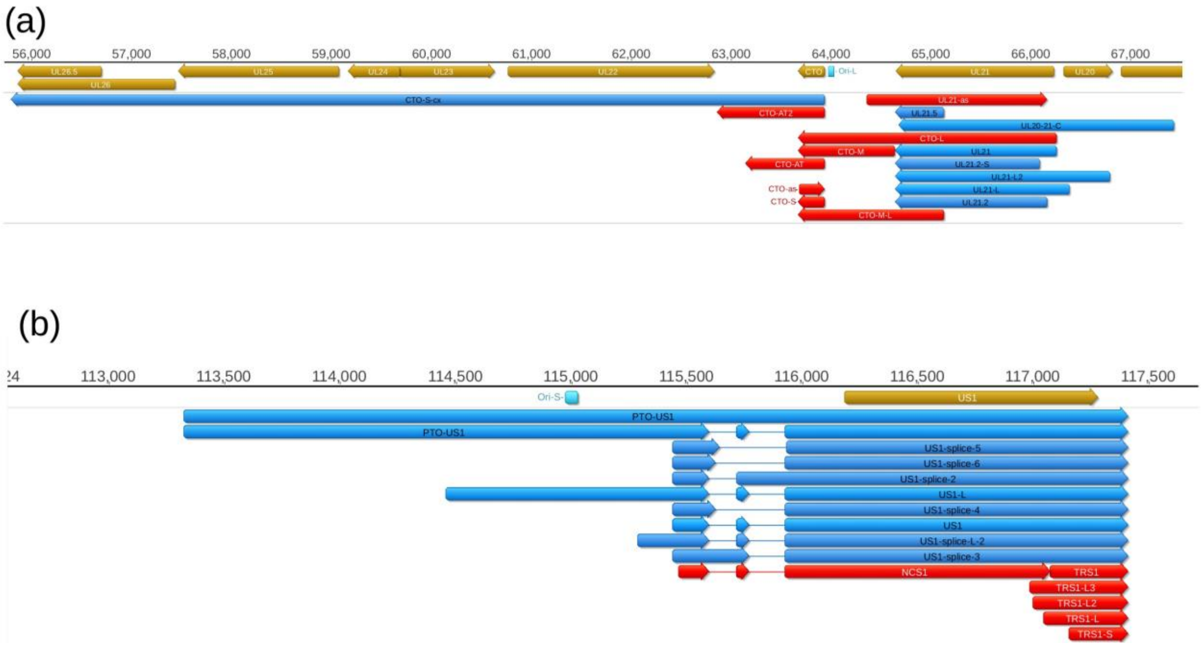


Figure 15. Replication origin-associated transcripts at the (**a**) Ori-L (a) and (**b**) the Ori-S genomic regions. These types of RNA molecules have been described in several viruses, including herpesviruses.

Except the CTO-S transcripts, these RNA molecules overlap the replication origins through either their 3′-UTR (CTO-L), or their 5′-UTR (PTO-US1). Both the raRNAs and the transcripts of adjacent genes are overlapped by antisense RNAs of which some are controlled by separate promoters. Color code: Light brown: Coding or non-coding genes, blue: mRNAs, red: ncRNAs, light blue: Origin of replications (Ori-L and Ori-S) of the virus.

7.5 Non-Coding Transcript

In this study, we annotated new ncRNAs using LRS platforms, which fall into the long non-coding RNAs (lncRNAs) category (>200 bp) (Figure 15). These ncRNAs, such as NOIR-1, ELIE, and AZURE, have been previously identified, but we detected new isoforms for these RNAs. These RNAs are localized in the PRV US region.

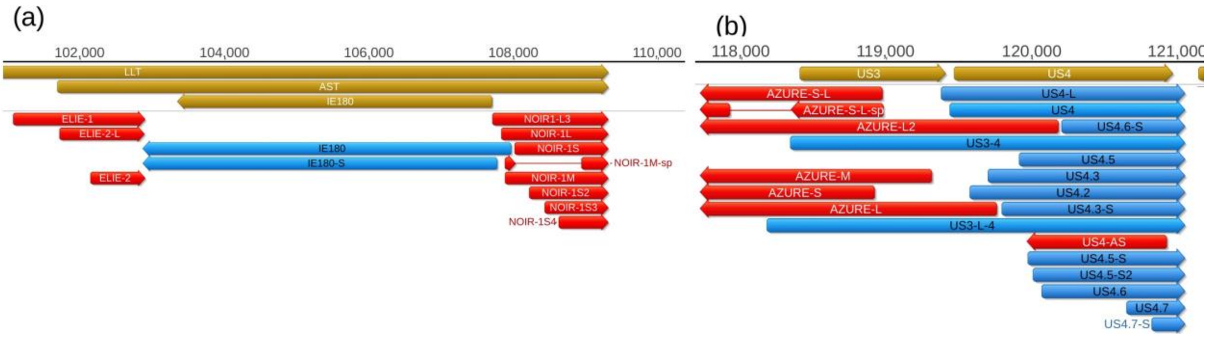


Figure 15. Coding and non-coding RNA molecules at the ie180-us4 genomic region. A high density of non-coding transcripts can be observed at this genomic region. Color code: Light brown: Coding, or non-coding genes, blue: mRNAs, red: ncRNAs, light brown: ORFs.

The NOIR-1 ncRNAs are in the opposite orientation to the IE180 gene, and their 5' end partially overlaps with it. The 3' end terminates in a common coterminal region with Antisense Transcripts (AST) and Long Latency Transcripts (LLT). While the NOIR-1M, NOIR-1L, and NOIR-1S transcripts have been previously identified, we found a spliced variant for NOIR-1M, which was confirmed by dRNA sequencing. Additionally, we annotated three shorter forms for NOIR-1S and a longer new form for NOIR-1L. The AZURE, located in antisense to the US3 gene, has its longest isoform overlapping with the US4 gene. Four RNAs, namely AZURE-M, AZURE-L, AZURE-L2, and AZURE-S, have been previously identified. Furthermore, we found a spliced and a longer isoform for AZURE-S, which were also validated by dRNA. Additionally, we reannotated the uncertain TSS for US4-AS in the US4 region.

Illumina SRS revealed antisense transcriptional activity across the entire genome. These antisense ncRNAs, such as US4 and AZURE, may be regulated by their own promoters or arise from convergent (tail-to-tail orientation; →←) or divergent (head-to-head direction; ←→) oriented gene pairs.

