



OPEN

## Identification of herpesvirus transcripts from genomic regions around the replication origins

Gábor Torma<sup>1,6</sup>, Dóra Tombácz<sup>1,2,6</sup>, Zsolt Csabai<sup>1,2,6</sup>, Islam A. A. Almsarrhad<sup>1,6</sup>, Gergely Ármin Nagy<sup>1</sup>, Balázs Kakuk<sup>1,2</sup>, Gábor Gulyás<sup>1,2</sup>, Lauren McKenzie Spires<sup>3</sup>, Ishaan Gupta<sup>4</sup>, Ádám Fülöp<sup>1</sup>, Ákos Dörmő<sup>1,2</sup>, István Prazsák<sup>1,2</sup>, Máté Mizik<sup>1</sup>, Virág Éva Dani<sup>1</sup>, Viktor Csányi<sup>1</sup>, Ákos Harangozó<sup>1</sup>, Zoltán Zádori<sup>5</sup>, Zsolt Toth<sup>3</sup> & Zsolt Boldogkői<sup>1,2</sup>✉

Long-read sequencing (LRS) techniques enable the identification of full-length RNA molecules in a single run eliminating the need for additional assembly steps. LRS research has exposed unanticipated transcriptomic complexity in various organisms, including viruses. Herpesviruses are known to produce a range of transcripts, either close to or overlapping replication origins (Oris) and neighboring genes related to transcription or replication, which possess confirmed or potential regulatory roles. In our research, we employed both new and previously published LRS and short-read sequencing datasets to uncover additional Ori-proximal transcripts in nine herpesviruses from all three subfamilies (alpha, beta and gamma). We discovered novel long non-coding RNAs, as well as splice and length isoforms of mRNAs. Moreover, our analysis uncovered an intricate network of transcriptional overlaps within the examined genomic regions. We demonstrated that herpesviruses display distinct patterns of transcriptional overlaps in the vicinity of or at the Oris. Our findings suggest the existence of a 'super regulatory center' in the genome of alphaherpesviruses that governs the initiation of both DNA replication and global transcription through multilayered interactions among the molecular machineries.

### Abbreviations

αHV	Alphaherpesvirus
BoHV-1	Bovine alphaherpesvirus 1
dRNA-Seq	Direct RNA sequencing
EBV	Epstein-Barr virus
HCMV	Human cytomegalovirus
IE	Immediate-early
IRL	Internal repeat of UL region
LAT	Latency-associated transcript
LRS	Long-read sequencing
ncRNA	Non-coding RNA
ORF	Open reading frame
PRV	Pseudorabies virus
SRS	Short-read sequencing
TF	Transcription factor
TR	Transcription regulator
US	Unique short
asRNA	Antisense RNA
DBP	DNA-binding protein
dcDNA-Seq	Direct cDNA sequencing

<sup>1</sup>Department of Medical Biology, Albert Szent-Györgyi Medical School, University of Szeged, Szeged, Hungary. <sup>2</sup>MTA -SZTE Lendület GeMiNI Research Group, University of Szeged, Szeged, Hungary. <sup>3</sup>Department of Oral Biology, University of Florida College of Dentistry, Gainesville, FL, USA. <sup>4</sup>Department of Biochemical Engineering and Biotechnology, Indian Institute of Technology, Delhi, India. <sup>5</sup>HUN-REN Veterinary Medical Research Institute HU, Budapest, Hungary. <sup>6</sup>These authors contributed equally: Gábor Torma, Dóra Tombácz, Zsolt Csabai and Islam A. A. Almsarrhad. ✉email: boldogkoi.zsolt@med.u-szeged.hu

EHV-1	Equid alphaherpesvirus 1
HHV-6	Human herpesvirus 6
ICP	Infected cell polypeptide
IRS	Internal repeat of US region
LLT	Long latency transcript
L/ST	L/S junction-spanning transcript
ONT	Oxford nanopore technologies
Ori	Replication origin
raRNA	Replication origin-associated RNA
SVV	Simian varicella virus
TI	Transcript isoform
TSS	Transcript start site
UTR	Untranslated region
$\beta$ HV	Betaherpesvirus
DNP	DNA polymerase
E	Early
$\gamma$ HV	Gammaherpesvirus
HSV-1	Herpes simplex virus 1
IR	Inverted repeat
L	Late
lncRNA	Long noncoding RNA
miRNA	Micro RNA
ORC	Origin recognition complex
PacBio	Pacific biosciences
RNP	RNA polymerase
TES	Transcript end site
TO	Transcriptional overlap
UL	Unique long
VZV	Varicella-zoster virus

### Regulation of herpesvirus transcription

The lytic transcription of herpesviruses follows a sequential order, which is divided into three distinct temporal phases: immediate-early (IE), early (E), and late (L)<sup>1</sup>. The IE gene products activate the transcription of both E genes, which encode enzymes of DNA replication, and L genes that produce structural proteins. Four IE proteins participate in the transcription regulation of herpes simplex virus type 1 (HSV-1), a representative member of alphaherpesviruses ( $\alpha$ HVs). The essential ICP4 viral protein, encoded by *rs1* (*icp4*) serves as the main transcription regulator (TR), recruiting cellular transcription factors (e.g. TFIID) to viral promoters to enhance transcription initiation<sup>2</sup>. ICP22, encoded by *us1* promotes the transcription elongation<sup>3</sup>. The *us1* gene of HSV-1 is located in the unique short (US) genomic region in a single copy, while its promoter is represented in duplicate in the inverted repeat (IR) region (the other IR copy controls the expression of *ul12* gene). In the Varicellovirus genus, the *us1* gene is translocated to the IR region making it present in duplicate. ICP0, encoded by *rl2* (*icp0*), is not a transcription factor in the strict sense, but it can promote transcription by affecting pre-chromatin interactions before histones bind to viral DNA<sup>4</sup>. ICP27, encoded by *ul54* participates in recruiting RNA polymerase (RNP) to viral promoters<sup>5</sup> and DNA synthesis<sup>6</sup>.

The mechanism controlling the lytic cycle in other  $\alpha$ HVs is comparable to that of HSV-1. One significant difference is the expression kinetics of *icp0*, *us1* and *ul54* orthologs, which have evolved to be expressed in the E kinetic phase in pseudorabies virus (PRV) and equid alphaherpesvirus type 1 (EHV-1). Additionally, the *icp0* gene has been structurally and functionally simplified in these viruses. Betaherpesviruses ( $\beta$ HVs) and gamma-herpesviruses ( $\gamma$ HVs) employ similar mechanisms as  $\alpha$ HVs for controlling genome-wide viral transcription. In human cytomegalovirus (HCMV), the prototype member of  $\beta$ HVs, two major IE genes (*ie1* and *ie2*) regulate global viral transcription<sup>7</sup>. The IE proteins of Epstein-Barr virus (EBV; the representative member of  $\gamma$ HVs), designated as BZLF1 and BRLF1, are transactivators that initiate the transcription of E viral genes<sup>8</sup>.

