University of Szeged

Albert Szent-Györgyi Medical School

Doctoral School of Multidisciplinary Medical Sciences

Exploring the Transcriptome Profile of Caviid Gammaherpesvirus 1 using Long-Read Sequencing

PhD Thesis

Ákos Dörmő

Supervisors: Dr. habil. Dóra Tombácz and prof. Dr. Zsolt Boldogkői

Szeged,
2026

Table of contents

List of publications

Publication directly related to the subject of the thesis

1. Torma Gábor ; **Dörmő Ákos***; Fülöp Ádám ; Tombácz Dóra ; Mizik Máté ; Pretory Amanda M. ; Lee See-Chi ; Toth Zsolt** ;Boldogkői Zsolt
   Long-read transcriptomics of caviid gammaherpesvirus 1: compiling a comprehensive RNA atlas
   MSYSTEMS 10: 3 Paper: e0167824 , 19 p. (2025)
   Folyóirat szakterülete: Scopus - Ecology, Evolution, Behavior and Systematics SJR indikátor: D1

Publications not related to the subject of the thesis

2. Tombácz Dóra ; Kakuk Balázs ; Torma Gábor ; Fülöp Ádám ; **Dörmő Ákos** ; Gulyás Gábor ; Csabai Zsolt ; Boldogkői Zsolt
   Mapping the temporal transcriptomic signature of a viral pathogen through CAGE and nanopore sequencing
   PLOS ONE 20 : 4 Paper: e0320439 , 28 p. (2025)
   Folyóirat szakterülete: Scopus - Multidisciplinary SJR indikátor: Q1

3. Tombácz Dóra ; Maróti Zoltán* ; Oláh Péter* ; **Dörmő Ákos** ; Gulyás Gábor ; Kalmár Tibor ; Csabai Zsolt ; Boldogkői Zsolt
   Temporal transcriptional profiling of host cells infected by a veterinary alphaherpesvirus using nanopore sequencing
   SCIENTIFIC REPORTS 15 : 1 Paper: 3247 , 12 p. (2025)
   Folyóirat szakterülete: Scopus - Multidisciplinary SJR indikátor: Q1

4. Gulyás Gábor ; Kakuk Balázs* **; Dörmő Ákos*** ; Járay Tamás ; Prazsák István ; Csabai Zsolt ; Henkrich Miksa Máté ; Boldogkői Zsolt ; Tombácz Dóra
   Cross-comparison of gut metagenomic profiling strategies
   COMMUNICATIONS BIOLOGY 7 : 1 Paper: 1445 , 22 p. (2024)
   Folyóirat szakterülete: Scopus - Agricultural and Biological Sciences (miscellaneous) SJR indikátor: D1

5. Nagy Gergely Ármin ; Tombácz Dóra* ; Prazsák István ; Csabai Zsolt ; **Dörmő Ákos** ; Gulyás Gábor ; Kemenesi Gábor ; Tóth Gábor E. ; Holoubek Jiří ; Růžek Daniel et al.
   Exploring the transcriptomic profile of human monkeypox virus via CAGE and native RNA sequencing approaches
   MSPHERE 9 : 9 Paper: e00356-24 , 22 p. (2024)
   Folyóirat szakterülete: Scopus - Microbiology SJR indikátor: Q1

6. Prazsák István ; Tombácz Dóra* ; Fülöp Ádám* ; Torma Gábor ; Gulyás Gábor ; **Dörmő Ákos** ; Kakuk Balázs ; McKenzie Spires Lauren ; Toth Zsolt ; Boldogkői Zsolt
   KSHV 3.0: A State-of-the-Art Annotation of the Kaposi's Sarcoma-Associated Herpesvirus Transcriptome Using Cross-Platform Sequencing
   MSYSTEMS 9 : 2 Paper: e01007-23 , 19 p. (2024)

Folyóirat szakterülete: Scopus - Ecology, Evolution, Behavior and Systematics SJR indikátor: D1

7. Tombácz Dóra ; Torma Gábor* ; Gulyás Gábor* ; Fülöp Ádám ; **Dörmő Ákos** ; Prazsák István ; Csabai Zsolt ; Mizik Máté ; Hornyák Ákos ; Zádori Zoltán et al.
   Hybrid sequencing discloses unique aspects of the transcriptomic architecture in equid alphaherpesvirus 1
   HELIYON 9 : 7 Paper: e17716 , 16 p. (2023)
   Folyóirat szakterülete: Scopus - Multidisciplinary SJR indikátor: Q1

8. Torma Gábor ; Tombácz Dóra* ; Csabai Zsolt* ; Almsarrhad Islam A. A.* ; Nagy Gergely Ármin ; Kakuk Balázs ; Gulyás Gábor ; Spires Lauren McKenzie ; Gupta Ishaan ; Fülöp Ádám ; **Dörmő Ákos** ; Prazsák István et al.
   Identification of herpesvirus transcripts from genomic regions around the replication origins
   SCIENTIFIC REPORTS 13 : 1 Paper: 16395 , 25 p. (2023)
   Folyóirat szakterülete: Scopus - Multidisciplinary SJR indikátor: D1

9. Kakuk Balázs ; **Dörmő Ákos*** ; Csabai Zsolt ; Kemenesi Gábor ; Holoubek Jiří ; Růžek Daniel ; Prazsák István ; Dani Virág Éva ; Dénes Béla ; Torma Gábor et al.
   In-depth Temporal Transcriptome Profiling of Monkeypox and Host Cells using Nanopore Sequencing
   SCIENTIFIC DATA 10 : 1 Paper: 262 , 12 p. (2023)
   Folyóirat szakterülete: Scopus - Computer Science Applications SJR indikátor: D1

10. Tombácz Dóra ; **Dörmő Ákos** ; Gulyás Gábor ; Csabai Zsolt ; Prazsák István ; Kakuk Balázs ; Harangozó Ákos ; Jankovics István ; Dénes Béla ; Boldogkői Zsolt
    High temporal resolution Nanopore sequencing dataset of SARS-CoV-2 and host cell RNAs
    GIGASCIENCE 11 Paper: giac094 , 11 p. (2022)
    Folyóirat szakterülete: Scopus - Computer Science Applications SJR indikátor: D1

11. Maróti Zoltán ; Tombácz Dóra* ; Moldován Norbert* ; Torma Gábor ; Jefferson Victoria A. ; Csabai Zsolt ; Gulyás Gábor ; **Dörmő Ákos** ; Boldogkői Miklós ; Kalmár Tibor et al.
    Time course profiling of host cell response to herpesvirus infection using nanopore and synthetic long-read transcriptome sequencing
    SCIENTIFIC REPORTS 11 : 1 Paper: 14219 , 11 p. (2021)
    Folyóirat szakterülete: Scopus - Multidisciplinary SJR indikátor: D1

12. Olasz Ferenc ; Tombácz Dóra ; Torma Gábor ; Csabai Zsolt ; Moldován Norbert ; **Dörmő Ákos** ; Prazsák István ; Mészáros István; Magyar Tibor ; Tamás Vivien et al.
    Short and Long-Read Sequencing Survey of the Dynamic Transcriptomes of African Swine Fever Virus and the Host Cells
    FRONTIERS IN GENETICS 11 Paper: 758 , 7 p. (2020)
    Folyóirat szakterülete: Scopus - Genetics (clinical) SJR indikátor: Q2

List of abbreviations

asRNA	antisense RNA
AXP	AMPure XP beads
CaGHV-1	Caviid GammaHerpesVirus 1
E	Early
EBV	Eppstein-Barr Virus
GPHLV	Guinea Pig Herpes-Like Virus
gpRTA	Caviid gammaherpesvirus 1 RTA
HSV-1	Herpes Simplex Virus 1
HVS	HerpesVirus Saimiri
IE	Immediate-Early
KSHV	Kaposi's Sarcoma-associated HerpesVirus
L	Late
LANA	Latency-Associated Nuclear Protein
LoRTIA	Long-read RNA-Seq Transcript Isoform Annotator
MHV-68	Murine gammaherpes 68
miRNA	microRNA
ncRNA	non-coding RNA
NFW	Nuclease-Free Water
NHP	Non-Human Primate
ONT	Oxford Nanopore Technologies
ORF	Open Reading Frame
PAN RNA	PolyAdenylated Nuclear RNA
PAS	PolyAdenylation Signal
Poly(A) RNA	Polyadenylated RNA
pre-miRNA	precursor miRNA
raRNA	replication origin-associated RNA
RFHV	Retroperitoneal Fibromatosis-associated HerpesVirus
RRV	Rhesus macaque RhadinoVirus
RTA	Replication and Transcription Activator

TES	Transrciption End Site
TSS	Transcription Start Site
UTR	UnTranslated region
VERO	African green monkey kidney cells
vFLIP	viral FLICE -Inhibitory Protein
vIL-6	viral InterLeukin-6

1. Introduction

Viruses are the most common[1] non-cellular obligate intracellular microorganisms[2] on Earth, and reproduce without cell division. One of the largest taxonomic groups of viruses is the herpesviruses, which belong to the order *Herpesvirales*[3] and infect vertebrates and a marine mollusc, which were primarily identified based on the structure of their virions. Later, related viruses were classified into subfamilies based on antigens and biological properties[4]. With the advancement of technology, this was replaced by sequence-based comparison[4–6], which is used as the primary approach for the taxonomic and phylogenetic systematization of new viruses. Approximately 130 different virus species have been identified and classified into three different families: *Orthoherpesviridae*, *Alloherpesviridae*, and *Malacoherpesviridae*[6–8].

Members of *Orthoherpesviridae* (formerly *Herpesviridae*) are large 150–200 nm enveloped viruses with linear double-stranded DNA and a genome size of approximately 125–241 kbp[3,6,8–11], and based on their organization, direct or inverted repeats may have been formed within the gene or at the terminal end. The genome contains 70–170 genes[8]. The *Orthoherpesviridae* family currently includes 118 herpesvirus species, of which nine are human pathogens[6,12,13]. It is divided into three further subfamilies based on tropism, pathogenicity, and laboratory cultivation, namely *Alphaherpesvirinae*, *Betaherpesvirinae*, and *Gammaherpesvirinae*[3,6,10,12]. Approximately 90% of humans are infected with some type of herpesvirus[14]. All herpesviruses are characterized by a lytic and latent cycle[10,12,13]. The host cell was entered by the virion either through membrane fusion or endocytosis via the cell surface. The capsid was transported to the nuclear membrane, and the viral DNA was delivered into the cell nucleus. Here, the productive, or lytic cycle was initiated, during which three regulated gene expression cascades are distinguished: immediate-early (IE), early (E), and late (L). The lytic cycle is initiated by the expression of IE genes. These genes encode regulatory proteins. Viral DNA replication and the expression of late genes are required by E genes. Virion proteins are encoded by late genes[8,15]. Virus DNA synthesis is carried out through a rolling circle replication mechanism, followed by packaging into the capsid and release by exocytosis through the trans-Golgi network[6,8,15]. Primary infection is followed by a latent phase, during which no virion production occurs and the latent virus is expressed, with occasional reactivation[6,8,10].

1.1. Gammaherpesvirues

Latent infection is established in the host organism by viruses belonging to the *Gammaherpesvirinae* subfamily, similarly to other herpesviruses, followed by lifelong persistence in the nucleus of infected cells as episomes or through integration into host cell chromosomes. In gammaherpesviruses, latency is established in B or T cells, with the exception of *Bovine gammaherpesvirus* (BoHV-4), by which monocytes/macrophages are infected[16]. The establishment of latency, as well as the investigation of the effects of non-coding RNAs (ncRNAs) on lytic and latent infection, has represented a dynamically and rapidly growing research field, which investigates the persistence of gammaherpesviruses in human populations. This includes two oncogenic human pathogenic gammaherpesviruses, *Epstein-Barr virus* (EBV, or HHV-4), which belongs to the lymphocryptoviruses, and *Kaposi's sarcoma-associated herpesvirus* (KSHV, or HHV-8), which belongs to the rhadinoviruses[17–19]. Multiple proteins and non-coding RNAs are encoded by EBV and KSHV in order to counteract the immune processes of the host organism[17]. These viruses are classified as the World Health Organization's number one carcinogenic viruses[20,21]. Approximately 200,000 new cancer cases worldwide each year are attributed to these two viruses, representing a significant health burden, particularly in countries with limited healthcare infrastructure[22,23]. The viruses are primarily spread through saliva via oral transmission, with B-cells being infected first in the oropharynx and subsequently in the epithelial and submucosal lymphoid tissues of the tonsils[21,23]. Numerous lymphoproliferative and neoplastic diseases are caused by these viruses (Figure 1.)[18].

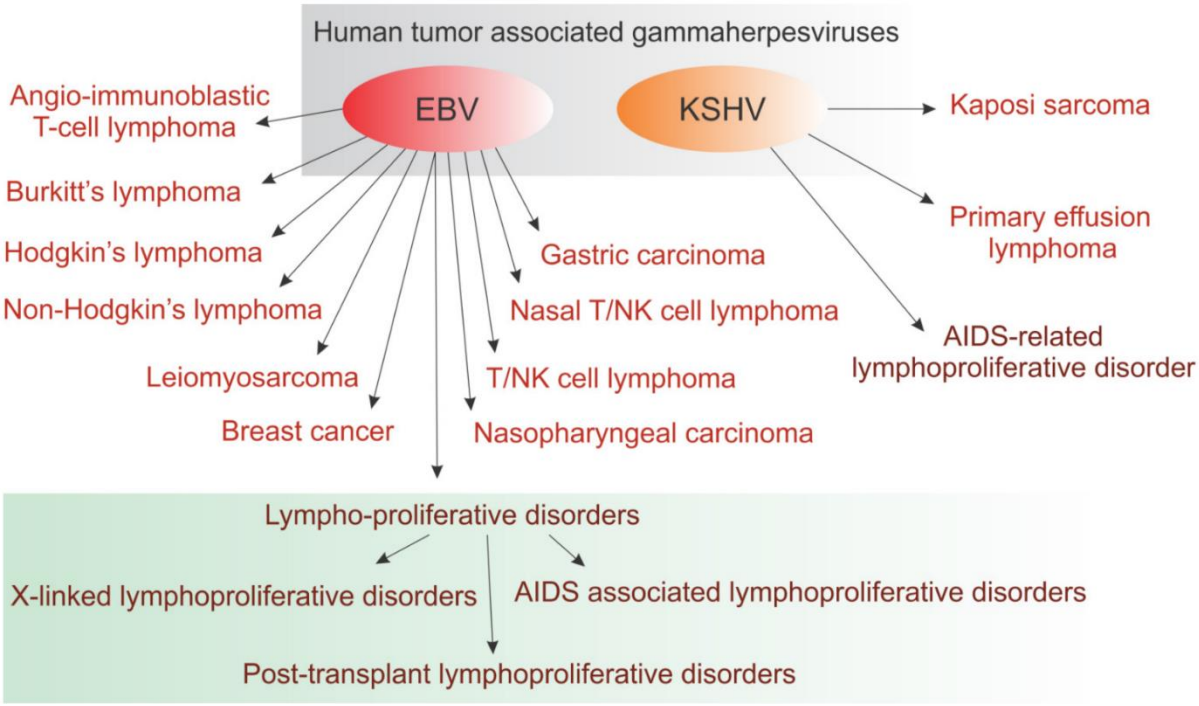


Figure 1. *The role of EBV and KSHV in malignant tumors*[18].

Several cancers, including Burkitt's lymphoma, nasopharyngeal carcinoma, localized or non-localized Hodgkin's lymphoma, and post-transplant lymphoproliferative disease (PTLD), have been linked to EBV. In addition, several non-malignant diseases, such as infectious mononucleosis, oral hairy leukoplakia, and X-linked immunodeficiency, are caused by EBV. Kaposi's sarcoma, primary effusion lymphoma, and Castleman's disease are caused by *human herpesvirus 8*[24].

A large number of cancers worldwide are caused by human gamma herpesviruses. However, no FDA-approved vaccine against them is currently available. Their effectiveness in humans is maintained by the use of different protein expression profiles in different cell types during different cell cycles[17]. A decisive role in tumor virology is played by oncogenic human gamma herpesviruses and B-cell differentiation. The understanding of these diseases is supported by the study of virus-host interactions and the immune responses that drive pathogenesis[23].