In Alphaherpesviruses, the structure of the UL15 gene is unique because the UL16 and UL17 genes are located in this ORF in antisense polarity. It was previously observed that the downstream region of the UL15 gene, fORF15, can also be transcribed, containing no in-frame ORF[104]. We identified new length isoforms for this fORF15, including one that overlaps with the UL15 intron region.

7.6 Novel Multigenic Transcripts

Using LRS analysis, we annotated 87 new polycistronic transcripts. Among them, there were 59 bicistronic, 19 tricistronic, 8 tetracistronic, and 1 pentacistronic variants. The pentacistronic RNA was the longest (7130 bps), transcribing the UL49.5, UL49, UL48, UL47, and UL46 genes.

In addition to the polycistronic transcripts, we also detected complex transcripts that transcribe multiple genes in opposing orientations. The longest complex transcript was the CTO-S-cx mentioned in replication-associated RNAs (8135 bps). For 9 out of 24 complex transcripts, the annotation of TSSs could not be precisely determined, as they were only found in dRNS-Seq. In the case of PacBio and ONT cDNA sequencing, it is difficult to detect very long multicistronic RNA molecules, due to the fragility of the RNAs and errors in RT, PCR reactions.

7.7 Transcriptional Overlaps

Due to the newly annotated 5' and 3' UTR isoforms and multicistronic transcripts, our analysis revealed an even more complex network of transcriptional overlaps. For divergent genes, the overlaps are caused by overlapping 5' UTR isoforms, while for convergent genes, transcriptional read-through events are responsible. In this analysis, we demonstrated that overlaps occur between all convergent gene pairs, likely due to transcriptional read-through by RNAPII, which has been shown to continue transcribing the RNA molecule beyond transcription termination, resulting in antisense segments against the adjacent gene. Another interesting result is that TES isoform RNAs are produced from some genes that span the intergenic region and terminate exactly in the TES of the adjacent convergent gene (UL27-AT, UL35-AT, UL44-AT, CTO-S-AT, US2-AT).

8. Discussion

The third-generation ONT-minION and PacBio platforms have revolutionized transcriptomic research[12]. With these technologies, our group and others have revealed an extremely high complexity of viral transcriptomes[12,125,126]. Different LRS platforms and chemistries have different advantages and disadvantages. Therefore, the combination of various sequencing approaches is crucial for the accurate reconstruction of transcriptomic architecture. In this work, I analyzed the global transcriptomes of African Swine Fever and Pseudorabies viruses using long-read sequencing. In both cases, we performed both amplified and non-amplified cDNA sequencing, as well as native dRNA sequencing, to eliminate errors arising from reverse transcription (RT) and template switching. Additionally, we used the LoRTIA software package developed by our group to determine the TSSs and TESs of RNAs. This proved to be a useful approach in removing false 5' ends generated due to mRNA degradation and eliminating false ends resulting from template switching and false priming. Using these techniques, we identified numerous new viral RNAs in ASFV and PRV, including classes that are challenging to study at the genome level with SRS technologies due to their short read lengths. These include read-through RNAs, multicistronic RNAs, TSS and TES RNAs isoforms, as well as putative embedded genes. Furthermore, LRS is capable of identifying ncRNAs, including intergenic and antisense RNAs. Recently a group of ncRNAs transcribing the replication origin of viruses and in its environment was also described using this platform.[32]

In our analysis, we annotated 465 TSSs and 57 TESs in PRV, resulting in a total of 619 categorized transcripts (of which 410 have not been published before). In the case of ASFV, we identified 202 TSSs and 220 TESs, resulting in 311 transcripts. From the SGS results of Cackett and colleagues, we were able to validate 98 TSS and 57 TES using LRS[74]. Among these, 14 long, 2 short 5’ UTR and 57 3’ UTR variants were identified. In contrast to SGS approaches, we could connect and annotate these TSSs and TESs as transcripts. Additionally, we were able to associate uORFs found in the 5' UTRs with transcripts. One biological role of these 5’ and 3’ UTR variants is to create overlaps. The 5’ UTR variants may contain uORFs, which can have a strong impact on translation[127]. On the other hand, the 3’ UTR variants may contain important sequences for miRNAs, allowing them to regulate the translation of mRNAs[81].

The identification of smaller genes embedded within larger genes revealed a broad spectrum of viral RNA complexity in both viruses[40,44]. In many viruses, truncated RNAs at their 5’ ends have been detected, which are transcripts containing a shared stop codon with canonical genes, but their ATG is located downstream of it. The SRS approach is not efficient enough in discovering these molecules, which is why they were previously unidentified[40]. In the case of PRV, we identified 209 5’ truncated mRNAs (of which 189 have not been published before). One source of these genes may be those embedded in larger canonical genes, and the other may be the 5' UTR isoforms of the transcripts. In our analysis, we considered these to be long isoforms, but it is possible that they are bicistronic molecules, since their UTR contains the in-frame ORF of the upstream gene. In the case of ASFV, 19 5' truncated RNAs were detected, from which Cackett and colleagues' CAGE results validated 3 TSSs. Additionally, we identified putative genes encoding proteins that contain RNAs with small ORFs and are localized in intergenic regions. Cackett and colleagues identified a similar putative gene encoding an RNA molecule, referred to as pNG6, for which they determined the TSS[80]. However, using LRS, we determined the complete 5’ and 3’ ends of the new nucleic acid molecules. Further investigations are needed for these 5’ truncated and intergenic RNA molecules, such as Northern blotting, as well as experiments assessing their coding capacity, such as Western blotting or mass spectrometry.

Polycistronism is a common feature among bacteria, which is also present in viruses, as they are capable of expressing multigenic RNA molecules. In our analysis, we detected such molecules in both PRV and ASFV viruses. In ASFV, it was previously assumed that only monocistronic molecules are expressed[75,122]. However, our analysis revealed that ASFV is also capable of expressing multicistronic molecules transcribing multiple genes in both opposite and the same polarities. In PRV, we identified a total of 87 polycistronic and 24 complex transcripts (of which 49 have not been published before), while in ASFV, we detected 51 new polycistronic and 22 new complex transcripts.