An alternative way herpesviruses operate is by establishing latency, during which the majority of the viral genome remains transcriptionally silent, with only a few specific viral RNAs being expressed. The latency-associated transcript (LAT) of HSV-1 is the sole viral gene product that is highly expressed during the latent state<sup>9</sup>. This non-coding RNA (ncRNA) represses the lytic gene expression through blocking the activity of *icp4*<sup>10</sup> and facilitating heterochromatin formation on the HSV-1 genome<sup>11</sup>. Although LAT contains numerous open reading frames (ORFs), it likely does not encode any polypeptides<sup>12,13</sup>. Other long ncRNAs (lncRNAs) are also expressed during latency, such as the long-latency transcript (LLT; overlapping *icp0* and *icp4* genes<sup>14</sup>) and the L/S junction-spanning transcripts (L/STs; overlapping *icp34.5* and *icp4* genes<sup>15</sup>).

Members of the non-coding NOIR-1 transcript family, described in  $\alpha$ HVs [PRV<sup>16</sup>, varicella-zoster virus (VZV)<sup>17</sup>, and EHV-1<sup>18</sup>] are 3'-coterminal with the LLT transcripts and are expressed during the lytic cycle. The low-abundance NOIR-2 RNA, which is transcribed in a reverse orientation relative to NOIR-1, has only been detected in PRV<sup>16</sup>. ELIE, another lytic lncRNA of PRV<sup>19</sup>, partially overlaps the long isoform of HSV-1 L/ST.

## DNA replication

While the prokaryotic genomes contain a single start site of DNA synthesis (designated as replication origin or Ori) defined by consensus sequences<sup>20</sup>, eukaryotic genomes typically have tens of thousands Oris specified by their chromatin structure<sup>21,22</sup>. Viruses have a single or a few Oris, which are specified by a combination of structural properties and sequence specificity of the particular DNA segment<sup>23</sup>. The replication of eukaryotic genomes is initiated by the binding of the origin recognition complex (ORC) to the Ori<sup>24</sup>. The function of ORC is to serve as a platform for the assembly of the replisome, which is composed of a wide range of proteins such as DNA helicase, DNA polymerase (DNP), topoisomerase, primase, DNA gyrase, single-stranded DNA-binding protein (ssDBP), RNase H, DNA ligase, and telomerase enzymes.

Herpesviruses encode several replication proteins needed for DNA synthesis. For example, HSV-1 codes for an origin-binding protein (OBP) (*ul9*), an ssDBP (*ul29*), two DNPs (*ul30* and *ul42*), and three helicase/primase molecules (*ul5*, *ul8*, and *ul52*)<sup>25,26</sup>. Several viral auxiliary factors play roles in the nucleotide metabolism, such as ribonucleotide reductase (*ul39/40*), thymidine kinase (*ul23*), uracil-DNA glycosylase (*ul2*), deoxyuridine triphosphatase (*ul50*), alkaline nuclease (*ul12*), allowing herpesvirus replication in non-dividing cells<sup>27</sup>.

HSV-1 contains three Oris, two within the IRs surrounding the US region (termed OriS), and one in the unique long (UL) region (termed OriL) (Supplementary Fig. 1). In Simplexviruses, OriL is located between two E genes, *ul29* and *ul30*, involved in DNA replication, while the two copies of OriS are surrounded by the IE genes, *icp4* and *us1*, within the internal repeat of the US region (IRS), and by *icp4* and *us12* genes within the terminal repeat of the US region (TRS). Some members of Varicelloviruses, such as VZV and Bovine alphaherpesvirus 1 (BoHV-1) lack an OriL. However, in other Varicelloviruses like PRV and EHV-1, the position of OriL has shifted to the intergenic region between the *ul21* and *ul22* gene pair. HCMV has only a single replication origin (OriLyt), located at a semi-orthologous position, next to *ul57* homologous to the HSV-1 *ul29*. Human gamma-herpesviruses, EBV and Kaposi's sarcoma herpesvirus (KSHV), have two lytic (OriLyt-L and OriLyt-R), and a latent replication origin (termed OriP in EBV and terminal repeat in KSHV)<sup>28–31</sup>. The EBV OriP comprises two main components: the dyad symmetry (DS) and the family of repeats (FR). Each of them hosts multiple binding sites for the protein EBNA-1. When EBNA1 binds to the DS, it functions as an Ori, facilitating the recruitment of ORC<sup>32</sup>. DNA replication in its lytic phase relies on seven EBV replication proteins. BZLF1 binds to the OriLyt. It has the ability to bind multiple viral replication proteins and thereby initiating the lytic phase. The BALF5 protein encodes the catalytic subunit of DNP and interacts with the helicase-primase complex. BALF2 acts as the ssDBP. The BMRF1 protein is a DNP accessory subunit, which can serve as a coactivator of BZLF1 and it interacts with BALF5, forming the DNP holoenzyme<sup>33</sup>.

## Overlapping viral transcripts

Recent studies<sup>34</sup> have shown that every herpesvirus gene forms various transcriptional overlaps (TOs), including divergent (head-to-head), convergent (tail-to-tail), and parallel (tail-to-head). Tandem genes form parallel-overlapping multigenic, 3'-coterminal transcripts, representing the archetypal genomic organization of herpesviruses. Moreover, many viral genes express 5'-truncated transcripts with different transcription start sites (TSSs) but the same transcription end site (TES), containing 'nested' open reading frames (ORFs) which encode N-terminally-truncated polypeptides<sup>35,36</sup>. Most co-located divergent genes produce 'hard' TOs where the canonical transcripts overlap each other. However, in a few cases, only the long transcript isoforms (TIs), create head-to head TOs ('soft' TOs), not the canonical transcript. Convergently oriented genes form 'soft' TOs through transcriptional read-through, but only a few 'hard' TOs can be observed (e.g. in  $\alpha$ HVs, the *ul7/8*, *ul30/31*, and *ul50/51* gene pairs<sup>37</sup>).

## Non-coding RNAs regulating DNA replication

There is growing evidence that certain non-coding transcripts, including short ncRNAs, such as microRNAs (miRNAs<sup>38</sup>), and lncRNAs, play essential roles in regulating DNA replication<sup>39</sup>. For instance, a specific type of lncRNAs encoded by sequences near the Oris has been found in all three domains of life and in viruses over the past decade. A recent survey revealed that approximately 72% of mammalian ORC1s are associated with active promoters, more than half of them controlled by ncRNAs<sup>40</sup>. Replication RNAs have several modes for controlling DNA replication. These include the regulation of RNA primer synthesis through hybridization with DNA sequences<sup>9</sup>, or the formation of hybrids with mRNAs. This latter process initiates their degradation by RNase H, thereby inhibiting the translation of replication proteins. Additionally, these transcripts can help recruit ORC to the Ori<sup>41</sup>.