1.1.1. Kaposi's sarcoma-associated herpesvirus (KSHV)

Kaposi's sarcoma-associated herpesvirus (KSHV), also known as *human herpesvirus 8* (HHV-8), is classified as an oncogenic gammaherpesvirus that has been linked to three types of neoplastic diseases. Kaposi's sarcoma, which is described as an inflammatory, heterogeneous endothelial cell tumor, is recognized as one of the most common types of cancer in HIV-positive

individuals[25]. Primary effusion lymphoma, which is defined as a B-cell lymphoma affecting body cavities, and multicentric Castleman's disease, which is characterized as a lymphoproliferative disorder causing pathological changes in the lymph nodes, are also associated with KSHV infection[26–30]. The seroprevalence of the virus is observed to vary by geographic region, with the highest rates being reported in sub-Saharan Africa[26,29–32]. Due to the reduced immune function caused by HIV infection, Kaposi's sarcoma remains the most common cancer in sub-Saharan Africa, as well as in the United States and Europe[24].

Kaposi's sarcoma-associated herpesvirus is described as a large double-stranded DNA virus with a genome length of approximately 165 kb and around 90 open reading frames (ORFs), by which B lymphocytes are infected and lifelong infection is established following latency. As is typical for herpesviruses, a replicative lytic and a quiescent latent life cycle are also exhibited by KSHV[28,29,31,33,34]. During latency, the linear genome is circularized and is maintained in the cell nucleus as an episome. It has been demonstrated that several transcripts are expressed in latently infected cells, including the latency-associated nuclear protein (LANA)[35] encoded by ORF73, vCyclin encoded by ORF72, the viral FLICE protein (vFLIP) encoded by ORF71, Kaposin encoded by ORFK12, and microRNAs (miRNAs) encoded by certain precursor miRNAs (pre-miRNAs)[28–31,34,36] as well as the K1 and K15 genes encoding transmembrane proteins, and viral IL-6 (vIL-6). Since latent genes are expressed in most infected tumor cells, their contribution to tumorigenesis is assumed[30]. An extremely important role in the establishment of latency is played by LANA. The binding of the viral episome to the host chromosome in the cell nucleus is mediated by LANA, thereby allowing replication to occur together with the host genome during normal cell division[29,30,33]. Numerous other functions are attributed to it, such as the inhibition of tumor suppressors and the modulation of the host immune response[29], and a transcriptional regulatory role is played in the inhibition of the expression of the replication and transcription activator (RTA) protein. In addition, together with v-Cyclin and v-FLIP, oncogenic processes and cell proliferation are influenced. During the latent cycle, miRNAs encoded by precursor miRNAs (pre-miRNAs) participate in the regulation of the viral cycle and the cell cycle by post-transcriptionally inhibiting the expression of the RTA protein, thereby promoting viral persistence[36].

The latency phase is exited by the virus in response to various stimuli, and entry into the replication phase is achieved, leading to the production of new infectious virus particles and the lysis of the infected cell[36]. Similar to other herpesviruses, an orderly cascade of gene expression regulation is followed during the lytic cycle of KSHV, which is classified into three kinetic

classes, namely immediate-early (IE), early (E), and late (L) genes. The immediate-early RTA is encoded by the ORF50 gene and is responsible for the initiation of the lytic cycle. Binding of the RTA protein to the RBP-Jκ transcription factor is mediated, and upon activation, its own promoter is activated, resulting in the expression of sufficient levels of RTA required to maintain the lytic cycle[34,36]. One of the most abundant transcripts encoded by ORF-K7, a long non-coding RNA referred to as polyadenylated nuclear RNA (PAN RNA), is suggested to play a role in viral reactivation[34,36,37]. The early genes are required for DNA replication and gene expression and are encoded by ORF9, ORF21, and ORF36, which encode DNA polymerase, viral thymidine kinase, viral phosphotransferase, and bZIP, the viral interferon regulatory factor (vIRF-1), viral IL-6 (v-IL-6), virus-encoded chemokines (v-CCL), and the viral G-protein-coupled receptor. This is followed by the expression of late genes, during which virus particles and viral maturation proteins are produced. This is followed by viral assembly in the cell nucleus. The genome is incorporated into the capsids, a tegument is subsequently acquired in the cytoplasm, an envelope is formed through the host membrane, and escape from the host cell is achieved[36].

Since KSHV was first identified, substantial advances have been made in the understanding of the virus's mechanisms[38]. However, no suitable animal model for KSHV is currently available. All existing models are limited to representing only parts of the disease and do not recapitulate the entire infection cycle. Therefore, the creation of an inexpensive and easily manipulated model is considered important, as it would facilitate a better understanding of pathogenesis and enable the further development and testing of antiviral therapies[24].

1.2. Model organisms of KSHV

Currently, no perfect model for human viruses is available; only one aspect of the disease is represented, but not the entire infection cycle[24]. Humans are regarded as the only natural hosts of KSHV gammaherpesviruses[24,39,40]. Due to the narrow host specificity of the virus, the creation of animal models is greatly complicated. Several different approaches are being used to overcome these limitations and to establish animal models[39,40].

First, non-human primates (NHPs) were experimentally infected with human pathogenic gammaherpesvirus[39]. Rhesus macaques were shown to be infectable with KSHV, however, replication was observed to be very weak and no detectable viral gene expression was annotated. In contrast, common marmosets (*Callithrix jacchus*) were found to be susceptible to

KSHV infection following inoculation and Kaposi's sarcoma-like lesions were observed. In addition, an antibody response and LANA expression were detected, and viral DNA was identified in peripheral blood mononuclear cells and various tissues[24,38–41].

In the second study, laboratory animals that had been naturally infected with the virus, or with gammaherpesviruses related to KSHV infecting NHPs, were examined. These viruses are characterized by a genome structure and properties similar to those of the human pathogenic virus. Mouse models are also included in this group[39,42]. The rhadinovirus subgroup of gammaherpesviruses has been divided into two lineages. One lineage is rhadinovirus 1 (RV1), which includes KSHV itself and an extinct mammalian virus referred to as *retroperitoneal fibromatosis herpesvirus* (RFHV), which causes retroperitoneal fibromatosis and was isolated from Kaposi's sarcoma-like tumors in macaques with monkey AIDS[24,43,44]. Similarity to Kaposi's sarcoma is observed in that cell proliferation is typically spindle-shaped. Conserved genes, such as ORF73 (LANA), which is responsible for the maintenance of latency, are encoded by both RFHV and KSHV, and additional genes involved in cell cycle regulation have also been identified. Currently, this virus is considered to be the most closely related to KSHV; however, RFHV has not been successfully cultured in vitro, which makes its use as a viral model difficult[39,40,43]. The other lineage is rhadinovirus 2 (RV2)40[39], which is composed of *rhesus rhadinovirus* (RRV) and *herpesvirus saimiri* (HVS)[24]. An aggressive infection is exhibited by HVS in New World non-human primates (NHPs), resulting in fulminant lymphoma. However, it is not considered a suitable model, because T cells are infected instead of B cells, which are infected by KSHV[38]. RRV is naturally harbored by rhesus macaques. Two strains have been sequenced, RRV-H26-95 and RRV-17577. The former strain has not been associated with any disease[45], whereas the 17577 strain, whose genome is highly homologous to that of KSHV, although to a lesser extent than the previously described RFHV, has been linked to disease[39]. Although RRV is widely distributed in macaque populations, disease is rarely caused. However, under conditions induced by *simian immunodeficiency virus* (SIV), strain 17577[41,46,47] has been associated with B-cell lymphoma and Kaposi's sarcoma-like lesions[45].

Murine gammaherpesvirus 68 (MHV-68, *Murid gammaherpesvirus 4*) is regarded as the murine homologue of KSHV[48]. Efficient replication is achieved in cell culture, infection is readily established, and latency is formed in B cells[24,40,49,50]. No visible symptoms are observed in immunocompetent mice; however, lymphoproliferative disease and B-cell lymphoma can be induced in immunodeficient mice. The greatest advantage and success of these models is

attributed to the possibility of introducing mutations into the viral genome, thereby enabling the investigation of tissue tropism, immune responses to MHV-68, or latency in the absence of specific genes in the mouse[24,45].

Thirdly, the group of humanized mice is described, in which immunodeficient mice are implanted with functional human tissues or cells[39]. Immunodeficient strains have been developed that serve as acceptors for the implantation of human cells[51]. SCID (severe combined immunodeficiency) mice, and subsequently NOD/SCID mice, were first used for the establishment of KSHV infection, either as a model of persistent infection in vivo or, more recently, through infection of BLT-NSG mice (bone marrow, liver, and thymus), which allowed the detection of lytic and latent viral DNA transcripts by measurement in different tissues[40,51–53].

In summary, non-human primates are considered the most suitable models for the study of KSHV pathogenesis, as KSHV-like diseases, such as multicentric Castleman's disease, B-cell lymphoma, and persistent infection, are caused by infection with RRV, and RFHV is regarded as the virus most homologous to KSHV. However, the maintenance of these models is extremely costly, and the number of infected animals is limited; therefore, alternative approaches must be sought. Although insight into the pathogenesis of KSHV can be provided by MHV68, it is not necessarily considered an ideal model for the study of KSHV-associated cancers, as lymphoma development is rare. Lymphoproliferative disease can be induced as a result of MHV68 infection; however, in humans, multicentric Castleman disease is a condition associated with HIV infection. Humanized mouse models are also not suitable for the complete reproduction of cancerous lesions[54]. Current models are characterized by limitations that hinder a comprehensive understanding of the virus and the diseases it causes[54]. Therefore, further alternatives are required that are inexpensive to maintain, easy to manipulate, and suitable for the testing of antiviral or anticancer agents[24]. To this end, the investigation of *Caviid gammaherpesvirus 1* as a potential new model organism was initiated.

1.3. *Caviid gammaherpesvirus 1* (CaGHV-1)

Caviid gammaherpesvirus 1, formerly known as *Guinea pig herpes-like virus* (GPHLV), was isolated from the buffy coat layer of strain 2 guinea pigs, in which acute lymphoblastoid leukemia developed spontaneously[54–56]. This virus was shown to replicate efficiently in rabbit cells[54,56]. This indicated that a relatively narrow host specificity was exhibited; however, later

attempts were successful in broadening the host tropism of the virus, allowing infection in vitro in rat cells, African green monkey kidney cells (VERO), mink cells, and cat cells[54,57]. However, with the emergence of MHV68, research on GPHLV was brought to an end, as *murine herpesvirus 68* became established as the new model organism for gammaherpesviruses, and, to date, only comparative analyses have been conducted to describe the genetic differences between *Guinea pig herpes-like virus*, GPCMV, and endogenous herpesviruses isolated from guinea pigs[54,58]. Therefore, the complete genome sequence of GPHLV was revealed and its phylogenetic affiliation was explored by Stanfield and colleagues (hereinafter, the currently accepted nomenclature, CaGHV-1, is used for the virus).

For virus propagation, VERO cells infected with the GPHLV LK-40 virus were used. Following bioinformatic processing of the sequencing data, phylogenetic analyses demonstrated that the virus is classified within the genus Rhadinovirus of the subfamily *Gammaherpesvirinae*[54]. After complete genome sequencing, de novo assembly revealed that the genome size of *Caviid gammaherpesvirus 1* is 103,374 base pairs (bp) (GenBank: OQ679822.1), and that its GC content is 35.45%. Prediction was performed and, based on these predictions, 75 open reading frames (ORFs) were identified, of which 63 exhibit sequence homology with KSHV, which belongs to the human pathogenic gammaherpesviruses (Figure 2)[54,59].

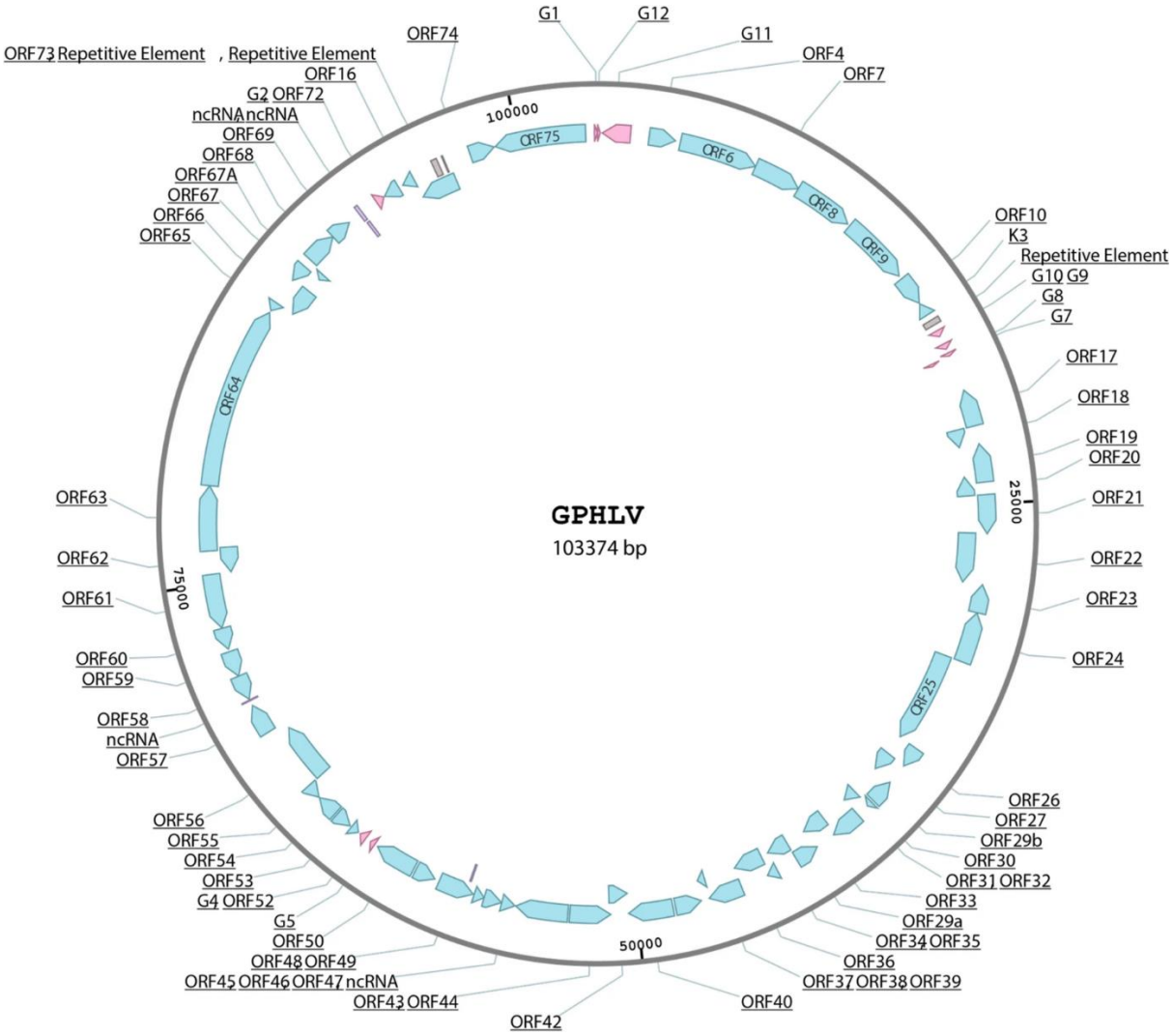


Figure 2. CaGHV-1 (previously known as GPHLV) genome. *The genome size is 103,374 (bp). The ORFs are highlighted in color, with genes homologous to KSHV shown in blue and virus-specific genes shown in pink*[54].

The predicted ORFs were annotated using the NCBI Herpesviridae non-repetitive database. Eighty-four percent of the previously described ORFs were shown to have significant relevance to KSHV. The CaGHV-1 proteome was compared to the proteomes of viruses currently used as model organisms for KSVH. These included RFHV, RRV, and MHV-68. The results demonstrated that the *Caviid gammaherpesvirus 1* proteome exhibited significantly greater homology with KSHV than MHV-68, which has been used for the past 30 years (Figure 3)[54].