LRS studies have yielded significant results, identifying numerous raRNAs molecules in alpha, beta, and gammaherpesviruses[33]. In Alphaherpesviruses such as PRV and EHV-1, there are three replication origins, among which OriS remains conserved and is localized between the ICP4 and US1 genes[33,40]. However, OriL differs among individual viruses, disappearing in VZV and BoHV-1 but persisting in EHV-1, HSV-1, and PRV. RNAs molecules transcribing and overlapping near OriL and OriS show a high degree of diversity in PRV. The raRNAs of OriL may include non-overlapping transcripts with Ori, such as CTO-S and CTO-AT, non-coding ones like CTO-M, and mRNA isoforms like CTO-L, which is a long 3' UTR variant of UL21. Our study revealed that CTO-S, located near OriL, is one of the most abundantly expressed RNA molecule in the viral genome. Additionally, we detected a new alternative termination of CTO-S, which overlaps the UL22 RNAs, and a complex RNA uses the promoter of CTO-S for its transcription. These raRNAs is supposed to participate in mechanisms regulating the initiation of virus replication and the orientation of replication fork. However, they also have poly(A) sequences, suggesting they may function as RNA molecules in the virus life cycle. Besides these roles, they might be involved in other functions, such as forming DNA-RNA hybrids around the ORI[128]. OriS raRNAs, which are probably non-coding like ELIE and NOIR-1, or mRNA isoforms like PTO-US1, were also identified. ELIE and NOIR-1 transcripts overlap with the LLT/AST genes, which raises the possibility that they somehow inhibit the expression of these latency genes during lytic infection. However, further analyses are required. In this study, new NOIR and AZURE length and splice variants, as well as the spliced version of PTO-US1 were identified. In ASFV, raRNAs molecules were also identified, localized in terminal repeats at the genome ends. Six such molecules were identified, likely functioning as non-coding RNA molecules, with one potentially functioning as an mRNA as it shares a promoter with the DP60R gene. Additionally, we cannot rule out the possibility that each of them also has mRNA functionality, as they all contain poly(A) sequences.

Our work revealed a genome-wide network of overlaps in both ASFV and PRV. These transcriptional overlaps can occur between convergent, divergent, and parallel genes[12,31]. In herpesviruses, tandem genes are organized into gene clusters where the transcribing polycistronic molecules create parallel overlaps. This arrangement follows the pattern: 'abcd', 'bcd', 'cd', and 'd', where 'a' is the most upstream gene and 'd' is the most downstream. The possible function of these molecules could be for ribosomes to use the uORF on them during translation[129,130]. Therefore, if a specific RNA's 5’ UTR region contains a uORF, it can influence the efficiency of translation of a downstream gene. This has been demonstrated in the case of the KSHV virus. However, these molecules may have other roles beyond the translation of downstream genes. Convergent overlaps arise from transcriptional read-throughs, for which the RNAPII enzyme is responsible[131]. Divergent overlaps result from overlapping 5’ UTR isoforms. Convergent and divergent transcripts create antisense segments in the oppositely polarized adjacent gene. Both ASFV and PRV viruses exhibit parallel, convergent, and divergent overlaps, though their roles are currently unclear. A suggested hypothesis is that the role of these overlaps is in the regulation of gene expression through influencing the transcriptional apparatus. This way, they could regulate the expression of virus genes in space and time. This hypothesis, known as TIN, has already been examined in various organisms[31,132,133]. However, the available results regarding the existence of this genetic regulation are currently inconclusive in the case of viruses.

9. Conclusions

To characterize the transcriptomes of African Swine Fever and Pseudorabies viruses in detail, we employed long-read sequencing platforms, which, unlike next-generation sequencing platforms, are capable of identifying full-length RNA. For both viruses, we prepared PCR-amplified and PCR amplification-free cDNA and RNA libraries. Our results indicate that the transcriptome profile of these viruses is much more complex than previously thought. Using the Long-Read Sequencing (LRS) approach, we were able to identify numerous new transcript isoforms, putative embedded genes, non-coding RNAs, polycistronic transcripts, and a new class of replication-associated RNAs. The newly identified transcript isoforms and polycistronic RNAs create a complex transcriptional overlap in the viral genome, which may play a crucial role in the regulation of their gene expression. Through these approaches, we obtained a more accurate picture of the transcriptome atlas of African Swine Fever and Pseudorabies viruses, which can be extremely useful in understanding the genetic regulation and gene expression of these viruses.

10. Acknowledgements

I would like to take this opportunity to express my gratitude and respect to everyone who contributed to the completion of this thesis. I want to express my thanks to Prof. Dr. Zsolt Boldogkői for making it possible for me to work in his research group and gain valuable professional experience.

I want to express my gratitude to my supervisor, Dr. Dóra Tombácz, who spared no effort in teaching, supporting, and providing much assistance and advice throughout the preparation of this work.

Additionally, I am thankful to all the staff at the Institute of Medical Biology for their assistance.

Finally, I would like to express my appreciation, heartfelt love, and gratitude to my family members, who provided support and love throughout the entire process.

11. 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. Clifton NJ* **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 Genet.* **34**, 666–681 (2018).
8. Marx, V. Method of the year: long-read sequencing. *Nat. Methods* **20**, 6–11 (2023).
9. Turnbull, C. *et al.* The 100 000 Genomes Project: bringing whole genome sequencing to the NHS. *BMJ* **361**, k1687 (2018).
10. Weimer, B. C. 100K Pathogen Genome Project. *Genome Announc.* **5**, e00594-17 (2017).
11. Metzker, M. L. Sequencing technologies — the next generation. *Nat. Rev. Genet.* **11**, 31–46 (2010).
12. Boldogkői, Z., Moldován, N., Balázs, Z., Snyder, M. & Tombácz, D. Long-Read Sequencing – A Powerful Tool in Viral Transcriptome Research. *Trends Microbiol.* **27**, 578–592 (2019).

13. Satam, H. *et al.* Next-Generation Sequencing Technology: Current Trends and Advancements. *Biology* **12**, 997 (2023).

14. Liu, L. *et al.* Comparison of next-generation sequencing systems. *J. Biomed. Biotechnol.* **2012**, 251364 (2012).

15. van Dijk, E. L., Auger, H., Jaszczyszyn, Y. & Thermes, C. Ten years of next-generation sequencing technology. *Trends Genet. TIG* **30**, 418–426 (2014).

16. Byrne, A., Cole, C., Volden, R. & Vollmers, C. Realizing the potential of full-length transcriptome sequencing. *Philos. Trans. R. Soc. B Biol. Sci.* **374**, 20190097 (2019).

17. Takahashi, H., Kato, S., Murata, M. & Carninci, P. CAGE- Cap Analysis Gene Expression: a protocol for the detection of promoter and transcriptional networks. *Methods Mol. Biol. Clifton NJ* **786**, 181–200 (2012).

18. Yu, F. *et al.* Poly(A)-seq: A method for direct sequencing and analysis of the transcriptomic poly(A)-tails. *PLOS ONE* **15**, e0234696 (2020).