## Replication origin-associated herpesvirus transcripts

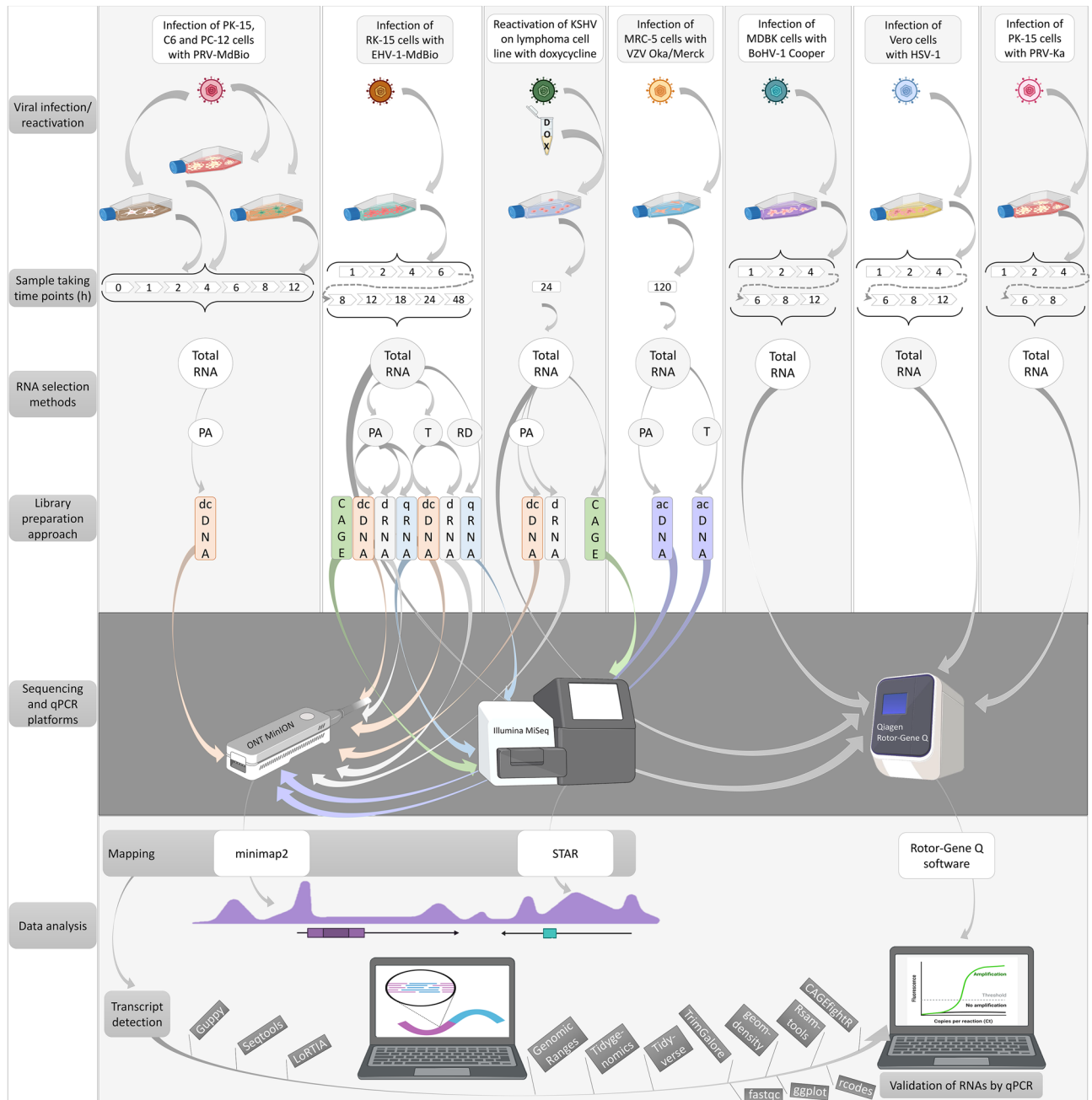
Replication origin-associated RNAs (raRNAs) have been identified in all three subfamilies of herpesviruses. While many of these transcripts have previously been described in  $\beta$ HVs and  $\gamma$ HVs, they were practically undetected in  $\alpha$ HVs until recently. For example, the non-coding RNA4.9 of HCMV, transcribed from the OriLyt, has been identified<sup>42</sup> and functionally characterized<sup>43</sup>. This transcript regulates viral DNA replication through the formation of a DNA:RNA hybrid and also affects the expression level of ssDBP encoded by the *ul57* gene. RNA4.9 is thought to have additional roles, operating in both *cis* and *trans*, such as promoting transcriptional repression of the major IE promoter during latency<sup>44</sup>. The discovery of RNA4.9 and other HCMV raRNAs led to the belief that this virus has a unique mode of replication regulation, which has not proven to be true. Other raRNAs (SRT, vRNA-1 and vRNA-2) overlapping the OriLyt have also been described in HCMV<sup>45,46</sup>. The formation of an RNA:DNA hybrid by the BHLF1 replication RNA at the EBV OriLyt region has been described earlier as well<sup>47</sup>. Additionally, a bidirectional promoter<sup>48</sup> and a highly structured RNA were identified within this region of EBV<sup>44</sup>. The function of this latter transcript is to assist the viral EBNA1 and HMGA1a proteins in recruiting ORC<sup>49</sup>. Incomplete sequences of two co-terminal lncRNAs near the HSV-1 OriS have been described previously<sup>50</sup>.

The emergence of long-read sequencing (LRS) technologies has greatly accelerated the discovery of novel viral transcripts and their TIs, including splice, TSS and TES variants. These investigations have identified numerous lncRNAs near both the OriS and OriL regions of  $\alpha$ HVs<sup>34,51–53</sup>. However, since no precise function has been linked to these transcripts, it would be inaccurate to term them as ‘replication RNAs’ as their role in DNA replication is still undisclosed. Moreover, genes surrounding the OriS have been shown to produce TIs with long 5'-untranslated regions (5' UTRs—TSS isoforms), or 3' UTRs (TES isoforms) that overlap the replication origin<sup>54,55</sup>.

## Results

### Multiplatform sequencing for characterization of herpesvirus transcripts

In this study, we applied newly generated (Fig. 1) and previously published sequencing data (Table 1) for analyzing the Ori-proximal transcripts of nine herpesviruses. The following approaches were used for data production:



**Figure 1.** Workflow. Here, we outline the techniques employed to generate new sequencing data, including infection of cells with various viruses, library preparation, sequencing and bioinformatics. The qRT-PCR validation workflow for several transcripts is also depicted. Abbreviations: PA: polyadenylated RNAs; T: Terminator-handled samples; RD: ribodepleted RNAs; dcDNA: direct cDNA-seq; CAGE: Cap Analysis Gene Expression-Seq; dRNA: direct RNA-Seq; qRNA: short-read RNA-Seq (library generated by qRNA-seq kit); acDNA: amplified cDNA-Seq.