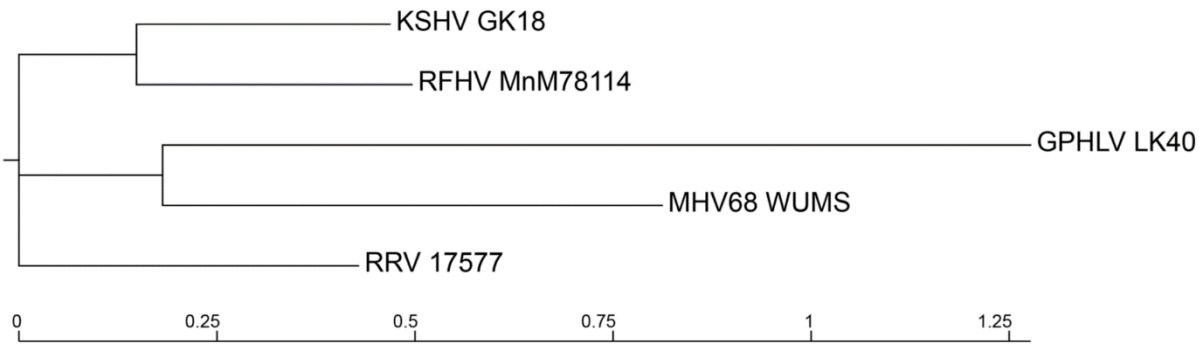


Figure 3. Phylogenetic analysis of KSHV model organism. *Genetically, RFHV is closest to KSHV, and the GPHLV LK-40 virus strain is a closer relative to MHV-68 than KSHV, but if we consider the distance between KSHV and CaGHV-1 and murine gammaherpesvirus 68, then caviid gammaherpesvirus shows greater homology in the genome than MHV-68*[54].

This is particularly evident for lytic antigens, in which the B glycoprotein (gB), H glycoprotein (gH), and L glycoprotein (gL), encoded by the ORF8, ORF22, and ORF47 genes, respectively, are shown to display a level of conservation comparable to that observed for gL in RRV in CaGHV-1. It can therefore be assumed that *Caviid gammaherpesvirus 1* is able to maintain the tissue tropisms and entry mechanisms identified in RRV. Similarity to KSHV is extended to most genes, including LANA and v-cyclin, which function as oncogenic factors, as well as RTA. Evidence is thus provided that the characterization and further investigation of the molecular mechanisms of CaGHV-1 as a relevant model organism are warranted[54].

1.4. Long-read sequencing in transcriptomics

The transcriptome is defined as the total amount of RNA present in a cell at a given time, process, or developmental stage. Its understanding is considered to be extremely important for the characterization of functional elements of the genome, development and disease, and molecular components. Its main purpose is defined as the categorization of all RNA species, including the determination of mRNAs, ncRNAs, small RNAs, transcript ends, splicing events, and gene expression[60]. RNA sequencing methods are now routinely applied. Although next-generation sequencing (NGS) plays an extremely important role in transcriptomics, short-read RNA sequencing cannot be applied adequately for transcriptome assembly, splice isoform annotation, and the annotation of novel genes, as most eukaryotic mRNAs are approximately 1–2 kb in length[61]. Bioinformatic methods can be applied for transcript assembly, however, transcript isoforms derived from the same expressed gene cannot be reliably identified due to

size limitations. In addition, various biases have been identified, arising from PCR amplification, GC content[62], or transcript quantification[63]. To overcome these errors and distortions, single-molecule third-generation long-read sequencing technologies, such as Pacific Biosciences (PacBio) and Oxford Nanopore Technologies (ONT), have been developed. Read lengths of approximately 15 kb for PacBio and more than 30 kb for ONT are achieved by these technologies[61,63,64]. Full-length sequencing of native RNA or complementary DNA (cDNA) without amplification is enabled by ONT nanopore sequencing[63,65]. Features such as differences in transcription start sites (TSSs) and poly(A) sites, as well as modifications of RNA molecules, are identified within a single read, thereby enabling the differentiation of highly similar transcripts. This has greatly expanded current knowledge of the transcriptome[65].

Based on nanopore sequencing, the pores are used as biosensors. These pores are embedded in a biomembrane, by which a channel is provided between the two sides of the lipid membrane. The pores were originally derived from α-hemolysin, which has a diameter of 2.6 nm. Currently, an MspA octamer is used by Oxford Nanopore Technologies, which has an inner diameter of 1 nm and represents a more stable and narrower pore suitable for higher single-nucleotide resolution. The cis ionic solution on one side of the membrane is brought into contact with the trans ionic solution through the nanopore. To ensure a stable electric field, electrodes are positioned on both sides of the sequencer. The detection mechanism itself is connected to a patch-clamp amplifier to record the ionic current signal, which has been miniaturized by ONT into portable ASIC chip systems. An extremely important role in the regulation of translocation speed is played by motor proteins. If the nucleic acid is translocated through the nanopore too rapidly, the passing nucleotide cannot be detected; however, if translocation occurs too slowly, excessive time is required for signal acquisition[66–68].

Long-read sequencing technologies have already been applied for the investigation of viral transcriptomes[69]. By this technology, previously unexpected complexities in the transcriptional landscape have been revealed, including the identification of transcript isoforms, novel splice variants, and transcriptional overlaps[70,71].

2. Aims

Although the genome of Caviid gammaherpesvirus 1 (CaGHV-1) has been sequenced, its transcriptome has not yet been characterized. Therefore, the aim of this study was to map the CaGHV-1 transcriptome and to annotate viral gene regulatory regions, thereby supporting the use of CaGHV-1 as a potential model organism for the investigation of human gammaherpesviruses.

The specific objectives of the study were:

1. To map the CaGHV-1 transcriptome using long-read sequencing approaches, including direct cDNA sequencing and native RNA sequencing, and to detect full-length viral RNA molecules.

2. To identify transcription start sites (TSSs) and transcription end sites (TESs) and to annotate regulatory elements controlling viral gene expression at nucleotide resolution.

3. To identify canonical transcripts, antisense RNAs, and non-coding RNAs, and to analyze transcriptional overlaps and overall transcriptome complexity.

4. To investigate the transcriptional activity of the ORF50-encoded replication and transcription activator (RTA) and to assess whether the CaGHV-1 ORF50 promoter can be activated by RTA proteins from other gammaherpesviruses.

3. Materials and methods

3.1. Cells and virus infection

Fibroblast 104C1 cells derived from guinea pig embryonic tissue [(CRL-1405) ATCC] and HEK293T cells [(Human Embryonic Kidney 293 cells SV40 T-antigen) ATCC] were cultured in RPMI-1640 and DMEM (Dulbecco's Modified Eagle Medium), respectively, supplemented with 10% FBS (Fetal Bovine Serum) and penicillin/streptomycin. CaGHV-1 (ATCC) was propagated in the 104C1 cell line, after which virions were concentrated by ultracentrifugation. Subsequently, 2×10^5 cells were infected with CaGHV-1 at eight different time points (4 h, 8 h, 16 h, 24 h, 48 h, 72 h, 96 h, and 120 h).

3.2. Cloning and luciferase reporter assay

Following PCR amplification of the ORF50 protein-coding sequence of *Caviid gammaherpesvirus 1*, the amplicon was cloned into the pCDH-CMV-MCS-EF1-puro expression vector using In-Fusion cloning (Takara). The N-terminally 3×FLAG-tagged CaGHV-1 RTA (Replication and Transcription Activator) was expressed as a protein in HEK293T cells by PEI (polyethylenimine) (Polysciences)-mediated transfection. ORF50 promoter fragments were cloned into the pGL4.15 luciferase reporter vector (Promega) using In-Fusion cloning following amplification. For reporter assays, 100 ng of reporter plasmid was transfected together with 400 ng of the RTA expression plasmid. Increasing amounts of expression plasmid (50 ng, 100 ng, 200 ng, and 400 ng) were used for transfection in order to evaluate the effects of varying RTA levels. The experiments were performed as previously described.[72]

3.3. RNA Isolation and Poly(A) selection

Total RNA was isolated using TRIzol™ Reagent (Invitrogen), modified and combined with the RNeasy® Kit (Qiagen). Following removal of the growth medium, 1 ml of TRIzol™ reagent was added to the infected cells, and after homogenization of the lysate, incubation was performed for 5 minutes. After lysis, 200 µl of chloroform was added, and the samples were thoroughly mixed by shaking. Following incubation for 2-3 minutes, samples were centrifuged for 15 minutes at 12,000 × g at 4°C, resulting in separation into an upper aqueous phase, an interphase, and a lower phenol–chloroform phase. For RNA purification, the RNeasy® protocol

was applied. Absolute ethanol was added to the upper aqueous phase containing RNA at a 1:1 ratio and mixed by pipetting. The samples were then transferred to RNeasy spin columns placed in 2 ml collection tubes and centrifuged for 15 seconds at ≥ 8,000 × g. Genomic DNA was digested using DNase I, followed by two additional washing steps in which 500 µl of RPE buffer was added to the column-bound RNA and centrifuged for 15 seconds at ≥ 8,000 × g. Total RNA was subsequently eluted from the column according to the previously described centrifugation steps, and RNA concentration was measured. Samples were stored at -80°C until further processing. Following quantitative and qualitative assessment, polyadenylated RNA was isolated using the Poly(A) RNA Selection Kit V1.5 (Lexogen) with a minor modification, whereby the maximum input RNA amount was increased from 5 µg to 10 µg, and all steps were proportionally scaled to the increased input amount in order to achieve a higher poly(A) RNA yield. After washing of the oligo(dT)-containing magnetic beads (MB) and elution in Hybridization Buffer (HYB), total RNA samples were denatured at 60°C for 1 minute in a thermocycler to disrupt secondary structures and enhance hybridization efficiency. Denatured RNA samples were mixed with pre-washed magnetic beads at a 1:1 ratio and incubated in a thermomixer at 25°C for 20 minutes at 1,250 rpm, allowing hybridization of 3′-polyadenylated RNAs to the oligo(dT) beads. Following incubation, samples were placed on a magnetic rack, and after clarification of the supernatant, it was removed and discarded. Beads were resuspended in 100 µl of Bead Wash Buffer (BW) and mixed for 5 minutes in a thermoshaker at 25°C and 1,250 rpm, followed by magnetic separation. This washing step was repeated twice, resulting in the removal of RNAs lacking poly(A) tails. Beads were subsequently resuspended in 12 µl of RNase-free water and incubated for 1 minute at 70°C to elute RNA from the beads. After placement on a magnetic rack, the purified eluates were transferred to new Eppendorf tubes. RNA concentration was measured, and samples were stored at -80°C until further downstream applications.

3.4. Direct RNA sequencing

The Oxford Nanopore Technologies (ONT) Direct RNA sequencing SKQ-RNA004 kit was used for sequencing native RNA molecules. During sequencing library preparation, RNA pooled from three biological replicates at eight time points was ligated to the Reverse Transcription Adapter (RTA) by the addition of T4 DNA Ligase (2M U/ml) (NEB) and NEBNext® Quick Ligation Reaction Buffer 5X (NEB), and RNaseOUT™ Recombinant Ribonuclease Inhibitor (Invitrogen) was added to inhibit RNase activity, followed by

incubation for 10 minutes at room temperature. Subsequently, the master mix required for reverse transcription was prepared using 10 mM dNTPs (NEB), 5X First Strand Buffer (Thermo Fisher Scientific), and 0.1 M DTT (Thermo Fisher Scientific). The RNA with the previously ligated RT Adapter was mixed with the master mix, after which SuperScript™ III Reverse Transcriptase (Thermo Fisher Scientific) was added for cDNA synthesis. The reaction was incubated in a thermal cycler at 50°C for 50 minutes, followed by incubation at 70°C for 10 minutes to terminate the reaction. Agencourt RNAClean XP beads (Beckman Coulter™) were added to the RNA–cDNA hybrid, after which the samples were placed on a magnetic rack, washed with 70% ethanol, and eluted in Nuclease-Free Water (NFW). In the subsequent step, the purified RNA–cDNA hybrid was ligated by the addition of the RNA Ligation Adapter (RLA), NEBNext Quick Ligation Reaction Buffer, and T4 DNA Ligase. Following the enzymatic reaction, the samples were washed with Wash Buffer (WSB) and eluted in RNA Elution Buffer. Finally, after concentration measurement, the prepared sequencing library was sequenced on a PromethION P2 Solo device using an RNA flow cell (FLO-PRO004RA).

3.5. Direct cDNA sequencing

For direct cDNA sequencing, the ONT Ligation Sequencing V14-Direct cDNA Sequencing Kit (SQK-LSK114) was used in combination with the Ligation Sequencing gDNA-Native Barcoding Kit 24 V14 (SQK-NBD114.24), as no barcode kit is provided by the manufacturer for sample separation during sequencing with the cDNA kit. RNAs derived from three parallel samples at eight time points (4 h, 8 h, 16 h, 24 h, 48 h, 72 h, 96 h, 120 h) were incubated for 5 minutes at 65°C in a thermocycler with 10 mM dNTPs and 2 µM VN Primer, which binds to the RNA poly(A) tail and initiates first-strand cDNA synthesis, followed by placement on a pre-cooled freezer block for 1 minute. In parallel, the master mix required for reverse transcription was prepared from 5× RT Buffer (Thermo Fisher Scientific), RNaseOUT, Nuclease-Free Water (NFW), and 10 µM Strand-Switching Primer (SSP). This master mix was added to the cooled samples, which were incubated in a thermal cycler at 42°C for 2 minutes, followed by a further incubation at 42°C for 90 minutes after the addition of 1 µl Maxima H Minus Reverse Transcriptase (Thermo Fisher Scientific). Enzyme inactivation was subsequently performed at 85°C for 5 minutes, followed by cooling to 4°C. Native RNA molecules were then digested by the addition of 1 µl RNase Cocktail Enzyme Mix (Thermo Fisher Scientific) and incubation at 37°C for 10 minutes. AMPure XP beads (AXP) were added for purification of the single-stranded cDNA, after which the samples were pelleted on a

magnetic rack, washed with 80% ethanol, and eluted in NFW. Following RNA degradation, 10 μM PR2 Primer, 2× LongAmp Taq Master Mix (NEB), and NFW were added to the samples for second-strand cDNA synthesis, followed by a single-cycle PCR reaction. The cycling conditions were as follows: 1 cycle at 94°C for 1 minute (denaturation), 1 cycle at 50°C for 1 minute (annealing), and 1 cycle at 65°C for 15 minutes (extension), followed by cooling to 4°C. After an additional ethanol wash, the NEBNext® Ultra™ II End Repair/dA-Tailing Module (NEB) was added to the double-stranded cDNA to generate blunt ends by filling in or trimming 5′ and 3′ overhangs, followed by the addition of deoxyadenosine (dA) to the 3′ blunt ends via terminal deoxynucleotidyl transferase (TdT) activity. NFW was subsequently added, and incubation was performed in a thermal cycler at 20°C for 5 minutes and then at 65°C for 5 minutes. After a final ethanol wash, concentration measurement was performed.

To enable sample separation during sequencing, barcode adapters containing unique dT tails were ligated to the previously generated 3′ dA tails using Blunt/TA Ligase Master Mix (NEB), followed by incubation at room temperature for 20 minutes, after which the reaction was terminated by the addition of EDTA. Following the addition of AXP beads and purification of the barcoded cDNA, samples originating from early time points (4 h, 8 h) and late time points (16 h, 24 h, 48 h, 72 h, 96 h, 120 h) were pooled separately. Native Adapter (NA) sequencing adapters were then ligated to the pooled samples using Quick T4 DNA Ligase (NEB) and NEBNext® Quick Ligation Reaction Buffer 5X. After incubation at room temperature for 20 minutes, AXP beads were added and the samples were washed with Short Fragment Buffer (SFB), followed by elution of the completed sequencing libraries in Elution Buffer (EB). After concentration measurement, libraries from early time points were sequenced on an ONT MinION Mk1B device using an R10.4.1 flow cell and a FLO-PRO114M flow cell due to potential barcode hopping, whereas libraries from late time points were sequenced on a PromethION P2 Solo device using a FLO-PRO114M flow cell.