19. Eid, J. *et al.* Real-time DNA sequencing from single polymerase molecules. *Science* **323**, 133–138 (2009).

20. Levene, M. J. *et al.* Zero-mode waveguides for single-molecule analysis at high concentrations. *Science* **299**, 682–686 (2003).

21. Travers, K. J., Chin, C.-S., Rank, D. R., Eid, J. S. & Turner, S. W. A flexible and efficient template format for circular consensus sequencing and SNP detection. *Nucleic Acids Res.* **38**, e159 (2010).

22. Lu, H., Giordano, F. & Ning, Z. Oxford Nanopore MinION Sequencing and Genome Assembly. *Genomics Proteomics Bioinformatics* **14**, 265–279 (2016).

23. Schneider, G. F. & Dekker, C. DNA sequencing with nanopores. *Nat. Biotechnol.* **30**, 326–328 (2012).

24. Goodwin, S., McPherson, J. D. & McCombie, W. R. Coming of age: ten years of next-generation sequencing technologies. *Nat. Rev. Genet.* **17**, 333–351 (2016).

25. Cocquet, J., Chong, A., Zhang, G. & Veitia, R. A. Reverse transcriptase template switching and false alternative transcripts. *Genomics* **88**, 127–131 (2006).

26. DeStefano, J. J., Mallaber, L. M., Rodriguez-Rodriguez, L., Fay, P. J. & Bambara, R. A. Requirements for strand transfer between internal regions of heteropolymer templates by human immunodeficiency virus reverse transcriptase. *J. Virol.* **66**, 6370–6378 (1992).

27. Pfeiffer, J. K. & Telesnitsky, A. Effects of limiting homology at the site of intermolecular recombinogenic template switching during Moloney murine leukemia virus replication. *J. Virol.* **75**, 11263–11274 (2001).

28. Garalde, D. R. *et al.* Highly parallel direct RNA sequencing on an array of nanopores. *Nat. Methods* **15**, 201–206 (2018).

29. Soneson, C. *et al.* A comprehensive examination of Nanopore native RNA sequencing for characterization of complex transcriptomes. *Nat. Commun.* **10**, 3359 (2019).

30. Tombácz, D. *et al.* Meta-analytic approach for transcriptome profiling of herpes simplex virus type 1. *Sci. Data* **7**, 223 (2020).

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

32. 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).

33. Torma, G. *et al.* Identification of herpesvirus transcripts from genomic regions around the replication origins. *Sci. Rep.* **13**, 16395 (2023).

34. Tikhanovich, I., Liang, B., Seoighe, C., Folk, W. R. & Nasheuer, H. P. Inhibition of human BK polyomavirus replication by small noncoding RNAs. *J. Virol.* **85**, 6930–6940 (2011).

35. 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).

36. Fülöp, Á. *et al.* Integrative profiling of Epstein-Barr virus transcriptome using a multiplatform approach. *Virol. J.* **19**, 7 (2022).

37. Kakuk, B. *et al.* Combined nanopore and single-molecule real-time sequencing survey of human betaherpesvirus 5 transcriptome. *Sci. Rep.* **11**, 14487 (2021).

38. Prazsák, I. *et al.* KSHV 3.0: A State-of-the-Art Annotation of the Kaposi's Sarcoma-Associated Herpesvirus Transcriptome Using Cross-Platform Sequencing. *BioRxiv Prepr. Serv. Biol.* 2023.09.21.558842 (2023) doi:10.1101/2023.09.21.558842.

39. Moldován, N. *et al.* Time-course profiling of bovine alphaherpesvirus 1.1 transcriptome using multiplatform sequencing. *Sci. Rep.* **10**, 20496 (2020).

40. Tombácz, D. *et al.* Hybrid sequencing discloses unique aspects of the transcriptomic architecture in equid alphaherpesvirus 1. *Heliyon* **9**, e17716 (2023).

41. Torma, G. *et al.* An Integrated Sequencing Approach for Updating the Pseudorabies Virus Transcriptome. *Pathog. Basel Switz.* **10**, 242 (2021).

42. Kakuk, B. *et al.* Nanopore Assay Reveals Cell-Type-Dependent Gene Expression of Vesicular Stomatitis Indiana Virus and Differential Host Cell Response. *Pathog. Basel Switz.* **10**, 1196 (2021).

43. Tombácz, D. *et al.* Time-Course Transcriptome Profiling of a Poxvirus Using Long-Read Full-Length Assay. *Pathog. Basel Switz.* **10**, 919 (2021).

44. Torma, G. *et al.* Dual isoform sequencing reveals complex transcriptomic and epitranscriptomic landscapes of a prototype baculovirus. *Sci. Rep.* **12**, 1291 (2022).

45. Torma, G. *et al.* Combined Short and Long-Read Sequencing Reveals a Complex Transcriptomic Architecture of African Swine Fever Virus. *Viruses* **13**, 579 (2021).

46. 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).

47. Dixon, L. K., Stahl, K., Jori, F., Vial, L. & Pfeiffer, D. U. African Swine Fever Epidemiology and Control. *Annu. Rev. Anim. Biosci.* **8**, 221–246 (2020).

48. Penrith, M.-L. History of ‘swine fever’ in Southern Africa. *J. S. Afr. Vet. Assoc.* **84**, 6 (2013).

49. Njau, E. P. *et al.* African Swine Fever Virus (ASFV): Biology, Genomics and Genotypes Circulating in Sub-Saharan Africa. *Viruses* **13**, 2285 (2021).

50. Boshoff, C. I., Bastos, A. D. S., Gerber, L. J. & Vosloo, W. Genetic characterisation of African swine fever viruses from outbreaks in southern Africa (1973–1999). *Vet. Microbiol.* **121**, 45–55 (2007).

51. Achenbach, J. E. *et al.* Identification of a New Genotype of African Swine Fever Virus in Domestic Pigs from Ethiopia. *Transbound. Emerg. Dis.* **64**, 1393–1404 (2017).

52. Andrés, G., Simón-Mateo, C. & Viñuela, E. Assembly of African swine fever virus: role of polyprotein pp220. *J. Virol.* **71**, 2331–2341 (1997).

53. 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.

54. Yáñez, R. J. *et al.* Analysis of the Complete Nucleotide Sequence of African Swine Fever Virus. *Virology* **208**, 249–278 (1995).

55. Dixon, L. K., Chapman, D. A. G., Netherton, C. L. & Upton, C. African swine fever virus replication and genomics. *Virus Res.* **173**, 3–14 (2013).

56. Chapman, D. A. G., Tcherepanov, V., Upton, C. & Dixon, L. K. Comparison of the genome sequences of non-pathogenic and pathogenic African swine fever virus isolates. *J. Gen. Virol.* **89**, 397–408 (2008).