Virus	Sequencing approach	Library	References	Data availability
<i>BoHV-1</i>	LRS ONT	direct RNA	53	ENA: PRJEB33511
	LRS ONT	direct cDNA		
	LRS ONT	amplified cDNA		
	Synthetic LRS Illumina	LoopSeq		
	LRS ONT	direct cDNA	74	ENA: PRJEB33511
<i>EBV</i>	LRS PacBio RSII	amplified cDNA	60,77	GSE: GSE79337
	SRS Illumina	amplified cDNA CAGE		
	LRS ONT	direct cDNA	84	ENA: PRJEB38992
		amplified cDNA		
<i>HCMV</i>	LRS PacBio RSII	amplified cDNA	78	ENA: PRJEB22072
	LRS PacBio Sequel	amplified cDNA	83	ENA: PRJEB25680
	LRS ONT	amplified cDNA		
	LRS ONT	direct RNA		
	LRS ONT	CAP-selected	36	ENA: PRJEB25680
<i>HSV-1</i>	SRS Illumina	amplified cDNA	56	GEO: GSE59717
		amplified cDNA	57	SRA: PRJNA505045
		amplified cDNA	58	SRA: PRJNA482043, PRJNA483305, and PRJNA533478
		amplified cDNA	59	GEO: GSE128324
	LRS PacBio Sequel	amplified cDNA	79	ENA: PRJEB25433
	LRS ONT	amplified cDNA direct RNA		
	LRS ONT	direct RNA	80	ENA: PRJEB27861
	LRS PacBio RSII	amplified cDNA	37	GEO: GSE97785
	<i>PRV</i>	SRS Illumina	amplified cDNA	62
LRS PacBio RSII		direct cDNA	16	ENA: PRJEB12867
		amplified cDNA		
LRS PacBio Sequel		amplified cDNA	37	ENA: PRJEB24593
LRS ONT		direct RNA		
LRS ONT		amplified cDNA		
LRS ONT		amplified Cap-selected		
LRS ONT		Terminator-handled amplified cDNA	55	ENA: ERP106430 and ERP019579
LRS ONT		direct cDNA		
LRS ONT	direct RNA			
<i>SVV</i>	LRS ONT	direct RNA	61	ENA: PRJEB42868
	SRS Illumina	amplified cDNA		
<i>VZV</i>	LRS ONT	amplified cDNA	17	ENA: PRJEB25401
		Targeted		
	LRS ONT	amplified Cap-selected	81	ENA: PRJEB25401
	SRS Illumina	amplified cDNA	75	ENA: PRJEB38829
	SRS Illumina	CAGE		
LRS ONT	direct RNA			

**Table 1.** Techniques and datasets from earlier publications used in this study. This table show the annotated transcripts mapping to the genomic loci examined in this study.

SRS on different Illumina platforms<sup>51,56–59</sup>, as well as LRS on ONT (MinION<sup>18</sup>), Pacific Biosciences (PacBio—RSII and Sequel<sup>16,60</sup>), and LoopSeq<sup>53</sup> platforms along with a wide range of library preparation techniques and Cap Analysis of Gene Expression sequencing (CAGE-Seq, conducted on an Illumina platform) for VZV<sup>61</sup> and EBV<sup>60</sup>. The terminology for orthologous genes varies among the  $\alpha$ HVs. To enhance comparability, we have adopted the naming convention used for HSV-1 (Supplementary Table 1). Our procedure for annotating and verifying transcripts was as follows. We first identified a sequencing read obtained by direct cDNA sequencing (dcDNA-Seq) across all three biological replicates. The TSS and TES of this read were identified by CAGE-Seq, RAMPAGE, and direct RNA sequencing (dRNA-Seq). Additionally, we required that the transcript itself be detectable by dRNA-Seq and quantitative reverse transcription PCR (qRT-PCR). For intron annotation, their detection via dRNA-Seq was necessary. However, as we also relied on previously published data, we could not employ all these



methods in every virus and dataset. In the respective figures, we indicated which techniques were utilized for the detection and validation of the given transcripts.

In the first part of this study, we precisely annotated the genomic locations of canonical transcripts and their TIs, including TSSs, TESs, and splice variants. Previous annotations were either verified, or in several instances, revised based on the integration and reassessment of sequencing datasets. We also identified cis-regulatory elements for numerous RNA molecules. Notably, we identified promoter elements (TATA boxes) within the Ori regions in all cases. Our objective was to offer a comprehensive view of the complexity of TOs in the investigated genomic regions. In viruses lacking either dRNA-Seq or CAGE data, we employed stricter criteria for transcript annotation, resulting in lower transcriptomic complexity in these cases. Relative transcript abundances were determined for viruses with sufficient data. Furthermore, a multi-time-point real-time RT-PCR (RT<sup>2</sup>-PCR) analysis was applied to monitor the expression kinetics of the three most important lncRNAs in PRV.

We also applied RT<sup>2</sup>-PCR for the validation of lncRNAs and longer mRNA isoforms of PRV, BoHV-1, EHV-1, HSV-1, and KSHV. Native RNA sequencing was employed to validate the results of cDNA sequencing. CAGE-Seq was applied for the validation of KSHV and EHV-1 TSSs obtained by ONT sequencing. It is worth noting that the size-biasing effect of library preparation and LRS techniques may lead to non-detection or considerable underestimation of the abundance of long (> 5 kb) transcripts. Our RT<sup>2</sup>-PCR analysis demonstrated that despite this bias, the long transcripts are indeed expressed at lower levels.

### raRNAs: transcripts proximal to replication origins

Transcripts that overlap or map proximal to the Oris include lncRNAs, as well as long 5' and 3' UTR isoforms of mRNAs (Supplementary Table 2 and FigShare: <https://doi.org/10.6084/m9.figshare.22339879.v1>).