3.6. Bioinformatic analysis

The raw signals generated by nanopore sequencing were translated into nucleotide base sequences (A, C, T, G) using the Dorado-0.8.2 basecaller developed by ONT. The resulting raw reads were mapped to the reference genome (GenBank: OQ679822.1) using the minimap2 software with the following parameters: -Y -C5 -ax splice -cs. SeqTools (https://github.com/moldovannorbert/seqtools) was used for statistical calculations and promoter element detection.

For the annotation of transcripts derived from direct cDNA sequencing data and for the identification of transcription start sites (TSS), transcription end sites (TES), and introns, the Long-read RNA-Seq Transcript Isoform Annotator (LoRTIA) v0.9.9 software package (https://github.com/zsolt-balazs/LoRTIA), which was developed by our research group, was applied. The first program used was Samprocessor.py, by which erroneous reads originating from false priming, template switching, and RNA degradation were removed. The script was executed with the following settings: -five_adapter GCTGATATTGCTGGG -five_score 14 -check_in_soft 15 -three_adapter AAAAAAAAAAAAAAA -three_score 14 input output. In the subsequent analysis, potential transcription start site (TSS) positions were identified using the command Stats.py -r genome -f r5 -b 10. For the determination of transcription end site (TES) positions, the command Stats.py -r genome -f r3 -b 10 was applied. Intron annotation was performed using the Stats.py -r genome -f in command, with introns identified from direct RNA sequencing data and confirmed using direct cDNA sequencing data. Random 5′ and 3′ positions originating from RNA degradation were filtered based on a Poisson distribution, for which the GFF_creator.py -s poisson -o script was used, and significance levels were corrected using the Bonferroni correction. Finally, transcript identification was carried out using the command Transcripts_annotator_two_wobbles.py -z 20 -a 10.

In addition, TES positions were identified from direct RNA sequencing data using the corresponding module of the ONT Dorado software package, applying default settings for the determination and identification of poly(A) tail lengths. Transcript annotation was performed using the NAGATA software[73]. The IntaRNA program was applied to assess the extent of replication origin–associated RNAs (raRNAs) and RNA-RNA interactions[74].

The Matplotlib library was used for the visualization of nucleotide distributions and line graphs. The Integrative Genomics Viewer (IGV) was applied for transcriptome visualization[75]. The Bedtools getfasta package (https://bedtools.readthedocs.io/en/latest/content/tools/getfasta.html) was used to extract nucleotide sequences based on the specified genome coordinates obtained from the reference FASTA file.

4. Results

4.1. Basic analysis of the CaGHV-1 transcriptome from sequencing data

Of the 71,651,155 reads generated by direct cDNA sequencing, 5,260,508 were mapped to the viral genome. The average length of these sequences was 575.33 nt. In the case of direct RNA sequencing, 1,208,372 viral reads were obtained from a total of 9,833,760 reads, with an average length of 1,013.77 nucleotides (Figure 4.). Erroneous Transcription Start Site (TSS) and Transcription End Site (TES) ends originating from false priming, reverse transcription artifacts caused by RNA degradation, as well as erroneous introns resulting from sequencing were filtered out and excluded by the LoRTIA program. For the 5′ end, following adapter alignment evaluated using the Smith-Waterman algorithm, the first nucleotide differing from the adapter sequence was marked as a potential TSS. During TES identification, the last nucleotide not matching the homopolymer A was designated as a potential TES. False ends originating from template switching or false priming, which were preceded by at least three adenines within the homopolymer A region, were removed, thereby reducing false positive results. As an additional filtering criterion, in the case of direct cDNA sequencing, a TSS was required to be present in at least three samples; otherwise, it was excluded. For TESs, confirmation of the 3′ end in the direct RNA sequencing dataset was required. Reads with fewer than two occurrences or with coverage below one percent were also filtered out. Introns were validated based on direct RNA sequencing data and were further verified using direct cDNA sequencing data. Among the most frequently occurring transcript isoforms or canonical transcripts of the virus, 278 were annotated, and the exact locations of the coding sequences were also determined.

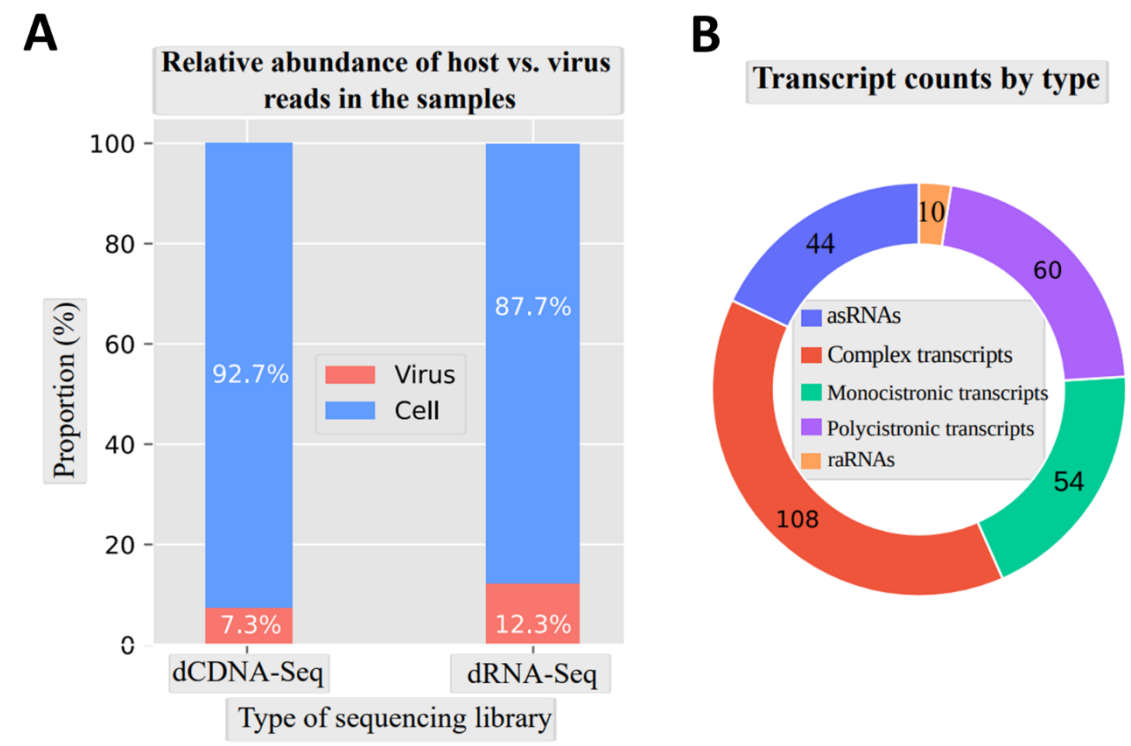


Figure 4. Statistical data. *These plots show the percentage of virus and host cell reads in the dcDNA and dRNA sequencing data compared to the total reads. The figure on the right shows the numerical composition of transcript types, omitting two categories which are a ncRNA and a putative canonical RNA.*

4.2. Promoters and TSSs identification

In comparison with the approximate eukaryotic consensus sequence[76], reduced conservation was observed for the initiator sequence of transcription start sites (TSSs) in the *Caviid gammaherpesvirus 1* genome (Py ANU/A). Predominance of G/A nucleotides was detected at the transcription start site, with guanine being extremely frequent at position +1, whereas T/C nucleotides were most frequently observed at position -1 upstream of the TSS (Figure 5.). The dominant occurrence of guanine nucleotides has previously been identified in other herpesvirus species, including *Bovine herpesvirus 1* (BoHV-1)[77] and *Epstein-Barr virus* (EBV)[78], as well as in the VP5 promoter initiator (Inr) element of *Herpes simplex virus 1* (HSV-1)[79,80].

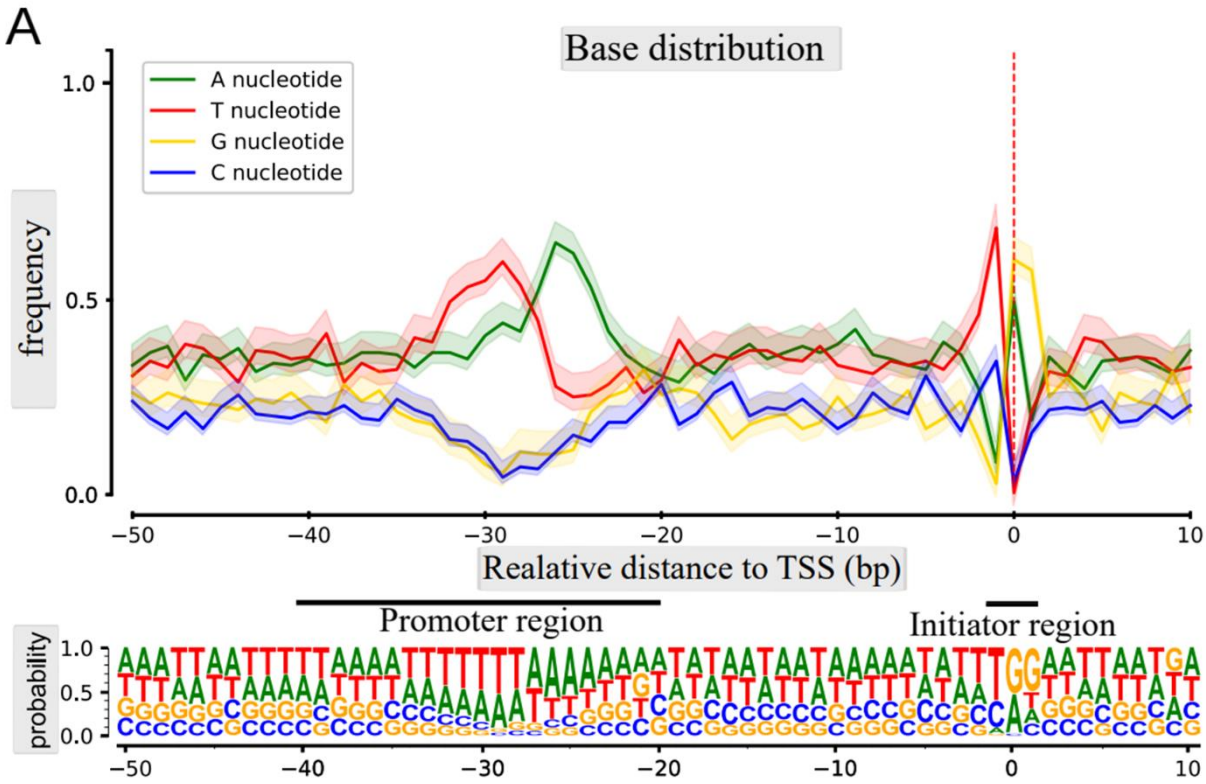


Figure 5. Analysis of Caviid gammaherpesvirus 1 TSSs. *It shows the distribution of nucleotide frequency in a section ranging from -50 to +10 base pairs, where G/A nucleotides are dominant at positions +1 and +2, and C/T nucleotides appear in greater numbers at position -1.*

A total of 162 canonical transcripts and 92 potential TATA boxes were identified in the *Caviid gammaherpesvirus 1* genome, with an average distance of 30.80 nucleotides from the TSS. In addition, five putative GC boxes were annotated at an average distance of 42 nucleotides, and 18 potential CAAT boxes were identified at an average distance of 112.33 nucleotides upstream of the TSS. T/A nucleotide-rich promoter elements were detected at positions -20 and -40 (Suppl. Table S03, Torma et al. 2025). Most of these promoter regions were found to contain a TATA box sequence with a TATTWAA motif, which has previously been demonstrated in KSHV to play an important role in the initiation of late gene transcription[71,81]. In this analysis, a large number of transcript ends were annotated, including the mRNAs of ORF75, ORF25, ORF26, and ORF59; however, the transcription start site of a non-coding RNA exhibited the highest read count, confirmed by 1,661,506 reads, indicating high transcriptional activity (Figure 6.).

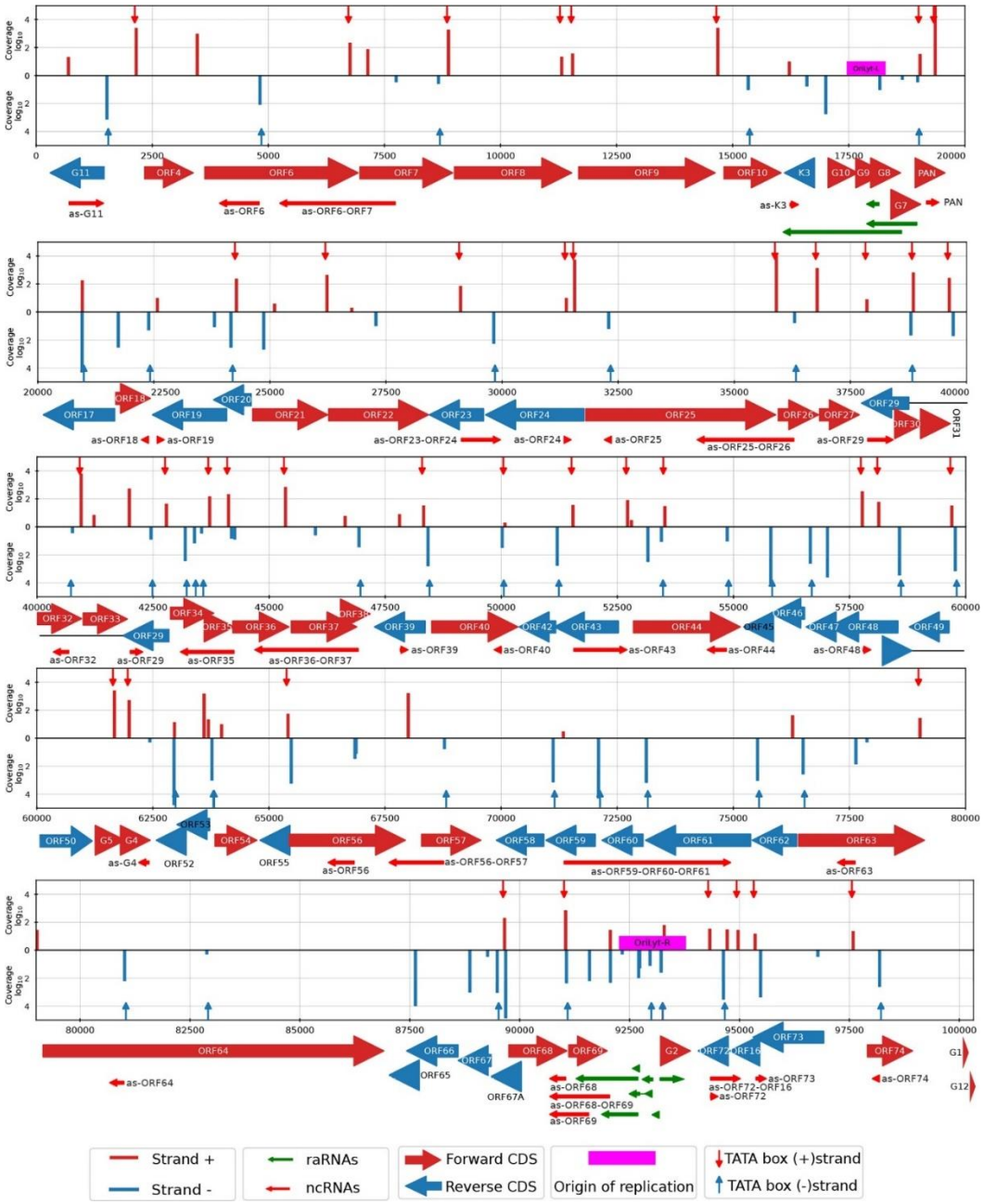


Figure 6. Distribution of TSSs on the CaGHV-1 genome. *The distribution of 5' transcript ends determined by the LoRTIA program is indicated by red color for positive strand (+) and blue color for strand negative (-) in the coverage boxes, and the height of the lines shows the amount of Transcription Start Site on a log10 scale. The locations of TATA boxes are also indicated by downward and upward arrows in the coverage diagram. Coding sequences (CDS) are indicated by large red 5'-3' and blue 3'-5' arrows below the boxes. Replication-associated RNAs (raRNAs) are shown in green, while non-coding RNAs are shown in red.*

4.3. Polyadenylation signals and TESs identification

During the analysis, data derived from direct cDNA sequencing were used for the determination of the 3′ ends of the viral genome, which were confirmed using direct RNA sequencing data. A total of 140 canonical transcription end sites were identified, of which 131 were associated with polyadenylation signals, with an average distance of 25.94 nucleotides between TESs and PASs (Suppl. Table S04, Torma et al. 2025) (Figure 7.). The PASs were predominantly located within a 50-nucleotide upstream region.