57. de la Vega, I., González, A., Blasco, R., Calvo, V. & Viñuela, E. Nucleotide Sequence and Variability of the Inverted Terminal Repetitions of African Swine Fever Virus DNA. *Virology* **201**, 152–156 (1994).

58. 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).

59. Enjuanes, L., Carrascosa, A. L., Moreno, M. A. & Viñuela, E. Titration of African Swine Fever (ASF) Virus. *J. Gen. Virol.* **32**, 471–477 (1976).

60. 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).

61. Alejo, A., Matamoros, T., Guerra, M. & Andrés, G. A Proteomic Atlas of the African Swine Fever Virus Particle. *J. Virol.* **92**, 10.1128/jvi.01293-18 (2018).

62. Karger, A. *et al.* An Update on African Swine Fever Virology. *Viruses* **11**, 864 (2019).

63. Yoo, D., Kim, H., Lee, J. Y. & Yoo, H. S. African swine fever: Etiology, epidemiological status in Korea, and perspective on control. *J. Vet. Sci.* **21**, (2020).

64. Lithgow, P., Takamatsu, H., Werling, D., Dixon, L. & Chapman, D. Correlation of cell surface marker expression with African swine fever virus infection. *Vet. Microbiol.* **168**, 413–419 (2014).

65. Sánchez-Torres, C. *et al.* Expression of porcine CD163 on monocytes/macrophages correlates with permissiveness to African swine fever infection. *Arch. Virol.* **148**, 2307–2323 (2003).

66. Alcami, A. & Viñuela, E. Fc receptors do not mediate african swine fever virus replication in macrophages. *Virology* **181**, 756–759 (1991).

67. Andrés, G. African Swine Fever Virus Gets Undressed: New Insights on the Entry Pathway. *J. Virol.* **91**, 10.1128/jvi.01906-16 (2017).

68. Basta, S., Gerber, H., Schaub, A., Summerfield, A. & McCullough, K. C. Cellular processes essential for African swine fever virus to infect and replicate in primary macrophages. *Vet. Microbiol.* **140**, 9–17 (2010).

69. Muñoz-Moreno, R., Galindo, I., Cuesta-Geijo, M. Á., Barrado-Gil, L. & Alonso, C. Host cell targets for African swine fever virus. *Virus Res.* **209**, 118–127 (2015).

70. Sánchez, E. G., Pérez-Núñez, D. & Revilla, Y. Mechanisms of Entry and Endosomal Pathway of African Swine Fever Virus. *Vaccines* **5**, 42 (2017).

71. Cuesta-Geijo, M. A. *et al.* Endosomal Maturation, Rab7 GTPase and Phosphoinositides in African Swine Fever Virus Entry. *PLOS ONE* **7**, e48853 (2012).

72. Hernáez, B., Guerra, M., Salas, M. L. & Andrés, G. African Swine Fever Virus Undergoes Outer Envelope Disruption, Capsid Disassembly and Inner Envelope Fusion before Core Release from Multivesicular Endosomes. *PLOS Pathog.* **12**, e1005595 (2016).

73. González, A., Talavera, A., Almendral, J. M. & Viñuela, E. Hairpin loop structure of African swine fever virus DNA. *Nucleic Acids Res.* **14**, 6835–6844 (1986).

74. Cackett, G. *et al.* The African Swine Fever Virus Transcriptome. *J. Virol.* **94**, e00119-20 (2020).

75. Rodríguez, J. M. & Salas, M. L. African swine fever virus transcription. *Virus Res.* **173**, 15–28 (2013).

76. Broyles, S. S. Vaccinia virus transcription. *J. Gen. Virol.* **84**, 2293–2303 (2003).

77. 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).

78. 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).

79. 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).

80. 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. *J. Virol.* **96**, (2022).

81. Cackett, G., Sýkora, M. & Werner, F. Transcriptome view of a killer: African swine fever virus. *Biochem. Soc. Trans.* **48**, 1569–1581 (2020).

82. Jaing, C. *et al.* Gene expression analysis of whole blood RNA from pigs infected with low and high pathogenic African swine fever viruses. *Sci. Rep.* **7**, 10115 (2017).

83. Pellett, P. E. *et al.* Order - Herpesvirales. in 99 (Elsevier, 2012). doi:10.1016/B978-0-12-384684-6.00005-7.

84. 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).

85. Xiang, S. *et al.* Complete Genome Sequence of a Variant Pseudorabies Virus Strain Isolated in Central China. *Genome Announc.* **4**, 10.1128/genomea.00149-16 (2016).

86. Zheng, H.-H. *et al.* Seroprevalence investigation and genetic analysis of pseudorabies virus within pig populations in Henan province of China during 2018–2019. *Infect. Genet. Evol.* **92**, 104835 (2021).

87. Shope, R. E. EXPERIMENTS ON THE EPIDEMIOLOGY OF PSEUDORABIES : II. PREVALENCE OF THE DISEASE AMONG MIDDLE WESTERN SWINE AND

THE POSSIBLE RÔLE OF RATS IN HERD-TO-HERD INFECTIONS. *J. Exp. Med.* **62**, 101–117 (1935).

88. Tu, L. *et al.* Assessing the Risk of Commercial Vaccines Against Pseudorabies Virus in Cats. *Front. Vet. Sci.* **9**, (2022).
89. Wang, H.-H. *et al.* Typical gene expression profile of pseudorabies virus reactivation from latency in swine trigeminal ganglion. *J. Neurovirol.* **26**, 687–695 (2020).
90. Zheng, H.-H., Fu, P.-F., Chen, H.-Y. & Wang, Z.-Y. Pseudorabies Virus: From Pathogenesis to Prevention Strategies. *Viruses* **14**, 1638 (2022).
91. Boldogköi, Z. *et al.* Novel tracing paradigms—genetically engineered herpesviruses as tools for mapping functional circuits within the CNS: present status and future prospects. *Prog. Neurobiol.* **72**, 417–445 (2004).
92. Nakamura, K. *et al.* Identification of Sympathetic Premotor Neurons in Medullary Raphe Regions Mediating Fever and Other Thermoregulatory Functions. *J. Neurosci.* **24**, 5370–5380 (2004).
93. Pomeranz, L. E. *et al.* Gene Expression Profiling with Cre-Conditional Pseudorabies Virus Reveals a Subset of Midbrain Neurons That Participate in Reward Circuitry. *J. Neurosci.* **37**, 4128–4144 (2017).
94. 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).
95. Zhou, M. *et al.* Characterization of a moderately pathogenic pseudorabies virus variant isolated in China, 2014. *Infect. Genet. Evol.* **68**, 161–171 (2019).
96. Klupp, B. G., Hengartner, C. J., Mettenleiter, T. C. & Enquist, L. W. Complete, Annotated Sequence of the Pseudorabies Virus Genome. *J. Virol.* **78**, 424–440 (2004).