#### *Alphaherpesviruses—OriS*

In this study, we detected novel transcripts and TIs near the OriSs of  $\alpha$ -HVs, including lncRNAs such as OriS-RNA of BoHV-1, OriS-RNA1 of HSV-1, NOIR-1 transcripts of PRV, EHV-1, VZV and Simian Varicellovirus (SVV), and NOIR-2 transcript of PRV (Figs. 2 and 3 and Supplementary Figs. 2–5). We found that in all examined  $\alpha$ HVs, the very long 5' TIs of transcription regulator genes (*us1* and *icp4*) of BoHV-1, EHV-1, HSV-1, and SVV overlap the OriS. We cannot exclude the possibility that this is the case for other  $\alpha$ HVs, but these putative transcripts might have gone undetected due to their low abundance. We detected a very complex splicing pattern of US1 transcripts in  $\alpha$ HVs. Novel lncRNAs oriented antisense to the HSV-1 OriS-RNA1 were also identified. The NOIR-1 family members display a divergent configuration in relation to the *icp4* gene. The canonical versions of these RNAs do not overlap *icp4*, whereas the longer NOIR-1 variants partially overlap this major TR gene. In EHV-1, VZV, SVV, and probably in PRV, a long TI of US1 originates from the promoter of the *noir-1* gene. In SVV, we discovered a long TSS variant of NOIR-1 that overlaps the canonical ICP4 transcript. Both this variant and the canonical NOIR-1 transcript overlap the OriS. NOIR-1 is expressed at a moderate level, whereas NOIR-2 exhibits very low-abundance. In the same genomic region of VZV, we detected five lncRNAs with distinct TSSs and TESs designated them as NOIR-1A, -1B, 1C, -1D, and -1E. We found TIs with TSSs mapping very close to the TATA boxes within the OriSs in all six  $\alpha$ HVs, suggesting that these promoter elements are functional.

#### *Alphaherpesviruses—OriL*

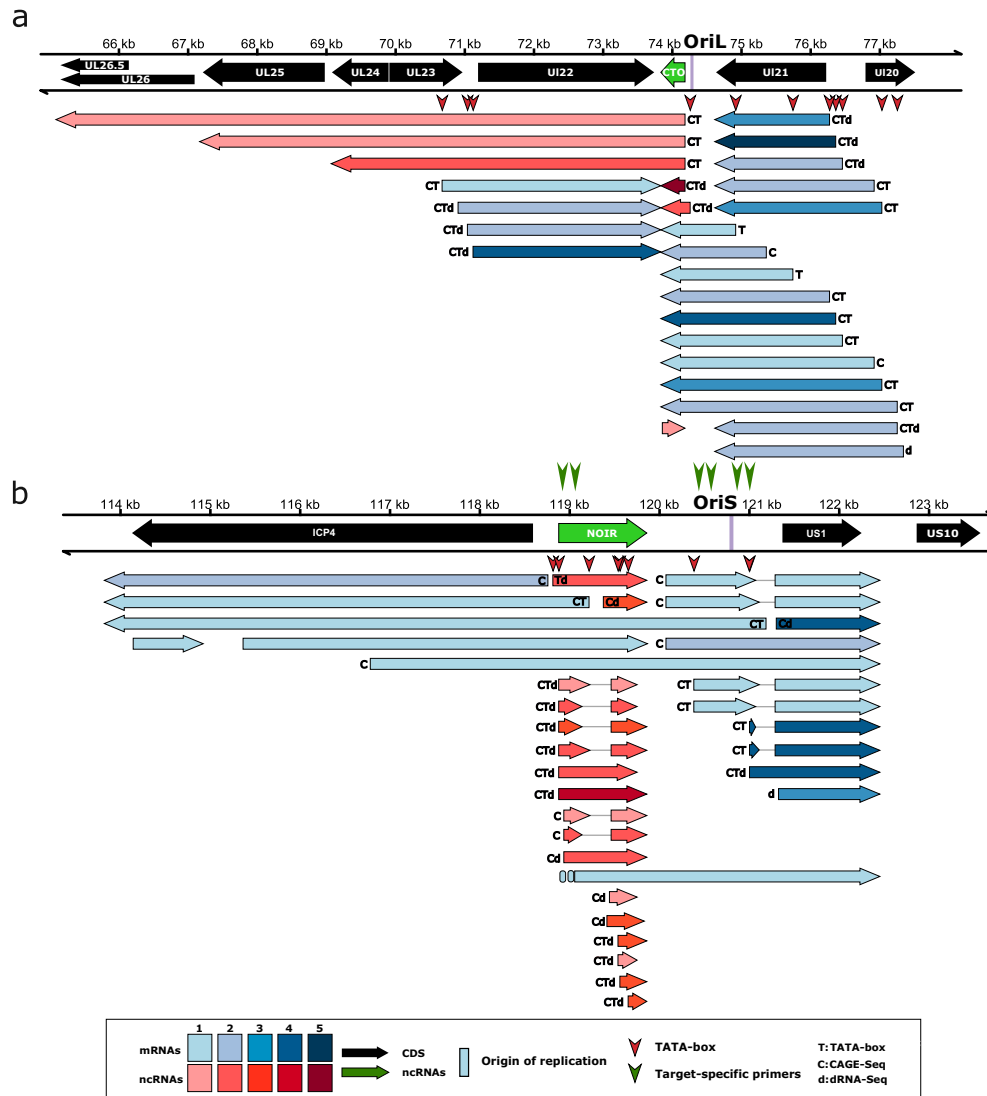
The CTO transcript family has been described in PRV in our earlier works<sup>16,62</sup>. In these initial publications, we identified three 3'-coterminal transcripts: the short CTO-S, the CTO-M [initiated near the poly(A) signal of *ul21* gene], and CTO-L [a transcriptional read-through isoform (3' UTR variant) encoded by the *ul21* gene]. It is worth noting that the long 3' UTR isoforms with unique TES are exceedingly uncommon in  $\alpha$ HV mRNAs. The transcript family list was later updated<sup>55</sup>, with a more comprehensive version provided in this current report (Supplementary Fig. 3). We also detected CTO transcripts in EHV-1 (Fig. 3), but not in other herpesviruses with annotated transcriptomes. Both PRV and EHV-1 exhibit a notably high CTO-S expression. We observed a tail-to-tail (convergent) transcriptional overlap (TO) between the 3' UTR isoforms of CTO-S and UL22 transcripts, and identified very long read-through CTO transcripts in both viruses. We examined two PRV strains, the laboratory strain Kaplan (PRV-Ka<sup>63</sup>) and a field isolate (strain MdBio: PRV-MdBio<sup>64</sup>). In HSV-1, the divergent *ul29-ul30* gene pair members produce long 5' UTR variants that overlap the OriL, but no lncRNA was found near the HSV-1 OriL.

#### *Betaherpesviruses*

Our data show that RNA4.9<sup>42</sup> is initiated from the OriLyt of HCMV (Supplementary Fig. 6). UL59<sup>65</sup>, SRT<sup>45</sup> and vRNA-2<sup>46</sup> were also identified using our previously published dataset<sup>36</sup>. Additionally, we detected two longer versions of UL58 lncRNA and a shorter isoform of UL59 lncRNA.