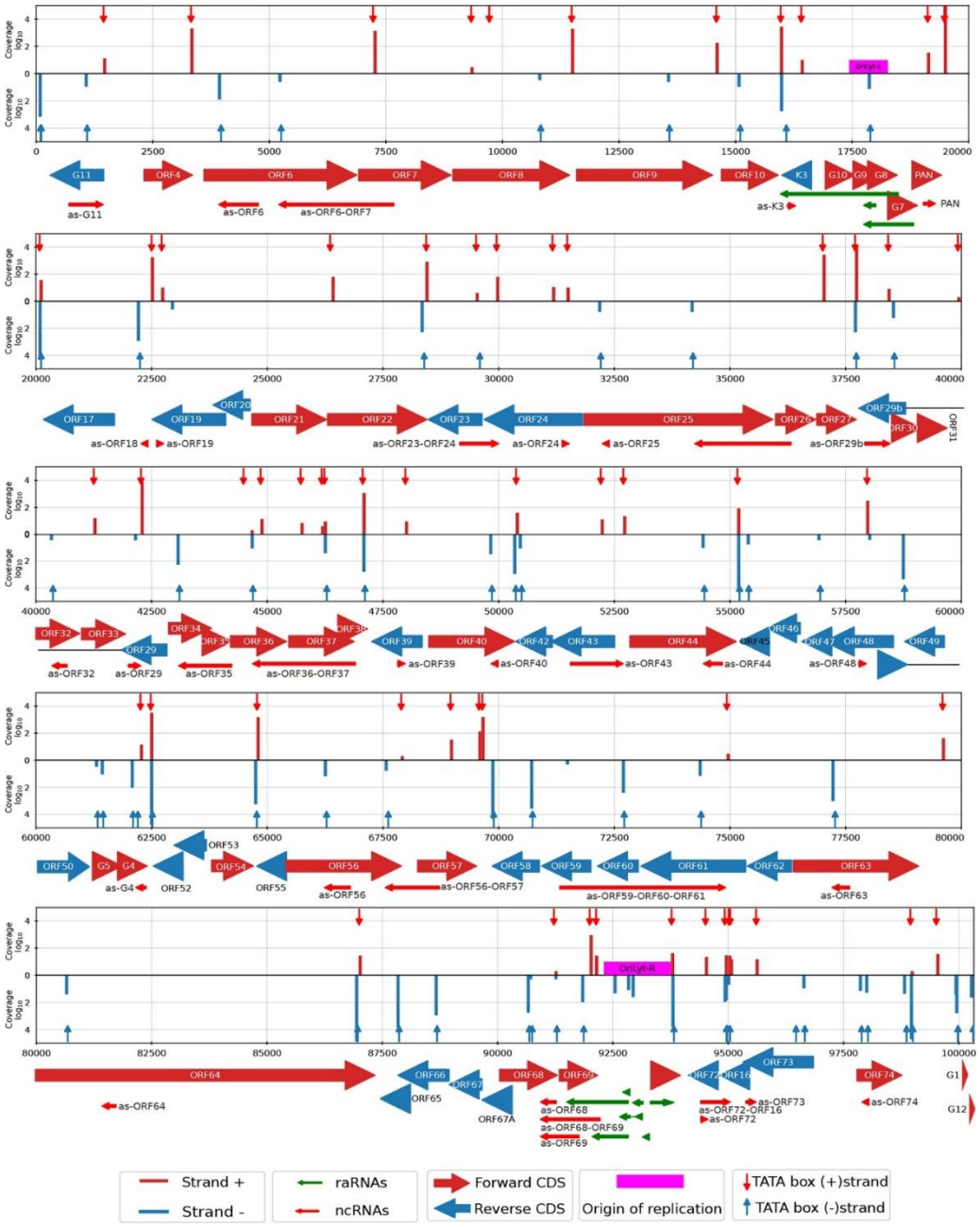


Figure 7. Distribution of TESs on the CaGHV-1 genome. *The distribution of 3′ transcript endpoints identified by LoRTIA software is indicated by red for positive (+) and blue for negative (-) strands in the coverage fields; the height of the lines represents the number of transcription endpoints on a log10 scale. The locations of PAS are also indicated by downward and upward arrows in the coverage diagram.*

4.4. Identification of introns and splice junctions occurring in viral transcripts

Direct cDNA sequencing data were used to confirm the splicing patterns of the CaGHV-1 transcriptome obtained from direct RNA sequencing data. Elevated levels of spliced transcripts were attributed to alternative splicing events, which were distributed across multiple genomic regions, including GPHLV-11, ORF20–21, ORF29, ORF31–ORF29b, ORF54–ORF57, and ORF63–ORF67. Furthermore, 56 introns were confirmed by both sequencing approaches; however, 280 introns were detected exclusively from direct cDNA sequencing data and were therefore considered likely artifacts generated during sequencing library preparation. To ensure the accuracy of the direct RNA sequencing data, they were compared with direct RNA sequencing data obtained from monkeypox virus, which does not undergo splicing. As no false splice sites were identified in the transcriptome, the data generated by native RNA sequencing and subsequent bioinformatic analyses were considered to be accurate. Additional validation was provided by introns identified using the NAGATA program, which confirmed all introns detected by the LoRTIA program and additionally identified six low-abundance splice sites that lacked the consensus sequence characteristic of canonical splicing. These results demonstrated that the detected spliced transcripts were of biological origin and that the majority of them likely represent transcriptional noise.

In mRNAs, most introns were found to be confined to UTRs. In contrast, introns detected within the coding regions of the ORF50 and ORF57 genes resulted in altered amino acid compositions at the N-terminal ends of the encoded proteins. Further analysis revealed that ORF29 was composed of two exons, which spanned four genes in the antisense direction, namely ORF30, ORF31, ORF32, and ORF33. This intron arrangement was shown to display conserved homology with KSHV; however, in comparison with HVS-1 and PRV, which belong to the alphaherpesvirus family and possess a homologous UL15 gene, only two genes were transcribed[82–84]. Several intron-containing non-coding RNAs were identified in the G4 and G5 regions of the genome, in which small ORFs were present, although their functional roles have not yet been determined. In addition, a non-coding replication origin–associated RNA located

in the vicinity of the OriLyt-R region was annotated, within which introns were detected. Furthermore, ncRNAs were annotated in the ORF63 and ORF64 gene regions, where splicing events occurred in the antisense portions of complex transcripts.

4.5. Monocistronic and Polycistronic transcripts

In viral transcriptomics, transcripts are commonly classified based on the number and relative orientation of the open reading frames (ORFs) they contain. Monocistronic transcripts encode a single ORF, whereas polycistronic transcripts encompass two or more ORFs arranged in tandem in the same orientation. Complex transcripts are defined as transcription units that contain ORFs or transcript segments with different orientations, resulting in bidirectional, overlapping, or antisense configurations.

During the analyses, 54 monocistronic, 60 multigene, and 108 complex mRNA transcripts were annotated as the most frequently occurring transcript variants associated with individual viral genes (Supp. Table S06, Torma et al.). One of the most abundant canonical transcripts was identified as ORF17.5, which is located within ORF17, contains a 5′-truncated open reading frame, and was detected more frequently than the full-length transcript. This phenomenon was not unique, as this embedded gene was also shown to be expressed by orthologous genes of viruses belonging to the alpha-, beta-, and gammaherpesvirus families[85]. In addition, an overlapping transcript was identified that initiated at the end of the G12 gene, spanned the genome junction, and terminated within the GPHLV-11 gene. Canonical polycistronic transcripts were annotated that contained at least two or more ORFs arranged in tandem in the same orientation. In contrast, complex transcripts were characterized as bidirectional, with at least one transcript oriented in the opposite direction. The longest multigene transcript was found to span a total of seven genes in the ORF4–ORF10 region, whereas the longest complex transcript encompassed 16 genes and shared a promoter with the PAN non-coding RNA, which is located between ORF17 and ORF33 in the viral genome.

Three short open reading frames located in the 5′ untranslated region, referred to as upstream ORFs, were identified upstream of the CaGHV-1 ORF35 gene, which has already been annotated in KSHV as ORF35–ORF37. In addition, the downstream gene of the ORF72–ORF71 polycistronic transcript was translated through a specific termination-reinitiation mechanism (Figure 8.)[86,87]. However, in comparison with KSHV, the uORFs identified in *Caviid gammaherpesvirus 1* were located at a greater distance from the ATG start codon.

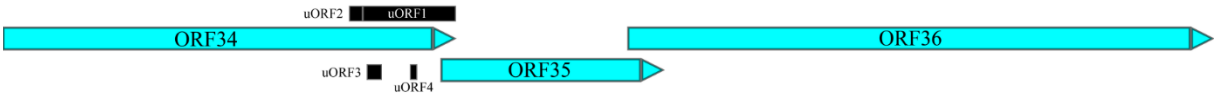
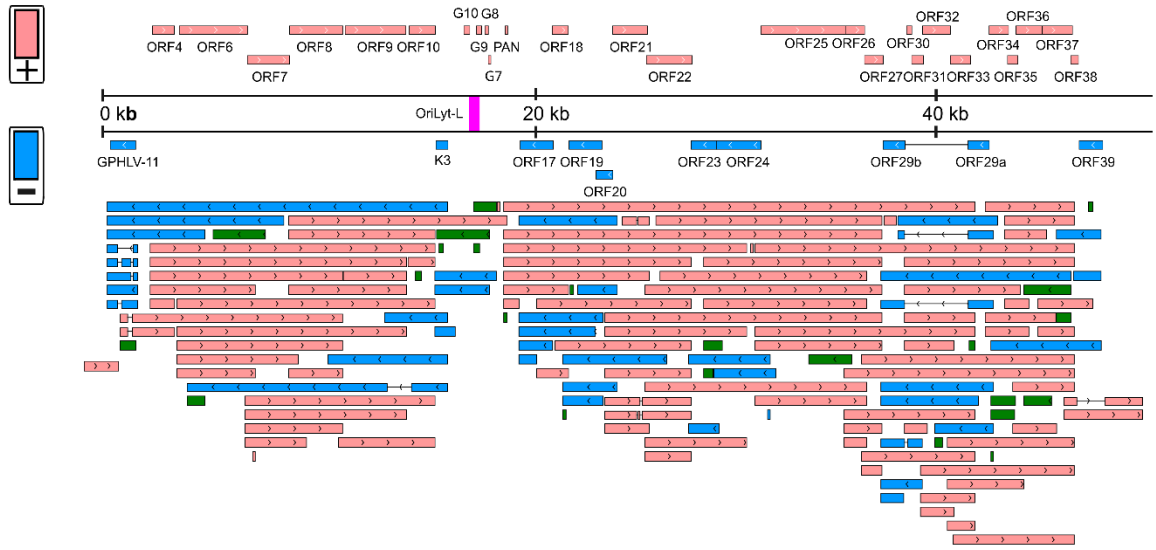


Figure 8. uORFs in the CaGHV-1 genome. *Three uORFs identified upstream of the ORF35 gene.*

4.6. Non-Coding transcripts

Large numbers of long 5′-UTR variants, antisense transcripts (asRNAs), and transcripts located in intergenic regions were contained within the ncRNA population. One of the most abundant non-coding RNAs was identified as polyadenylated nuclear RNA (PAN), similarly to KSHV[88,89]. Presumably, non-coding RNAs were represented by polygenic transcripts with large distances between the ATG start codon and transcription end sites, or by transcripts in which the most upstream genes of complex transcripts were positioned in the opposite orientation. A total of 44 asRNAs were identified, of which 33 antisense RNAs were associated with a single gene, 10 with two genes, and one asRNA with three genes (Figure 9.). The ratio of overlapping antisense RNA/mRNA pairs was also calculated (Supp. Table S06B, Torma et al.). Of the 26 TATA boxes identified, 15 were found to contain the TATTWAA motif. Based on further statistical analyses, the shortest non-coding RNA was determined to be 104 nucleotides in length, whereas the longest ncRNA was 3,788 nt long and was identified in the ORF63–ORF64 gene region. The average length of ncRNAs was calculated to be 758.78 nucleotides. Several antisense RNAs were annotated in the ORF63–ORF64 and ORF75 genomic regions, where they were detected in large numbers in both spliced and unspliced forms, similarly to observations made in the closely related *Murine gammaherpesvirus 68*[90].



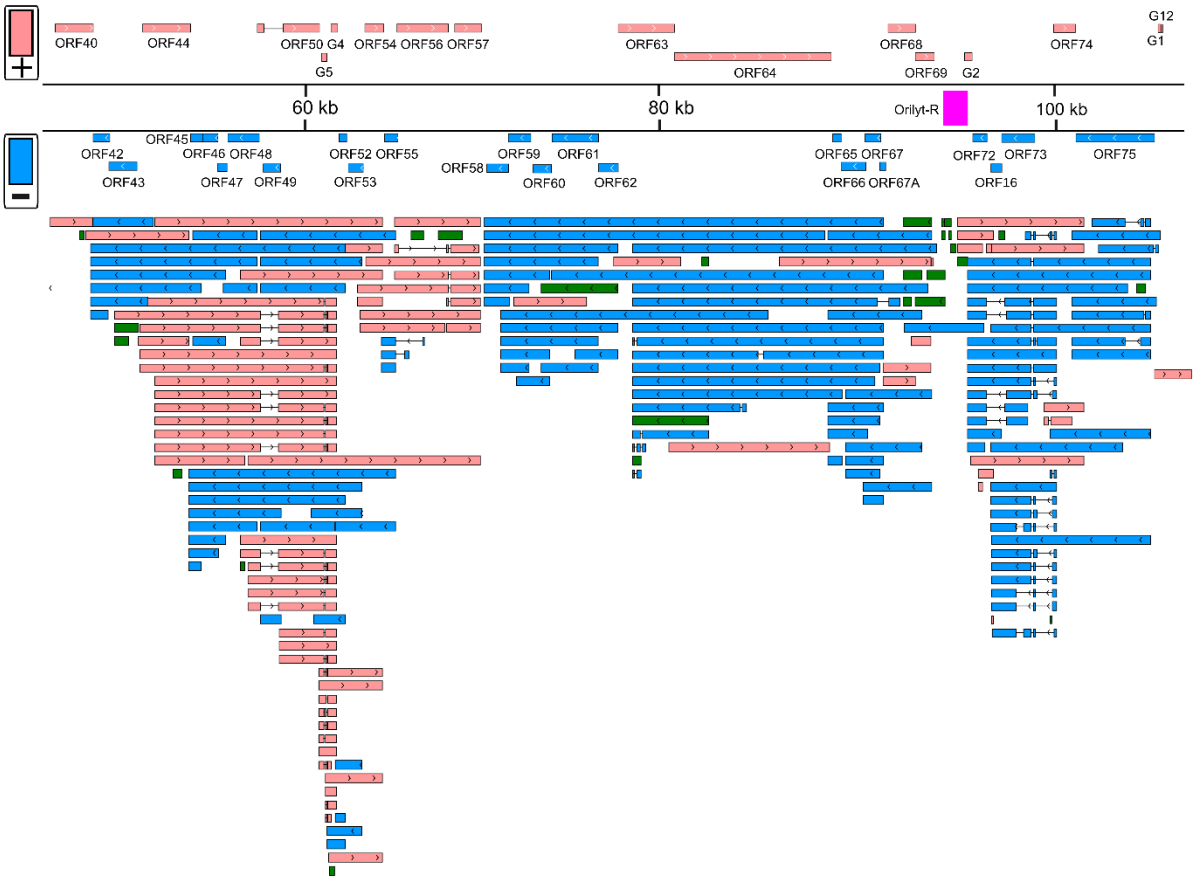


Figure 9. Canonical transcripts of the Caviid gammaherpesvirus 1 genome. *The 5'-3' orientation of mRNAs and non-coding RNAs is shown in pink, while the 3'-5' orientation is shown in blue. Antisense RNAs and raRNAs are shown in green. Transcripts that showed different splicing patterns were represented as independent canonical transcripts.*

4.7. Replication origin-associated RNAs

Several transcripts were identified in the vicinity of the replication origin, the majority of which were designated as replication origin–associated RNAs (raRNAs), representing a distinct class of non-coding RNAs. raRNAs have previously been identified in KSHV and EBV[91]. The replication origins of *Caviid gammaherpesvirus 1* were annotated based on sequence alignment using the KSHV reference genome and by mapping previously described KSHV replication origins onto the CaGHV-1 genome. Multiple non-coding RNAs were identified that overlapped the OriLyt-L region with their 5′ untranslated regions and terminated coterminally with K3 transcripts. In addition, a long transcript was identified that terminated coterminally with PAN and fully transcribed the replication origin. In the OriLyt-R genomic region, a transcription end site isoform expressed by ORF72 was identified, the 3′ untranslated region of which completely

overlapped the replication origin. At this stage, it remained unclear whether this transcript plays a role in translation or solely influences the process of DNA replication. The promoter regions of these RNAs were directly associated with OriLyt-R, and transcripts derived from the OriLyt-R region predominantly utilized the TATTWAA promoter motif. This consensus sequence was recognized by the LTF1 transcription factor encoded by the ORF24 gene, by which RNA polymerase II binding was rendered more efficient, thereby promoting global transcriptional regulation. Consequently, it was assumed that, during the late phase of infection, binding of LTF1 to the replication origin could influence DNA replication. The presence of long, complex transcripts overlapping the entire lytic origin within the genomic region was also demonstrated, as well as non-overlapping antisense RNAs located in the vicinity of OriLyt-R but associated with ORF69. A latent replication origin could not be detected in *Caviid gammaherpesvirus 1*; however, its existence cannot be excluded. For the prediction of in silico RNA/RNA interactions involving replication origin–associated RNAs (raRNAs), analysis performed using the IntaRNA program indicated that three raRNAs located in the OriLyt-L region exceeded the literature-defined threshold (−30 kcal/mol). Accordingly, potential interaction partners were identified as ORF9, ORF50, ORF64, and ORF73. To determine whether these high interaction scores possess true functional relevance, further investigations are required (Supp. Table S06, Torma et al.) (Figure 10.).