97. 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. *J. Gen. Virol.* **81**, 1539–1543 (2000).

98. 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).

99. 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).

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

101. 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).

102. 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).

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

104. Tombácz, D. *et al.* Full-Length Isoform Sequencing Reveals Novel Transcripts and Substantial Transcriptional Overlaps in a Herpesvirus. *PLOS ONE* **11**, e0162868 (2016).

105. Office International des Epizooties (OIE). https://biosecuritycentral.org/resource/core-guidance-and-recommendations/terrestrial-manual/ (2019).

106. Olasz, F. *et al.* Short and Long-Read Sequencing Survey of the Dynamic Transcriptomes of African Swine Fever Virus and the Host Cells. *Front. Genet.* **11**, (2020).

107. Olasz, F. *et al.* A Simple Method for Sample Preparation to Facilitate Efficient Whole-Genome Sequencing of African Swine Fever Virus. *Viruses* **11**, 1129 (2019).

108. Portugal, R. S., Bauer, A. & Keil, G. M. Selection of differently temporally regulated African swine fever virus promoters with variable expression activities and their application for transient and recombinant virus mediated gene expression. *Virology* **508**, 70–80 (2017).

109. Zsak, L. & Neilan, J. G. Regulation of Apoptosis in African Swine Fever Virus–Infected Macrophages. *Sci. World J.* **2**, 1186–1195 (2002).

110. Al, C., Mj, B. & P, de L. Methods for growing and titrating African swine fever virus: field and laboratory samples. *Curr. Protoc. Cell Biol.* **Chapter 26**, (2011).

111. C, H., Mj, B. & Al, C. The use of COS-1 cells for studies of field and laboratory African swine fever virus samples. *J. Virol. Methods* **164**, (2010).

112. Sánchez, E. G. *et al.* Phenotyping and susceptibility of established porcine cells lines to African Swine Fever Virus infection and viral production. *Sci. Rep.* **7**, 10369 (2017).

113. Bustos, M. J., Nogal, M. L., Revilla, Y. & Carrascosa, A. L. Plaque assay for African swine fever virus on swine macrophages. *Arch. Virol.* **147**, 1453–1459 (2002).

114. Tombácz, D. *et al.* Transcriptome-wide survey of pseudorabies virus using next- and third-generation sequencing platforms. *Sci. Data* **5**, 180119 (2018).

115. D, T. *et al.* Strain Kaplan of Pseudorabies Virus Genome Sequenced by PacBio Single-Molecule Real-Time Sequencing Technology. *Genome Announc.* **2**, (2014).

116. Li, H. Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics* **34**, 3094–3100 (2018).

117. Lex, A., Gehlenborg, N., Strobelt, H., Vuillemot, R. & Pfister, H. UpSet: Visualization of Intersecting Sets. *IEEE Trans. Vis. Comput. Graph.* **20**, 1983–1992 (2014).

118. Sessegolo, C. *et al.* Transcriptome profiling of mouse samples using nanopore sequencing of cDNA and RNA molecules. *Sci. Rep.* **9**, 14908 (2019).

119. Lee, S. *et al.* Global mapping of translation initiation sites in mammalian cells at single-nucleotide resolution. *Proc. Natl. Acad. Sci. U. S. A.* **109**, E2424-2432 (2012).

120. Balázs, Z. *et al.* Template-switching artifacts resemble alternative polyadenylation. *BMC Genomics* **20**, 824 (2019).

121. Prazsák, I. *et al.* Long-read sequencing uncovers a complex transcriptome topology in varicella zoster virus. *BMC Genomics* **19**, 873 (2018).

122. Almazán, F., Rodríguez, J. M., Angulo, A., Viñuela, E. & Rodriguez, J. F. Transcriptional mapping of a late gene coding for the p12 attachment protein of African swine fever virus. *J. Virol.* **67**, 553–556 (1993).

123. Crooks, G. E., Hon, G., Chandonia, J.-M. & Brenner, S. E. WebLogo: A Sequence Logo Generator. *Genome Res.* **14**, 1188–1190 (2004).

124. Tombácz, D., Tóth, J. S., Petrovszki, P. & Boldogkői, Z. Whole-genome analysis of pseudorabies virus gene expression by real-time quantitative RT-PCR assay. *BMC Genomics* **10**, 491 (2009).

125. Braspenning, S. E. *et al.* Decoding the Architecture of the Varicella-Zoster Virus Transcriptome. *mBio* **11**, e01568-20 (2020).

126. O'Grady, T. *et al.* Global transcript structure resolution of high gene density genomes through multi-platform data integration. *Nucleic Acids Res.* **44**, e145 (2016).

127. 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).

128. Tai-Schmiedel, J. *et al.* Human cytomegalovirus long noncoding RNA4.9 regulates viral DNA replication. *PLOS Pathog.* **16**, e1008390 (2020).