#### *Gammaherpesviruses*

The long 5' UTR isoform of EBV *BCRF1* gene overlaps the OriP<sup>66</sup> (Supplementary Fig. 7a). Likewise, the long 5' UTR variants of *BHRF1* gene overlap OriLyt-L. One of these transcripts is also a splice isoform encoded by this gene. The promoter of *BHLF1* gene is located within the OriLyt-L<sup>47</sup>. Here we describe novel isoforms of lncRNAs that either overlap the OriLyt-R with their introns or are initiated from the Ori. Our analyses revealed that several OriLyt-L-associated ncRNAs of varying length can be produced from the same TSS besides the 1.4-kb lncRNA during lytic reactivation of KSHV (Fig. 4). The OriLyt-L is flanked by short protein-coding genes such as K4.2, K4.1 and K4 on the left and K5, K6, and K7 on the right side<sup>67</sup>. Previous studies indicated that K4, K4.1, and K4.2 are transcribed as mono-, bi-, and tricistronic mRNAs<sup>68</sup>, but our analysis show a more complex expression pattern that includes unspliced RNAs of varying length and spliced RNA variants as well. We also found

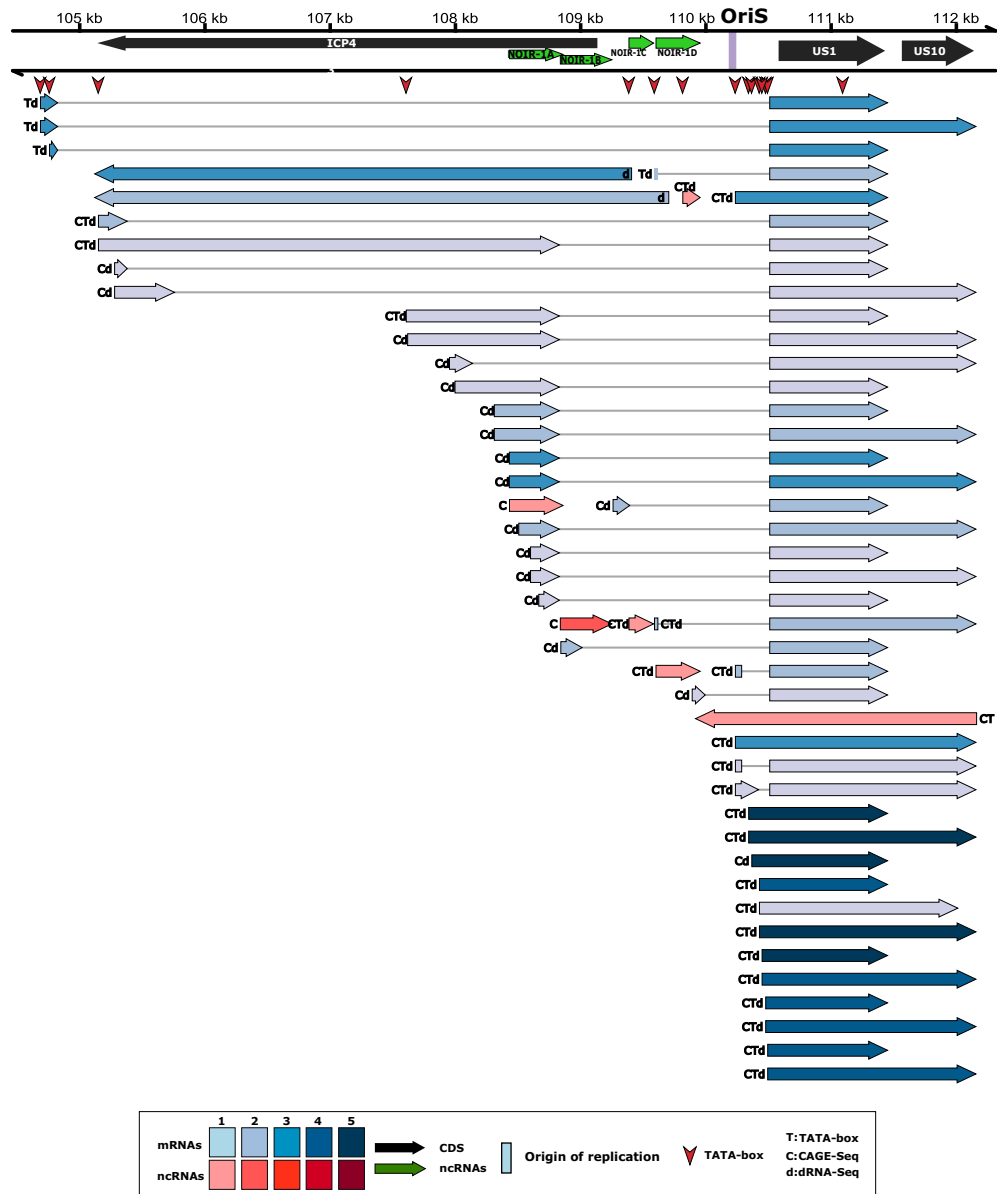


**Figure 2.** Ori-proximal transcripts of EHV-1. This figure illustrates the transcripts encoded by the OriL- (a) and OriS- (b) proximal regions of Equid alphaherpesvirus 1. For the library preparation, both polyA-selected and ribodepleted samples were used. However, in both cases, the RNAs were reverse transcribed to cDNA using an oligo(dT) primer. All the putative transcripts were identified by LoRTIA software using dcDNA datasets unless otherwise stated. Protein-coding genes are marked with black arrows, non-coding genes with green arrows, mRNAs with blue arrows, and ncRNAs with red arrows. For better comparability, we use the names for the genes applied in HSV-1 terminology. The relative transcript abundance is indicated by shading. Shades represents relative abundance: 1: 1–9 reads, 2: 10–49 reads, 3: 50–199 reads, 4: 200–999 reads, 5: > 1000 reads. CAGE-Seq was performed (transcripts detected by this technique are marked with a ‘C’). Transcripts with a proximal TATA box are labeled by a ‘T’ letter, and the vertical red arrows (▼) show TATA box positions on the genome. Transcripts also detected by dRNA-Seq are marked with a ‘d’ letter at upstream positions. Introns are represented by horizontal lines. The position of PCR primers are indicated by vertical green arrows (↕).

that K5/K6 genes are expressed not only individually but also through splicing, which results in mRNAs whose first exon is of different length. Importantly, our results are in line with previous transcriptomic studies<sup>69–71</sup> but also expand the number of different viral RNA transcripts produced from the OriLyt-L locus. The latency locus of KSHV encode four protein-coding genes [K12, K13, ORF72, ORF73 (LANA)] and twelve pre-micro RNAs (miRNAs)<sup>71–73</sup>. Here, we detected several lncRNAs that are antisense to the miRNA-coding genomic regions.