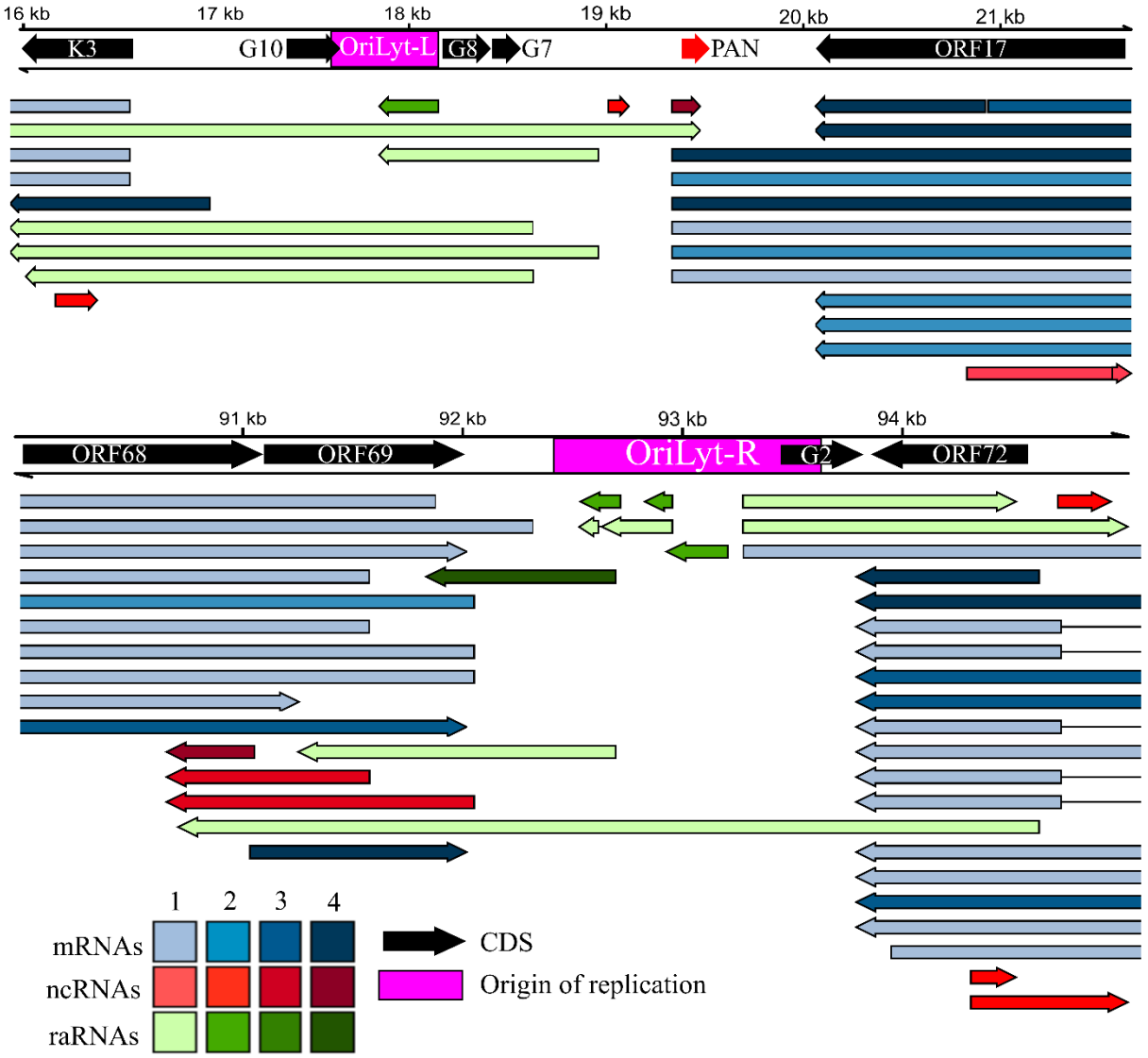


Figure 10. raRNAs. *In the regions around the two replication origins OriLyt-L and OriLyt-R, the red arrows indicate ncRNAs, the green arrows indicate rRNAs, and the blue arrows indicate monocistronic and multigene RNAs. The intensity of the colors indicates the frequency of the transcripts, which have been classified into four categories: 1(1-9), 2(10-49), 3(50-199), 4(>200).*

4.8. Transcriptional overlaps

Based on the overlap of raw reads obtained from direct RNA sequencing, transcripts of genes with different orientations, including convergent (→ ←), divergent (← →), and tandem (→ → or ← ←) arrangements, were found to overlap, indicating a high degree of complexity in the viral transcriptional pattern (Figure 11.). Similar transcript overlaps have previously been demonstrated in EBV and KSHV[34,92]. This observation indicated that the viral genome is

transcriptionally highly active. Convergent canonical transcript groups were identified in which the 3′ ends overlapped and were therefore classified as “hard” overlaps, where both transcripts were completely or largely overlapping. Examples included ORF18–19, ORF22–23, ORF27, ORF29b, ORF38–39, ORF40–42, ORF54–55, ORF64–65, G2–ORF72, and ORF74–75. In addition, convergent transcript clusters were annotated in which the 3′ ends did not terminate within each other; these were designated as “soft” overlaps, such as those observed between ORF10–K3, ORF44–45, G4–ORF52, and ORF57–58. In such genomic configurations, transcriptional readthrough was inferred to occur, resulting in occasional overlaps between tandem, convergent, or divergent gene pairs.

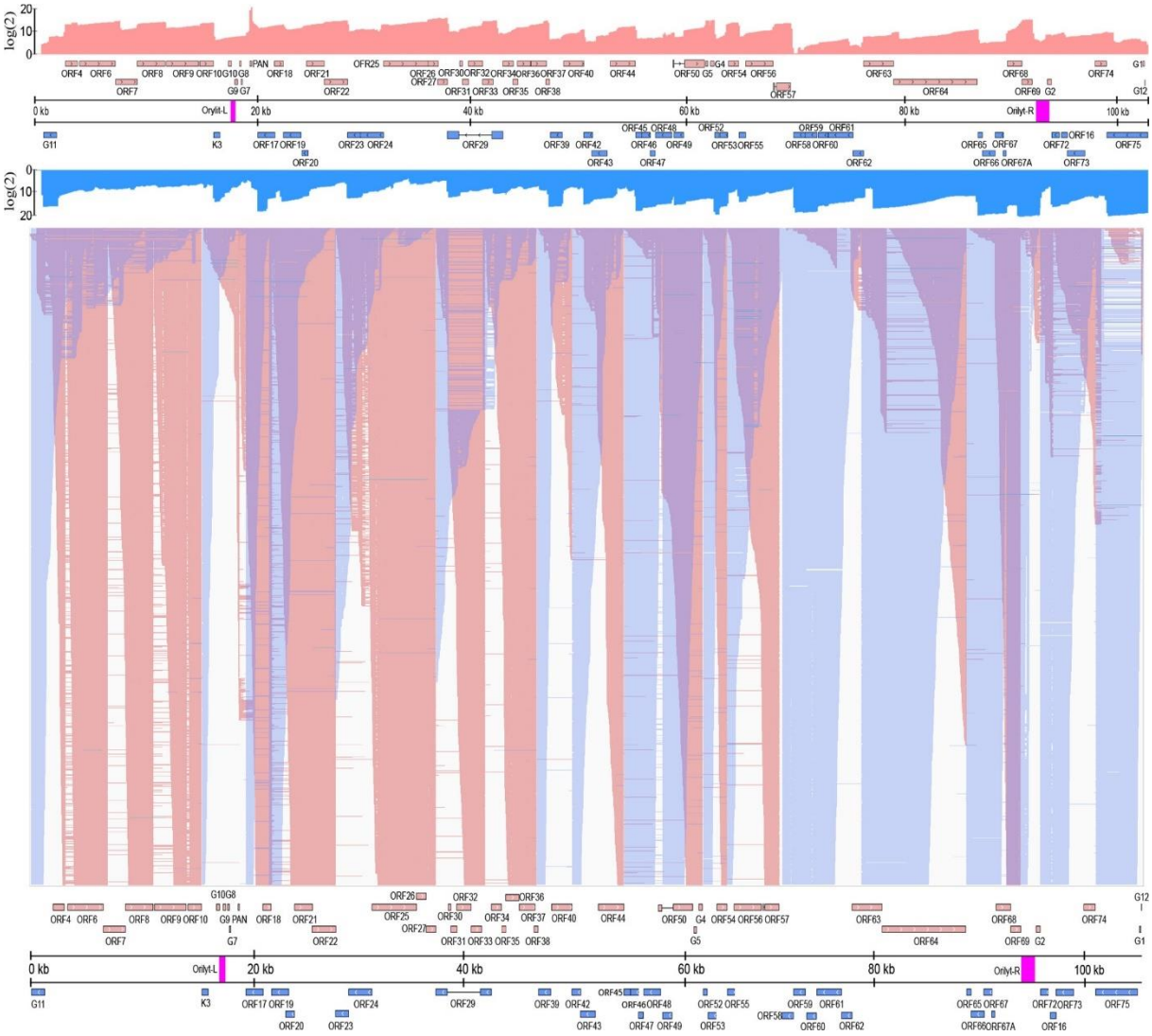


Figure 11. CaGHV-1 transcription overlaps. *CaGHV-1 transcriptional overlaps. The upper part of the figure shows the coverage scale on a log2 scale, which indicates global transcriptional activity. The 5'-3' orientation is marked in pink, and the 3'-5' orientation is*

marked in blue. The lower two-thirds of the figure show transcript overlaps, which are formed by divergent and convergent genes.

4.9. Replication and Transcription Activator (RTA) transcriptional activity and comparison

The RTA protein is a transcription factor conserved among gammaherpesviruses and is encoded by the ORF50 immediate-early (IE) gene, by which viral lytic reactivation from latency is regulated[93–97] and the ORF50 promoter, together with lytic genes, is activated. The results demonstrated that the most abundant mRNAs of the Caviid gammaherpesvirus 1 ORF50 gene contained four exons and originated from two adjacent transcription start sites. RTA, consisting of 643 amino acids encoded by the first two exons, was expressed as a 90 kDa protein. To investigate the transcriptional activity of Caviid gammaherpesvirus 1 RTA (gpRTA), a 3 kb DNA fragment of the viral promoter was cloned upstream of the ORF50 translation start site into a luciferase reporter vector. Transfection of HEK293T and 104C1 guinea pig fibroblast cell lines with increasing amounts of the 3×FLAG-gpRTA expression plasmid demonstrated that the ORF50 promoter was significantly activated by gpRTA in both cell lines in a concentration-dependent manner. To further examine the effect of gpRTA, the ORF50 promoter was truncated to 2 kb and 1 kb regions. Subsequent experiments showed that, in HEK293T cells, comparable activation was observed for the 3 kb, 2 kb, and 1 kb promoters, whereas in the fibroblast cell line, activation of the 2 kb promoter was reduced to approximately one quarter of that observed for the 3 kb promoter, indicating cell type– and/or species-specific differences. Additional luciferase assays were performed to assess gpRTA-mediated promoter induction in comparison with other gammaherpesvirus homologues. The 3 kb ORF50 promoter of *Caviid gammaherpesvirus 1* was compared with the corresponding replication and transcription activators of EBV, KSHV, HVS, MHV68, CaGHV-1, and RRV. While gpRTA efficiently induced the *Caviid gammaherpesvirus 1* ORF50 promoter in both HEK293T and fibroblast cell lines, the homologous proteins exhibited differential promoter activation patterns between the two cell types (Figure 12.). In conclusion, gpRTA was shown to be functionally similar to its homologues with respect to its ability to induce its own promoter.

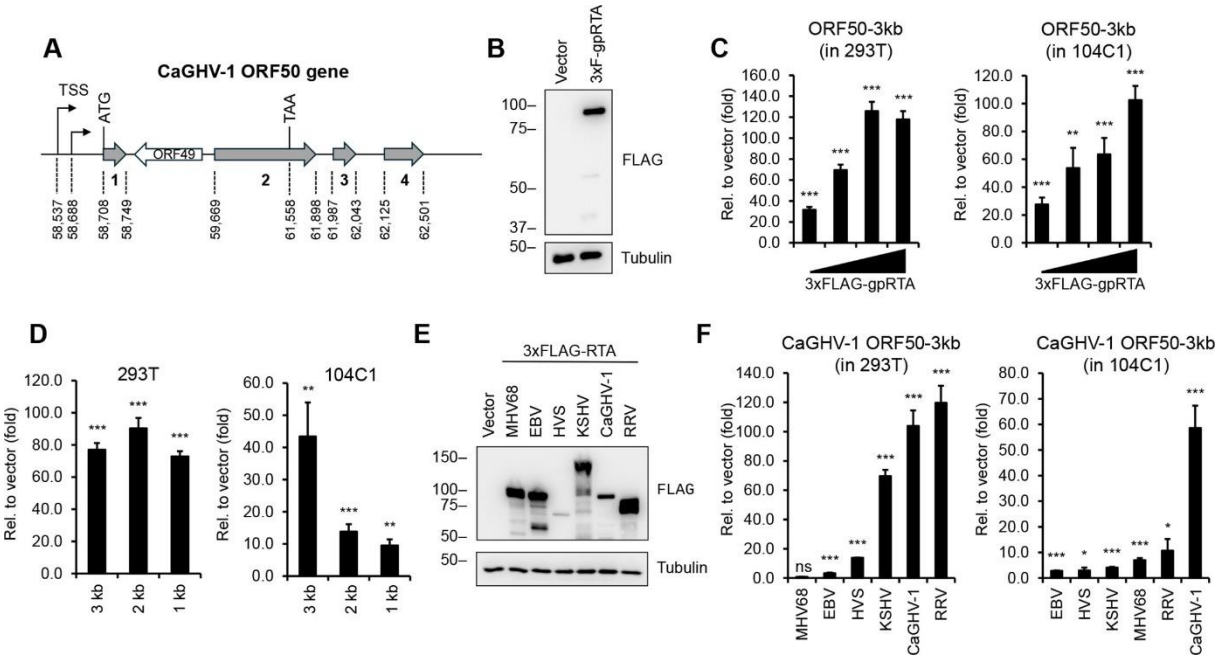


Figure 12. Transcriptional activity of CaGHV-1. *(A) ORF50 gene structure with four exons. (B) Expression of the 90 kDa 3×FLAG-tagged RTA protein in HEK293T cells. (C) Detection of the inducibility of the ORF50 promoter region in HEK293T and guinea pig fibroblast 104C1 cell lines. (D) Assessment of the effect of RTA on shorter regions of the ORF50 promoter. (E) Detection of RTA protein expression by Western blotting. (F) Analysis of the transcriptional activity of RTAs from different gammaherpesviruses on the CaGHV-1 ORF50 promoter region in the two different cell lines. Significance values determined by t-tests (n = 3) are interpreted as follows: ns = not significant, * = P < 0.05, ** = P < 0.01, *** = P < 0.001.*

5. Discussion

To generate a comprehensive transcriptome atlas of *Caviid gammaherpesvirus 1*, data obtained from Oxford Nanopore Technologies direct cDNA and direct RNA sequencing were analyzed using the NAGATA and LoRTIA bioinformatics pipelines. Numerous canonical monocistronic, multigene, fusion, and complex transcripts were identified, including intergenic transcripts such as PAN, non-coding RNAs, antisense RNAs, replication origin–associated RNAs, and transcripts containing truncated open reading frames. Cis-regulatory elements were annotated at single-nucleotide resolution, including transcription start sites and promoter elements, and polyadenylation signals, 3′ cleavage sites, and splice junctions were also identified. To eliminate reads containing false 5′ ends resulting from erroneous template switching and to accurately identify transcription start sites, the LoRTIA program was applied. Additional filtering criteria were introduced to further reduce false-positive ends, such that a transcription start site was accepted only if it was detected in all three biological replicates. As a result of this stringent filtering strategy, 99 TATA boxes and 165 transcription start sites were identified. The TATTWAA motif, which has previously been identified in other beta- and gammaherpesvirus species, including CMV, EBV, and KSHV, was detected upstream of several transcription start sites[98]. The TATTWAA consensus sequence functions as a binding site for LTF1 and the viral transcriptional activator (vTA), which are essential for the transcription of viral late genes[81]. Furthermore, the initiator (Inr) pattern was found to be consistent with results previously reported for other herpesviruses, indicating that the nucleotide composition of the TSS surrounding region is conserved[77,78,99].