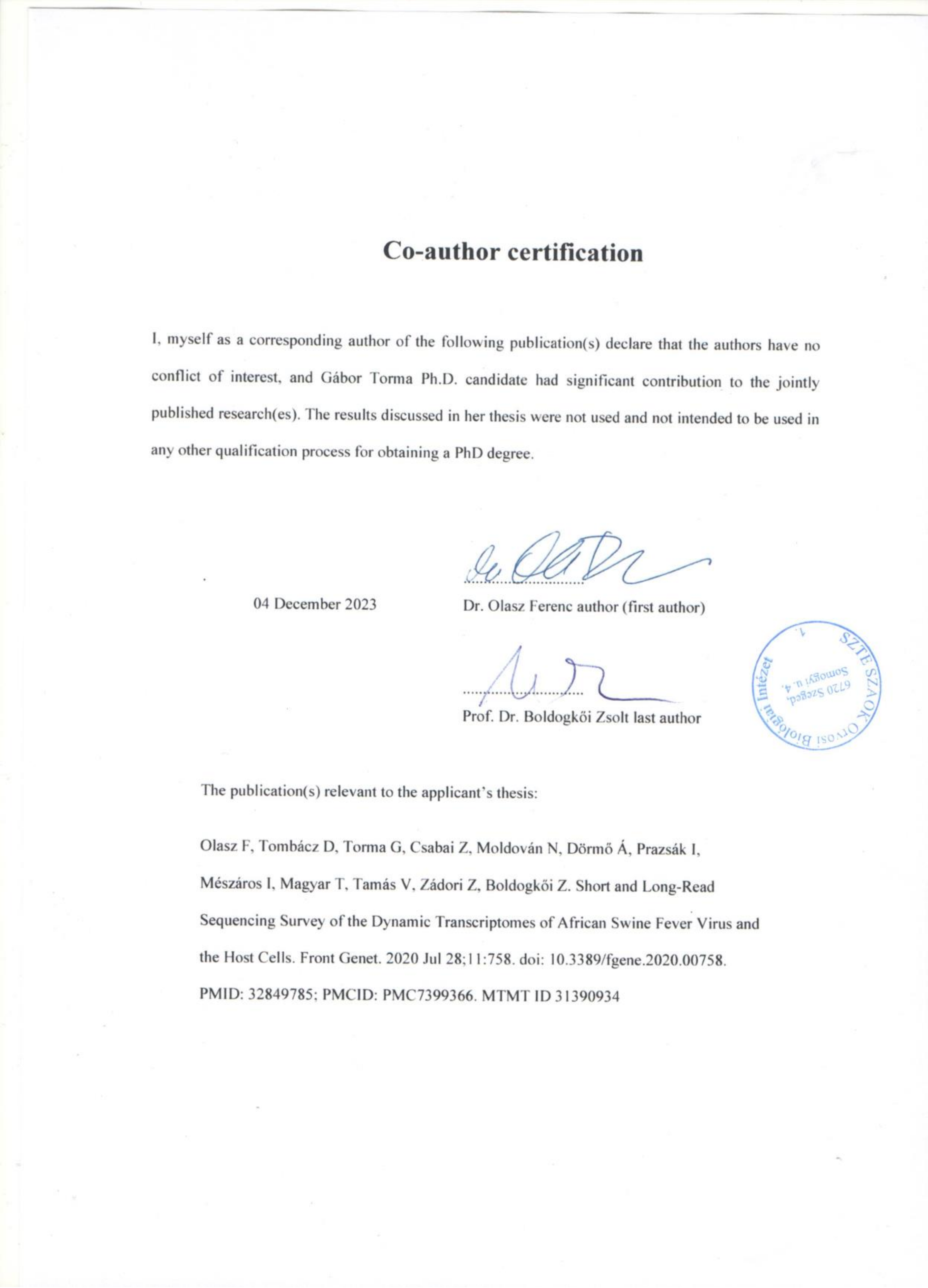
129. Kronstad, L. M., Brulois, K. F., Jung, J. U. & Glaunsinger, B. A. Dual Short Upstream Open Reading Frames Control Translation of a Herpesviral Polycistronic mRNA. *PLOS Pathog.* **9**, e1003156 (2013).

130. Vilela, C. & McCarthy, J. E. G. Regulation of fungal gene expression via short open reading frames in the mRNA 5'untranslated region. *Mol. Microbiol.* **49**, 859–867 (2003).

131. Proudfoot, N. J. Transcriptional termination in mammals: Stopping the RNA polymerase II juggernaut. *Science* **352**, aad9926 (2016).

132. Martens, J. A., Laprade, L. & Winston, F. Intergenic transcription is required to repress the Saccharomyces cerevisiae SER3 gene. *Nature* **429**, 571–574 (2004).

133. X, H. *et al.* Transcriptional interference among the murine beta-like globin genes. *Blood* **109**, (2007).



Co-author certification

I, myself as a corresponding author of the following publication(s) declare that the authors have no conflict of interest, and Gábor Torma Ph.D. candidate had significant contribution to the jointly published research(es). The results discussed in her thesis were not used and not intended to be used in any other qualification process for obtaining a PhD degree.

04 December 2023

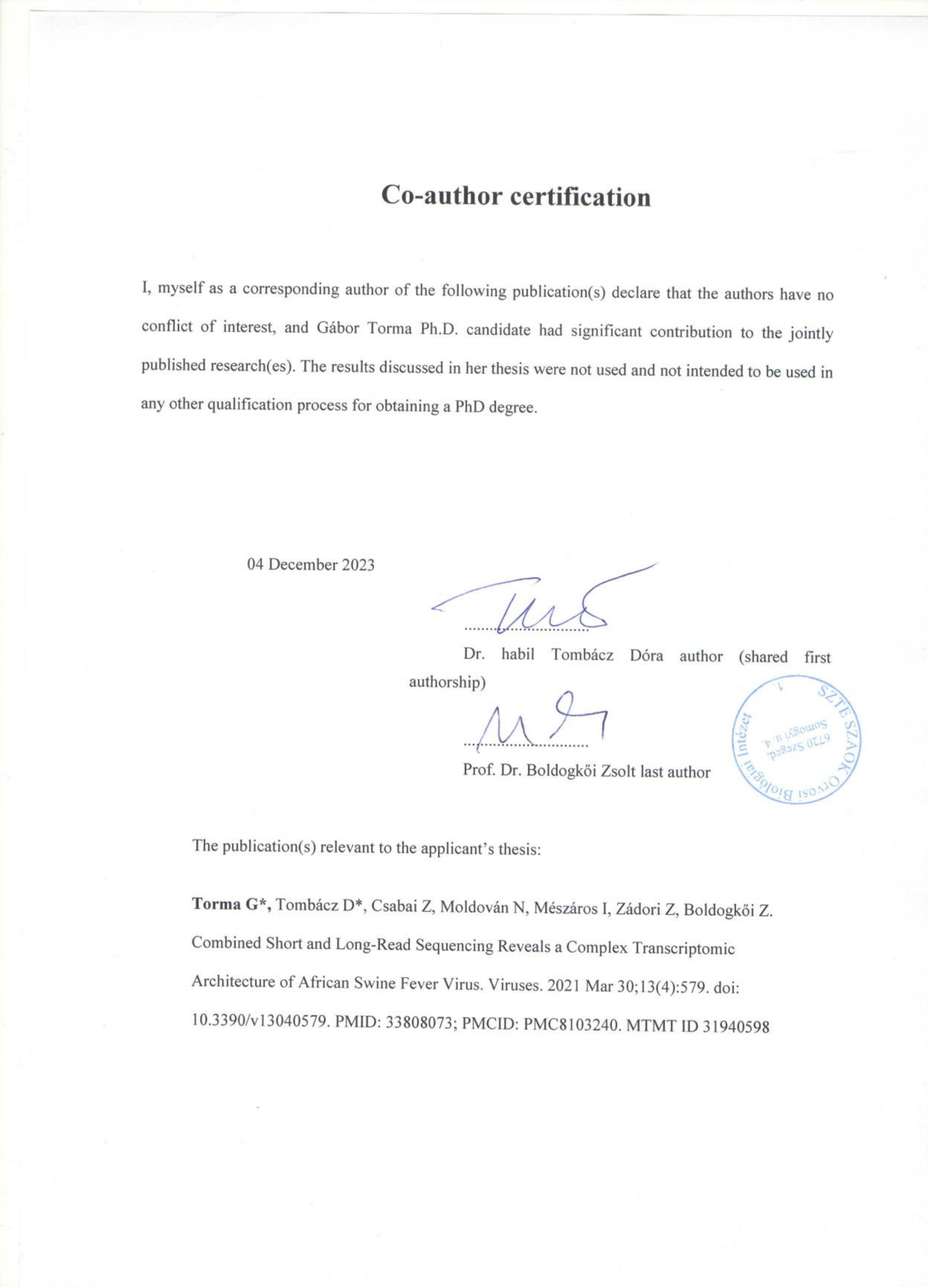
Dr. Olasz Ferenc author (first author)