### RNA isoforms of transcription regulator genes

The *us1* gene produces very long 5' UTR isoforms, which were observed to create head-to-head TOs with ICP4 transcripts in EHV-1 (Fig. 2), VZV (Fig. 3), and HSV-1 (Supplementary Fig. 5). In HSV-1, the 5' UTR isoforms of US10-12 polycistronic transcripts were found to establish a divergent TO with the *icp4* gene. Additionally, in HSV-1, we detected an ICP4 TIs with an extended 5' UTR that overlaps the *us1* gene. In BoHV-1, a 3' UTR



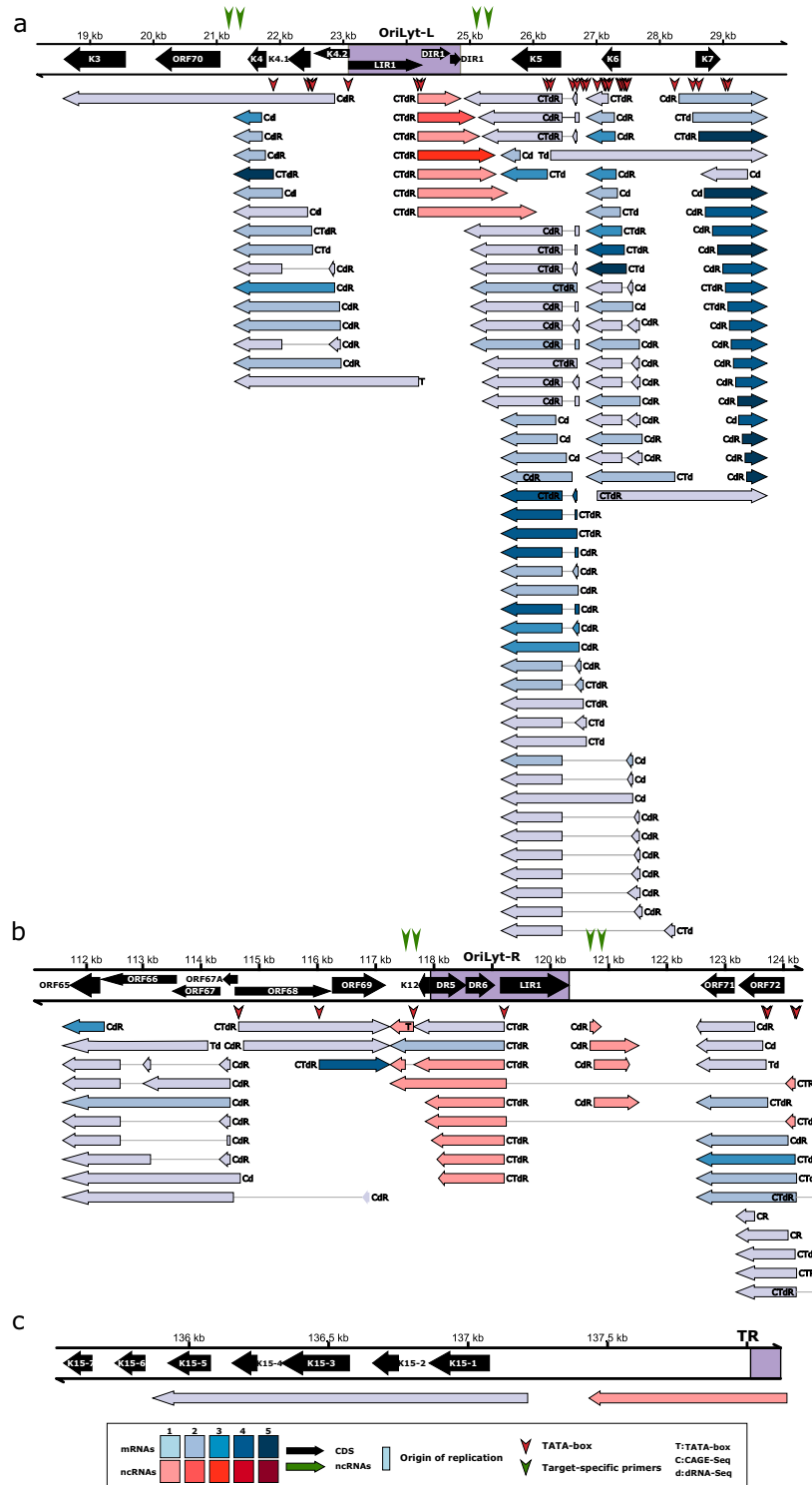
**Figure 3.** OriL-proximal transcripts of VZV. This figure shows the transcripts encoded by the OriS-proximal region of Varicella-zoster virus. All putative transcripts were identified by LoRTIA software using dcDNA datasets unless otherwise stated. Protein-coding genes are labeled with black arrows, non-coding genes with green arrows, mRNAs with blue arrows, and ncRNAs with red arrows. For better comparability, we have adopted the naming convention used for HSV-1 genes. Relative transcript abundance is indicated by shading. Shades represent relative abundance: 1: 1–9 reads, 2: 10–49 reads, 3: 50–199 reads, 4: 200–999 reads, 5: > 1000 reads. Transcripts with a proximal TATA box are marked by a ‘T’ letter, and those that were also detected by dRNA-Seq are marked with a ‘d’ letter at the upstream positions. Vertical red arrows (T) show the positions of TATA boxes on the genome. Transcripts detected by CAGE-Seq are marked with a ‘C’ letter. Introns are represented by horizontal lines.

isoform of *icp4* has been reported to create a parallel TO with the downstream *icp0* gene 1, but, the *icp4* ORF is spliced out from this transcript, resulting in a chimeric RNA containing the full-length *icp0* gene and a portion of 5’ UTR of the ICP4 transcript<sup>74</sup>. Additionally, ICP4 TIs were found to generate similar chimeric and bicistronic transcripts with the BoHV-1 CIRC RNA, which maps to the adjacent genomic locus in the circular or concatemeric viral genome.

**Non-coding RNAs mapping near the transcription regulator genes**

In this work, we detected antisense RNAs (asRNAs) overlapping the *us1* gene in BoHV-1 and PRV. Previously, ELIE was only identified in PRV, but we found a transcript with a similar genomic location in EHV-1 (Fig. 2). ELIE is situated between the *icp4* and *icp0* genes, with one of its TIs being 5’-coterminial with NOIR-1 transcripts.





**Figure 4.** Ori-proximal transcripts of KSHV. This image illustrates the transcripts specified by the Ori-proximal regions of Kaposi's sarcoma-associated herpesvirus (OriLyt-L: **a**; OriLyt-R: **b**; TR: **c**). All putative transcripts were identified by LoRTIA software using dcDNA datasets unless otherwise stated. Protein-coding genes are marked with black arrows, non-coding genes with green arrows, mRNAs with blue arrows, and ncRNAs with red arrows. Relative transcript abundance is indicated by shading. Shades represent relative abundance: 1: 1–9 reads, 2: 10–49 reads, 3: 50–199 reads, 4: 200–999 reads, 5: >1000 reads. Transcripts with a proximal TATA box are marked by a 'T' letter, and those that were also detected by dRNA-Seq are marked with a 'd' letter at the upstream positions. The vertical red arrows (▼) indicate the positions of TATA boxes on the genome. Transcripts detected by CAGE-Seq are marked by a 'C' letter. RAMPAGE data were also available for KSHV (transcripts detected by this technique are marked by an 'R' letter). Introns are represented by horizontal lines. The position of PCR primers are indicated by vertical green arrows (↕).