The LoRTIA program was applied for the annotation of 3′ ends. To avoid false priming events and erroneous ends resulting from template switching, only those transcription end sites that were detected in all three samples and contained at least three adenines at the 3′ end were accepted. This filtering strategy was further reinforced by validation using native RNA molecules derived from direct RNA sequencing, thereby reducing false-positive results. Based on these criteria, a total of 143 transcription end sites were identified, a large proportion of which contained the conserved AAUAA sequence characteristic of eukaryotic mRNAs. This motif has previously been demonstrated in other herpesvirus subfamilies, suggesting that 3′ end cleavage is modulated by host cellular mechanisms[100,101].

RNA molecules overlapping the lytic replication origin have been observed in several DNA viruses. Examples include RNA 4.9, the most abundant viral RNA in human cytomegalovirus

(HCMV) of the betaherpesvirus subfamily, which has been shown to play a role in viral replication, as well as the T1.5 transcript expressed in KSHV, a gammaherpesvirus, which is involved in the maintenance of cell survival and the modulation of immune processes[102,103]. Replication origin–associated RNAs (raRNAs) were also annotated in the OriLyt-L and OriLyt-R regions of *Caviid gammaherpesvirus 1*. Some raRNAs were found to originate directly from the replication origin, whereas others overlapped the Ori region and were identified as short non-coding RNAs or long multigene transcripts that included regulatory elements of neighboring genes. One of these overlaps extended to PAN from ORF8, which is notable because PAN is a long transcript; therefore, the use of the PAN polyadenylation site instead of an earlier termination site raises questions. One possible explanation is that, during the transcription of these long RNAs, conflicts may arise between DNA and RNA polymerases, and such overlaps may inhibit gene transcription, thereby contributing to the temporal and spatial separation of replication and transcription processes[104]. Another notable finding was that the TATTWAA consensus sequence required for late gene transcription was present in the OriLyt-R region, which has been shown to be involved in both DNA replication and transcriptional regulation. A more detailed investigation of these mechanisms may open new avenues for fundamental research and support the development of novel antiviral intervention strategies.

Several introns were identified based on data obtained from direct cDNA and direct RNA sequencing of CaGHV-1, both within untranslated regions (UTRs) of mRNA-coding transcripts and within non-coding RNAs. A splicing pattern similar to that observed in KSHV and MHV68 was detected in spliced transcripts, including the ORF50 transcript, which initiates and regulates the transition from the latent to the lytic phase of the viral life cycle, as well as ORF64[90,105,106]. Furthermore, a high degree of transcript isoform diversity was identified for ORF73, which encodes the latency-associated nuclear antigen (LANA), the expression of which is essential for the establishment and maintenance of latency following infection[107].

Numerous polycistronic, complex, and antisense transcripts were also identified, the majority of which overlapped within the viral genome, thereby demonstrating its high transcriptional complexity. Depending on their orientation, overlapping genes were classified as convergent, divergent, or tandem. These transcriptional overlaps were not restricted to internal genomic regions but also occurred at the terminal regions of the genome and spanned the circular genome junction, indicating that overlapping RNA expression was not limited solely to protein-coding genes.

Based on the homology of gene organization between the human pathogen KSHV and *Caviid gammaherpesvirus 1*, the CaGHV-1 ORF50 gene, which is conserved across all gammaherpesviruses, was cloned. This gene encodes the replication and transcription activator (RTA) transcription factor, by which entry of the virus into the lytic cycle is triggered[108]. One of the most important functions of RTA is the binding to and activation of the ORF50 promoter, thereby establishing a positive feedback loop that enhances the expression of RTA itself as well as other lytic genes. The results demonstrated that the *Caviid gammaherpesvirus 1* ORF50 promoter was strongly activated by the CaGHV-1 replication and transcription activator (RTA). In addition, homologous RTA transcription factors derived from other gammaherpesviruses were also shown to induce the CaGHV-1 ORF50 promoter, suggesting functional similarities in transcriptional activation among these viruses. Further studies are required to determine the extent and nature of these functional similarities between CaGHV-1 RTA and RTAs of human gammaherpesviruses, particularly with respect to the modulation of host and viral gene expression and the influence on protein degradation pathways.

The transcriptional mechanisms characteristic of gammaherpesviruses, which are mediated by viral proteins and non-coding RNAs, have been evolutionarily conserved and are considered to play a key role in shaping the biological properties and processes of gammaherpesviruses. Through these conserved mechanisms, effective viral replication and long-term latent persistence within the host organism are ensured, thereby enabling viral adaptation. One such example is polyadenylated nuclear RNA (PAN), a non-coding RNA that is most abundantly expressed during the lytic cycle of KSHV and plays an essential role in the regulation of viral gene expression. PAN has previously been identified in other gammaherpesviruses, including RRV and EHV-2[37,109,110]. Analyses of PAN ncRNAs derived from KSHV and RRV have revealed low sequence homology; however, functional interchangeability has been demonstrated[89]. An important observation is that, although MHV68 lacks a PAN homologue, *Caviid gammaherpesvirus 1* encodes PAN. This feature provides a unique opportunity to investigate the biological significance of PAN directly, both during infection and in pathogenesis, using a small animal model.

Based on the results obtained and the currently available knowledge, the identification of canonical transcripts, genes, and gene regulatory regions of *Caviid gammaherpesvirus 1*, together with the assessment of its tumorigenic potential, is considered to provide an excellent foundation for the establishment of CaGHV-1 as a novel model organism for the investigation of human gammaherpesvirus pathogenicity. Furthermore, the development of chimeric

CaGHV-1 strains capable of expressing KSHV glycoproteins could potentially enable the development of a vaccine against KSHV. Such model systems are expected to offer valuable opportunities for therapeutic development, thereby increasing the relevance of CaGHV-1 in both basic research and translational medicine.

6. Summary

In this study, the transcriptome of Caviid gammaherpesvirus 1 (CaGHV-1) was mapped using long-read sequencing combined with various bioinformatic programs, and the transcriptional activity of the replication and transcription activator (RTA) gene, which shows homology with other gammaherpesviruses and is encoded by ORF50, was investigated.

Following transcriptome mapping, numerous coding monocistronic, polycistronic, and complex mRNAs, as well as non-coding RNAs, including antisense, 5′-truncated, and replication origin–associated transcripts, were identified. In addition, cis-regulatory elements regulating the expression of individual genes were detected. Overlapping transcripts with different orientations were also identified, which, together with previous results, suggest that the CaGHV-1 genome represents a large-scale, transcriptionally complex gammaherpesvirus.

One of the most abundant non-coding RNAs, PAN, which is important for KSHV lytic viral gene expression and replication, was identified in CaGHV-1. Importantly, PAN was found to be absent from MHV-68, which further reinforced the role of CaGHV-1 as a potential model organism.

Based on the results of functional studies of CaGHV-1 RTA encoded by the ORF50 gene, activation of its own promoter was demonstrated, and homologous RTAs from other gammaherpesviruses were also shown to activate the CaGHV-1 ORF50 promoter, indicating functional similarity. This was reflective of the evolutionary conservation of transcriptional mechanisms.

Taken together, these results confirm that CaGHV-1, which shows homology with the human pathogen KSHV, offers new opportunities for studying the pathogenesis of human gammaherpesviruses as a new model organism, due to its non-human pathogenicity, less narrow host specificity, and tumorigenic effect.

7. Acknowledgements

The study of the CaGHV-1 transcriptome presented in the thesis was funded by the National Research, Development and Innovation Office (NRDIO), through the Researcher-initiated research projects (Grant number: K 142674) awarded to ZB. ZT was supported by the NIH grant R01DE028331. DT was supported by NRDIO FK 142676. The publication fee was covered by the University of Szeged, Open Access Fund: 7358.

First of all, I would like to thank my supervisors, Dr. Habil. Dóra Tombácz and Prof. Dr. Zsolt Boldogkői, for their guidance and support during my doctoral studies and before, both in my studies and in my research. It is an honor to have been given the opportunity and trust to work at this institute during my PhD studies.

I am grateful to my colleagues Dr. Gábor Gulyás and Dr. Zsolt Csabai for their support, assistance, and the many useful collaborations we have had over the past few years. I would also like to thank the members of Department of Medical Biology for their support and assistance in every way.

Last but not least, I cannot thank my beautiful fiancée Zília and my parents enough for their tireless support and perseverance, as well as my friends for their support. Thank you!

8. References

1. Nayfach, S. *et al.* CheckV assesses the quality and completeness of metagenome-assembled viral genomes. *Nat Biotechnol* **39**, 578–585 (2021).
2. Takács, M. A vírusok kémiai összetétele és szerkezete. in *Orvosi Virológia* 3 (Medicina Könyvkiadó Zrt., 2022).
3. Davison, A. J. *et al.* The order Herpesvirales. *Arch Virol* **154**, 171–177 (2009).
4. Davison, A. J. Evolution of the herpesviruses. *Veterinary Microbiology* **86**, 69–88 (2002).
5. McGeoch, D. J., Rixon, F. J. & Davison, A. J. Topics in herpesvirus genomics and evolution. *Virus Research* **117**, 90–104 (2006).
6. Dotto-Maurel, A., Arzul, I., Morga, B. & Chevignon, G. Herpesviruses: overview of systematics, genomic complexity and life cycle. *Virol J* **22**, 155 (2025).
7. Fu, M., Deng, R., Wang, J. & Wang, X. Whole-genome phylogenetic analysis of herpesviruses. *Acta Virol* **52**, 31–40 (2008).
8. Gatherer, D. *et al.* ICTV Virus Taxonomy Profile: Herpesviridae 2021. *J Gen Virol* **102**, 001673 (2021).
9. Connolly, S. A., Jardetzky, T. S. & Longnecker, R. The structural basis of herpesvirus entry. *Nat Rev Microbiol* **19**, 110–121 (2021).
10. Gruffat, H., Marchione, R. & Manet, E. Herpesvirus Late Gene Expression: A Viral-Specific Pre-initiation Complex Is Key. *Front. Microbiol.* **7**, (2016).
11. Morissette, G. & Flamand, L. Herpesviruses and chromosomal integration. *J Virol* **84**, 12100–12109 (2010).
12. Carneiro, V. C. de S., Pereira, J. G. & de Paula, V. S. Family Herpesviridae and neuroinfections: current status and research in progress. *Mem Inst Oswaldo Cruz* **117**, e220200 (2022).

13. Houldcroft, C. J. Human Herpesvirus Sequencing in the Genomic Era: The Growing Ranks of the Herpetic Legion. *Pathogens* **8**, 186 (2019).
14. Lan, K. & Luo, M.-H. Herpesviruses: epidemiology, pathogenesis, and interventions. *Virol Sin* **32**, 347–348 (2017).
15. Paludan, S. R., Bowie, A. G., Horan, K. A. & Fitzgerald, K. A. Recognition of herpesviruses by the innate immune system. *Nat Rev Immunol* **11**, 143–154 (2011).
16. Sorel, O. & Dewals, B. G. The Critical Role of Genome Maintenance Proteins in Immune Evasion During Gammaherpesvirus Latency. *Front. Microbiol.* **9**, 3315 (2019).
17. Weed, D. J. & Damania, B. Pathogenesis of Human Gammaherpesviruses: Recent Advances. *Curr Clin Micro Rpt* **6**, 166–174 (2019).
18. Jha, H., Banerjee, S. & Robertson, E. The Role of Gammaherpesviruses in Cancer Pathogenesis. *Pathogens* **5**, 18 (2016).
19. Blake, N. Immune evasion by gammaherpesvirus genome maintenance proteins. *Journal of General Virology* **91**, 829–846 (2010).
20. Münz, C. Human γ-Herpesvirus Infection, Tumorigenesis, and Immune Control in Mice with Reconstituted Human Immune System Components. *Front. Immunol.* **9**, 238 (2018).
21. Böni, M., Rieble, L. & Münz, C. Co-Infection of the Epstein–Barr Virus and the Kaposi Sarcoma-Associated Herpesvirus. *Viruses* **14**, 2709 (2022).
22. Guzha, B. T. *et al.* The impact of DNA tumor viruses in low-to-middle income countries (LMICS): A literature review. *Tumour Virus Res* **18**, 200289 (2024).
23. Gupta, A., Li, R., Shair, K. & Gao, S.-J. Model Systems of Gammaherpesvirus Infection, Immunity, and Disease. *J Med Virol* **97**, e70581 (2025).
24. Dittmer, D. P., Damania, B. & Sin, S.-H. Animal models of tumorigenic herpesviruses--an update. *Curr Opin Virol* **14**, 145–150 (2015).

25. Sin, S.-H. *et al.* The complete Kaposi sarcoma-associated herpesvirus genome induces early-onset, metastatic angiosarcoma in transgenic mice. *Cell Host Microbe* **32**, 755-767.e4 (2024).
26. Moorad, R. *et al.* Genome evolution of Kaposi sarcoma-associated herpesvirus (KSHV). *J Virol* **99**, e0195024 (2025).
27. Bechtel, J. T., Liang, Y., Hvidding, J. & Ganem, D. Host range of Kaposi's sarcoma-associated herpesvirus in cultured cells. *J Virol* **77**, 6474–6481 (2003).
28. Wen, K. W. & Damania, B. Kaposi sarcoma-associated herpesvirus (KSHV): molecular biology and oncogenesis. *Cancer Lett* **289**, 140–150 (2010).
29. Yu, C. J. & Damania, B. Molecular Mechanisms of Kaposi Sarcoma-Associated Herpesvirus (HHV8)-Related Lymphomagenesis. *Cancers (Basel)* **16**, 3693 (2024).
30. Cesarman, E. *et al.* Kaposi sarcoma. *Nat Rev Dis Primers* **5**, 9 (2019).
31. Ganem, D. KSHV and the pathogenesis of Kaposi sarcoma: listening to human biology and medicine. *J Clin Invest* **120**, 939–949 (2010).
32. Yang, W.-S. *et al.* Development of KSHV vaccine platforms and chimeric MHV68-K-K8.1 glycoprotein for evaluating the *in vivo* immunogenicity and efficacy of KSHV vaccine candidates. *mBio* **15**, e02913-24 (2024).
33. Broussard, G. & Damania, B. Regulation of KSHV Latency and Lytic Reactivation. *Viruses* **12**, 1034 (2020).
34. 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. *mSystems* **9**, e0100723 (2024).
35. Kedes, D. H., Lagunoff, M., Renne, R. & Ganem, D. Identification of the gene encoding the major latency-associated nuclear antigen of the Kaposi's sarcoma-associated herpesvirus. *J Clin Invest* **100**, 2606–2610 (1997).