Prof. Dr. Boldogkői Zsolt last author

SZTE SZAOK Orvosi Biológiai Intézet 1.
6720 Szeged, Somogyi u. 4.

The publication(s) relevant to the applicant's thesis:

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. PMID: 32849785; PMCID: PMC7399366. MTMT ID 31390934



Co-author certification

I, myself as a corresponding author of the following publication(s) declare that the authors have no conflict of interest, and Gábor Torma Ph.D. candidate had significant contribution to the jointly published research(es). The results discussed in her thesis were not used and not intended to be used in any other qualification process for obtaining a PhD degree.

04 December 2023

....

Dr. habil Tombácz Dóra author (shared first authorship)

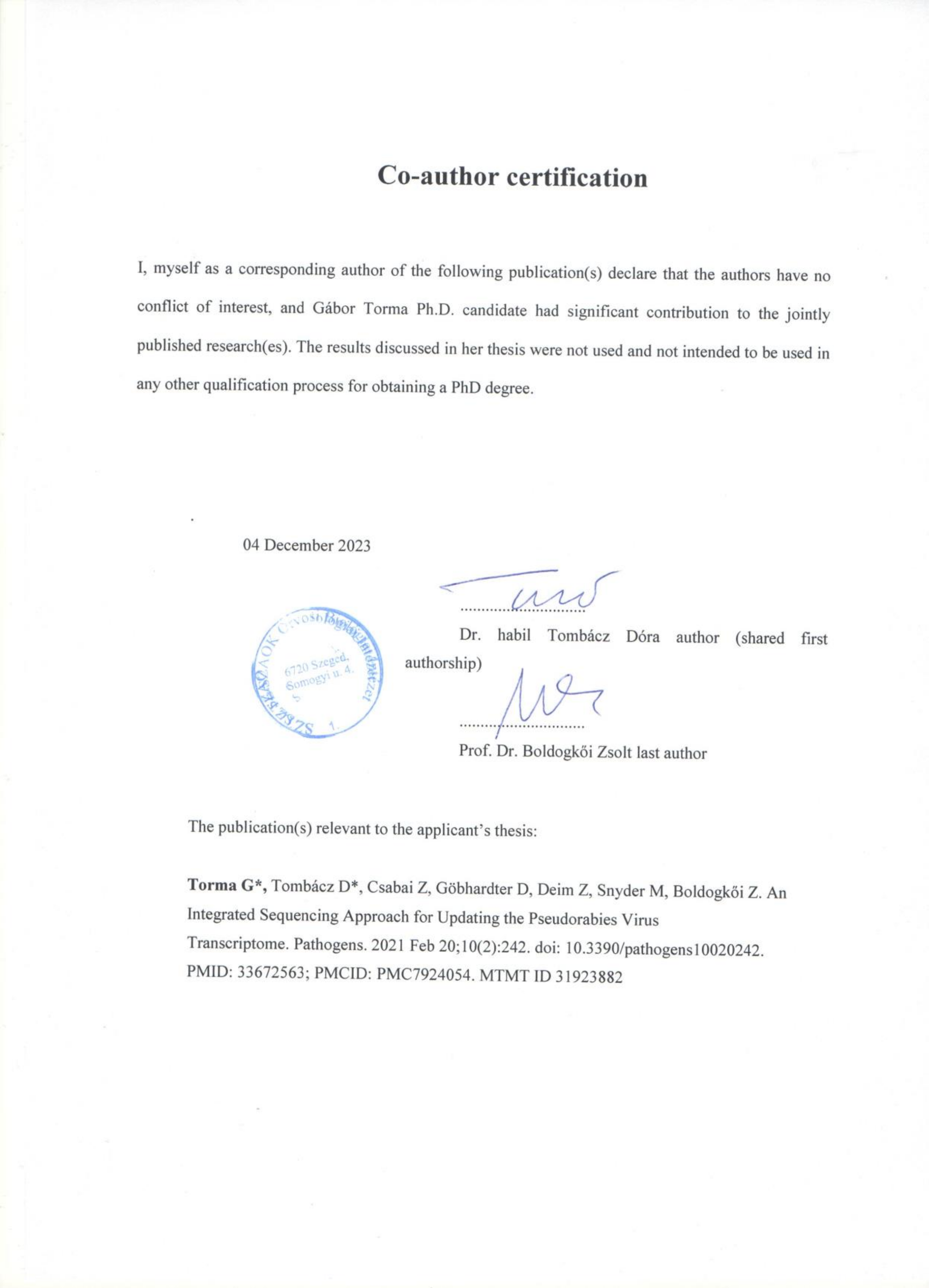
....

Prof. Dr. Boldogkői Zsolt last author

SZTE SZAOK Orvosi Biológiai Intézet 1. 6720 Szeged, Somogyi u. 4.

The publication(s) relevant to the applicant's thesis:

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. PMID: 33808073; PMCID: PMC8103240. MTMT ID 31940598



Co-author certification

I, myself as a corresponding author of the following publication(s) declare that the authors have no conflict of interest, and Gábor Torma Ph.D. candidate had significant contribution to the jointly published research(es). The results discussed in her thesis were not used and not intended to be used in any other qualification process for obtaining a PhD degree.

04 December 2023

....

Dr. habil Tombácz Dóra author (shared first authorship)

....

Prof. Dr. Boldogkői Zsolt last author

The publication(s) relevant to the applicant's thesis:

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. PMID: 33672563; PMCID: PMC7924054. MTMT ID 31923882