36. Jary, A. *et al.* Kaposi's Sarcoma-Associated Herpesvirus, the Etiological Agent of All Epidemiological Forms of Kaposi's Sarcoma. *Cancers (Basel)* **13**, 6208 (2021).
37. Rossetto, C. C., Tarrant-Elorza, M., Verma, S., Purushothaman, P. & Pari, G. S. Regulation of viral and cellular gene expression by Kaposi's sarcoma-associated herpesvirus polyadenylated nuclear RNA. *J Virol* **87**, 5540–5553 (2013).
38. Chang, H. *et al.* Non-human primate model of Kaposi's sarcoma-associated herpesvirus infection. *PLoS Pathog* **5**, e1000606 (2009).
39. Fujiwara, S. & Nakamura, H. Animal Models for Gammaherpesvirus Infections: Recent Development in the Analysis of Virus-Induced Pathogenesis. *Pathogens* **9**, 116 (2020).
40. Fujiwara, S. Animal Models of Human Gammaherpesvirus Infections. in *Human Herpesviruses* (eds Kawaguchi, Y., Mori, Y. & Kimura, H.) vol. 1045 413–436 (Springer Singapore, Singapore, 2018).
41. Estes, J. D., Wong, S. W. & Brenchley, J. M. Nonhuman primate models of human viral infections. *Nat Rev Immunol* **18**, 390–404 (2018).
42. Blossom, D. EBV and KSHV – related herpesviruses in non-human primates. in *Human Herpesviruses: Biology, Therapy, and Immunoprophylaxis* (eds Arvin, A. et al.) (Cambridge University Press, Cambridge, 2007).
43. Bruce, A. G. *et al.* Experimental co-transmission of Simian Immunodeficiency Virus (SIV) and the macaque homologs of the Kaposi Sarcoma-Associated Herpesvirus (KSHV) and Epstein-Barr Virus (EBV). *PLoS One* **13**, e0205632 (2018).
44. Dhingra, A. *et al.* Novel Virus Related to Kaposi's Sarcoma-Associated Herpesvirus from Colobus Monkey. *Emerg Infect Dis* **25**, 1548–1551 (2019).
45. Jia, Z., Zhang, D., Zhu, L. & Xue, J. Animal models of human herpesvirus infection. *Animal Model Exp Med* **8**, 615–628 (2025).

46. Orzechowska, B. U. *et al.* Rhesus macaque rhadinovirus-associated non-Hodgkin lymphoma: animal model for KSHV-associated malignancies. *Blood* **112**, 4227–4234 (2008).

47. Wong, S. W. *et al.* Induction of B cell hyperplasia in simian immunodeficiency virus-infected rhesus macaques with the simian homologue of Kaposi's sarcoma-associated herpesvirus. *J Exp Med* **190**, 827–840 (1999).

48. Dong, S., Forrest, J. C. & Liang, X. Murine Gammaherpesvirus 68: A Small Animal Model for Gammaherpesvirus-Associated Diseases. in *Infectious Agents Associated Cancers: Epidemiology and Molecular Biology* (eds Cai, Q., Yuan, Z. & Lan, K.) vol. 1018 225–236 (Springer Singapore, Singapore, 2017).

49. Cieniewicz, B., Santana, A. L., Minkah, N. & Krug, L. T. Interplay of Murine Gammaherpesvirus 68 with NF-kappaB Signaling of the Host. *Front. Microbiol.* **7**, (2016).

50. A. Blackman, M., Flaño, E., Usherwood, E. & L. Woodland, D. Murine γ-herpesvirus-68: a mouse model for infectious mononucleosis? *Molecular Medicine Today* **6**, 488–490 (2000).

51. Shifflett, K. W. & Dittmer, D. P. Mouse models of Kaposi sarcoma-associated herpesvirus (KSHV). *Virology* **603**, 110384 (2025).

52. Wu, W. *et al.* KSHV/HHV-8 infection of human hematopoietic progenitor (CD34+) cells: persistence of infection during hematopoiesis in vitro and in vivo. *Blood* **108**, 141–151 (2006).

53. Wang, L.-X. *et al.* Humanized-BLT mouse model of Kaposi's sarcoma-associated herpesvirus infection. *Proc Natl Acad Sci U S A* **111**, 3146–3151 (2014).

54. Stanfield, B. A., Ruiz, E., Chouljenko, V. N. & Kousoulas, K. G. Guinea pig herpes like virus is a gamma herpesvirus. *Virus Genes* **60**, 148–158 (2024).

55. Hsiung, G. D. & Kaplow, L. S. Herpeslike virus isolated from spontaneously degenerated tissue culture derived from leukemia-susceptible guinea pigs. *J Virol* **3**, 355–357 (1969).

56. Hsiung, G. D., Kaplow, L. S. & Booss, J. Herpesvirus infection of guinea pigs. I. Isolation, characterization and pathogenicity. *Am J Epidemiol* **93**, 298–307 (1971).

57. Rhim, J. S. Malignant transformation of rat embryo cells by a herpesvirus isolated from L2C guinea pig leukemia. *Virology* **82**, 100–110 (1977).

58. Bia, F. J., Summers, W. C., Fong, C. K. & Hsiung, G. D. New endogenous herpesvirus of guinea pigs: biological and molecular characterization. *J Virol* **36**, 245–253 (1980).

59. Truong, V. N. Y., Ellis, R. & Stanfield, B. A. Guinea Pig X Virus Is a Gammaherpesvirus. *Viruses* **17**, 1084 (2025).

60. Wang, Z., Gerstein, M. & Snyder, M. RNA-Seq: a revolutionary tool for transcriptomics. *Nat Rev Genet* **10**, 57–63 (2009).

61. Oikonomopoulos, S. *et al.* Methodologies for Transcript Profiling Using Long-Read Technologies. *Front Genet* **11**, 606 (2020).

62. Benjamini, Y. & Speed, T. P. Summarizing and correcting the GC content bias in high-throughput sequencing. *Nucleic Acids Res* **40**, e72 (2012).

63. Wongsurawat, T., Jenjaroenpun, P., Wanchai, V. & Nookaew, I. Native RNA or cDNA Sequencing for Transcriptomic Analysis: A Case Study on Saccharomyces cerevisiae. *Front Bioeng Biotechnol* **10**, 842299 (2022).

64. Udaondo, Z. *et al.* Comparative Analysis of PacBio and Oxford Nanopore Sequencing Technologies for Transcriptomic Landscape Identification of Penaeus monodon. *Life (Basel)* **11**, 862 (2021).

65. Monzó, C., Liu, T. & Conesa, A. Transcriptomics in the era of long-read sequencing. *Nat Rev Genet* **26**, 681–701 (2025).

66. Deamer, D., Akeson, M. & Branton, D. Three decades of nanopore sequencing. *Nat Biotechnol* **34**, 518–524 (2016).

67. Zheng, P. *et al.* Nanopore sequencing technology and its applications. *MedComm (2020)* **4**, e316 (2023).

68. MacKenzie, M. & Argyropoulos, C. An Introduction to Nanopore Sequencing: Past, Present, and Future Considerations. *Micromachines (Basel)* **14**, 459 (2023).

69. Boldogkői, Z., Moldován, N., Balázs, Z., Snyder, M. & Tombácz, D. Long-Read Sequencing – A Powerful Tool in Viral Transcriptome Research. *Trends in Microbiology* **27**, 578–592 (2019).

70. Moldován, N. *et al.* Third-generation Sequencing Reveals Extensive Polycistronism and Transcriptional Overlapping in a Baculovirus. *Sci Rep* **8**, 8604 (2018).

71. Shekhar, R., McMahon, S., Tibbetts, S. A., Flemington, E. K. & Renne, R. Cross-Species Insights Into Gamma Herpesvirus Transcriptomes: Long-Read and Multi-Omics Perspectives. *J Med Virol* **98**, e70802 (2026).

72. Spires, L. M., Wind, E., Papp, B. & Toth, Z. KSHV RTA utilizes the host E3 ubiquitin ligase complex RNF20/40 to drive lytic reactivation. *J Virol* **97**, e0138923 (2023).

73. Abebe, J. S. *et al.* Nanopore Guided Annotation of Transcriptome Architectures. *bioRxiv* 2024.04.02.587744 (2024) doi:10.1101/2024.04.02.587744.

74. Mann, M., Wright, P. R. & Backofen, R. IntaRNA 2.0: enhanced and customizable prediction of RNA-RNA interactions. *Nucleic Acids Res* **45**, W435–W439 (2017).

75. Thorvaldsdóttir, H., Robinson, J. T. & Mesirov, J. P. Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. *Brief Bioinform* **14**, 178–192 (2013).

76. Javahery, R., Khachi, A., Lo, K., Zenzie-Gregory, B. & Smale, S. T. DNA sequence requirements for transcriptional initiator activity in mammalian cells. *Mol Cell Biol* **14**, 116–127 (1994).

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

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

79. Huang, C. J., Petroski, M. D., Pande, N. T., Rice, M. K. & Wagner, E. K. The herpes simplex virus type 1 VP5 promoter contains a cis-acting element near the cap site which interacts with a cellular protein. *J Virol* **70**, 1898–1904 (1996).

80. Tombácz, D. *et al.* Multiple Long-Read Sequencing Survey of Herpes Simplex Virus Dynamic Transcriptome. *Front Genet* **10**, 834 (2019).

81. Nandakumar, D. & Glaunsinger, B. An integrative approach identifies direct targets of the late viral transcription complex and an expanded promoter recognition motif in Kaposi's sarcoma-associated herpesvirus. *PLoS Pathog* **15**, e1007774 (2019).

82. Renne, R., Blackbourn, D., Whitby, D., Levy, J. & Ganem, D. Limited transmission of Kaposi's sarcoma-associated herpesvirus in cultured cells. *J Virol* **72**, 5182–5188 (1998).

83. Iwaisako, Y., Shimizu, K., Suzuki, Y., Nakano, T. & Fujimuro, M. Characterization of the Kaposi's sarcoma-associated herpesvirus terminase complex component ORF29. *Microbiol Spectr* **13**, e0330824 (2025).

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

85. Tombácz, D., Balázs, Z., Csabai, Z., Snyder, M. & Boldogkői, Z. Long-Read Sequencing Revealed an Extensive Transcript Complexity in Herpesviruses. *Front. Genet.* **9**, 259 (2018).

86. 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).

87. 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).

88. Rossetto, C. & Pari, G. PAN's Labyrinth: Molecular Biology of Kaposi's Sarcoma-Associated Herpesvirus (KSHV) PAN RNA, a Multifunctional Long Noncoding RNA. *Viruses* **6**, 4212–4226 (2014).

89. Withers, J. B., Li, E. S., Vallery, T. K., Yario, T. A. & Steitz, J. A. Two herpesviral noncoding PAN RNAs are functionally homologous but do not associate with common chromatin loci. *PLoS Pathog* **14**, e1007389 (2018).

90. O'Grady, T. *et al.* Genome-wide Transcript Structure Resolution Reveals Abundant Alternate Isoform Usage from Murine Gammaherpesvirus 68. *Cell Rep* **27**, 3988-4002.e5 (2019).

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

92. 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).

93. Santana, A. *et al.* RTA Occupancy of the Origin of Lytic Replication during Murine Gammaherpesvirus 68 Reactivation from B Cell Latency. *Pathogens* **6**, 9 (2017).

94. Hong, Y., Qi, J., Gong, D., Han, C. & Deng, H. Replication and transcription activator (RTA) of murine gammaherpesvirus 68 binds to an RTA-responsive element and activates the expression of ORF18. *J Virol* **85**, 11338–11350 (2011).

95. Luan, Y. *et al.* Linear ubiquitination regulates the KSHV replication and transcription activator protein to control infection. *Nat Commun* **15**, 5515 (2024).

96. Combs, L. R., Spires, L. M., Alonso, J. D., Papp, B. & Toth, Z. KSHV RTA Induces Degradation of the Host Transcription Repressor ID2 To Promote the Viral Lytic Cycle. *J Virol* **96**, e0010122 (2022).

97. Guito, J. & Lukac, D. M. KSHV Rta Promoter Specification and Viral Reactivation. *Front Microbiol* **3**, 30 (2012).

98. Dremel, S. E. & Didychuk, A. L. Better late than never: A unique strategy for late gene transcription in the beta- and gammaherpesviruses. *Semin Cell Dev Biol* **146**, 57–69 (2023).

99. Tombácz, D. *et al.* Dynamic transcriptome profiling dataset of vaccinia virus obtained from long-read sequencing techniques. *Gigascience* **7**, giy139 (2018).

100. Proudfoot, N. J. & Brownlee, G. G. 3′ Non-coding region sequences in eukaryotic messenger RNA. *Nature* **263**, 211–214 (1976).

101. Vijayakumar, A., Park, A. & Steitz, J. A. Modulation of mRNA 3'-End Processing and Transcription Termination in Virus-Infected Cells. *Front Immunol* **13**, 828665 (2022).

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

103. Wang, Y., Tang, Q., Maul, G. G. & Yuan, Y. Kaposi's sarcoma-associated herpesvirus ori-Lyt-dependent DNA replication: dual role of replication and transcription activator. *J Virol* **80**, 12171–12186 (2006).

104. Helmrich, A., Ballarino, M. & Tora, L. Collisions between replication and transcription complexes cause common fragile site instability at the longest human genes. *Mol Cell* **44**, 966–977 (2011).

105. Purushothaman, P., Uppal, T. & Verma, S. C. Molecular biology of KSHV lytic reactivation. *Viruses* **7**, 116–153 (2015).

106. Shekhar, R. *et al.* High-density resolution of the Kaposi's sarcoma associated herpesvirus transcriptome identifies novel transcript isoforms generated by long-range transcription and alternative splicing. *Nucleic Acids Res* **52**, 7720–7739 (2024).

107. Allen, R. D., Dickerson, S. & Speck, S. H. Identification of spliced gammaherpesvirus 68 LANA and v-cyclin transcripts and analysis of their expression in vivo during latent infection. *J Virol* **80**, 2055–2062 (2006).

108. Staudt, M. R. & Dittmer, D. P. The Rta/Orf50 transactivator proteins of the gamma-herpesviridae. *Curr Top Microbiol Immunol* **312**, 71–100 (2007).

109. Tycowski, K. T., Shu, M.-D., Borah, S., Shi, M. & Steitz, J. A. Conservation of a triple-helix-forming RNA stability element in noncoding and genomic RNAs of diverse viruses. *Cell Rep* **2**, 26–32 (2012).

110. Borah, S., Darricarrère, N., Darnell, A., Myoung, J. & Steitz, J. A. A viral nuclear noncoding RNA binds re-localized poly(A) binding protein and is required for late KSHV gene expression. *PLoS Pathog* **7**, e1002300 (2011).

Co-author certification

I, myself as a corresponding author of the following publication(s) declare that the authors have no conflict of interest, and Ákos Dörmő 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 Feb 2026

Date

Dr. Torma Gábor
shared first author

Dr. Toth Zsolt
shared last author

Prof. Dr. Boldogkői Zsolt
last author

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

Gábor Torma, **Ákos Dörmő**, Ádám Fülöp, Dóra Tombácz, Máté Mizik, Amanda M. Pretory, See-Chi Lee, Zsolt Toth, Zsolt Boldogkői.

Long-read transcriptomics of caviid gammaherpesvirus 1: compiling a comprehensive RNA atlas. mSystems 10:e01678-24. 2025 February 27. https://doi.org/10.1128/msystems.01678-24