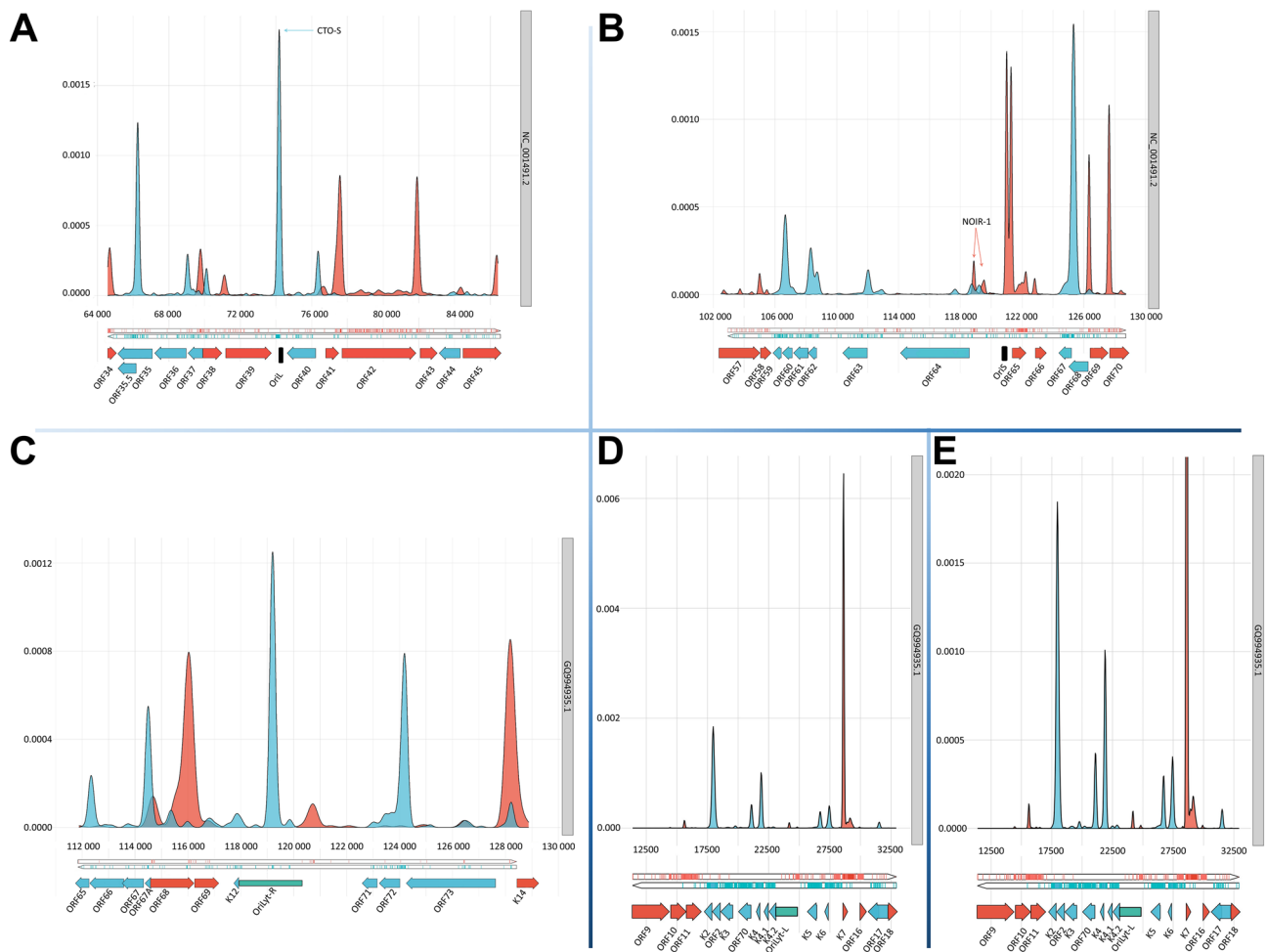
We also identified an asRNA in EHV-1, named as64, which shares the same orientation as PRV ELIE but it is located within the *icp4* gene. Here, we also describe an HSV-1 transcript that starts at the 3' end of *icp4* gene and terminated at the *us1* gene. Moreover, we detected a TSS of long 5' UTR isoform of the VZV *us1* gene, downstream of *icp4* gene, at a position homologous to the TSS of PRV ELIE. In PRV, another lncRNA called AZURE is oriented opposite to the *us1* gene.

### Transcriptional overlaps of replication genes

In Simplexviruses, the long 5' UTR isoforms of divergent *ul29-ul30* gene pair not only overlap the OriL but also each other (Supplementary Fig. 5). Intriguingly, both genes encode the main regulators of DNA replication. In both HCMV (*ul57*) and human herpesvirus type 6 (HHV-6) (*ul42*), the *ul29* orthologs are found adjacent to the OriLyt. Remarkably, in  $\alpha$ HVs, three 'hard' TOs between gene pairs exist, with one partner always being a gene involved in viral replication. These TOs include *ul30/ul31*, *ul6-7/ul8-9*, *ul50/ul51* (*ul30*: DNP; *ul8*: DNA helicase; *ul9*: OBP; *ul50*: deoxyuridine triphosphatase).

### Defining the TSS patterns of examined genomic regions

Determining the TSS of RNA molecules is a significant challenge in transcriptome research. In addition to using various LRS and SRS approaches, we conducted CAGE-Seq for EHV-1 and KSHV (Fig. 5) and utilized CAGE data from others for VZV<sup>75</sup> and EBV<sup>60</sup> for the analysis (Figs. 2, 3 and 4). We compared our KSHV TSS



**Figure 5.** TSS Distribution in examined genomic regions determined by CAGE-Seq. TSS distributions are illustrated in the following genomic regions of EHV-1 and KSHV: (A) EHV-1 OriL; (B) EHV-1 OriS of the IRS; (C) KSHV OriLyt-R; (D) KSHV OriLyt-L; (E) KSHV OriLyt-L. A higher resolution is used for the better visibility of the low-abundance TSSs. CTO-S transcript is highly expressed (A), whereas NOIR-1 is a group of relative low-expressed transcripts (B). Smoothed density plots of the 5' ends in the CAGE data. The y-axis shows the probability estimation of the 5' ends using a probability density function (details in the Materials and Methods section). Coding sequence annotations for the respective genomes (displayed with the accession number on the right) are visualized in the lower part. Positive strand coverage and the coding sequence annotation are shown in red, while in KSHV they are depicted in blue in the negative strand. The Ori regions in EHV-1 are shown in black, whereas in KSHV they are depicted in green. In the latter case, an accompanying white box displays the 20-nt binding site for the DNA replication origin-binding protein.

CAGE-Seq results with RAMPAGE-Seq results published by others<sup>76</sup>. We described a total of 199 KSHV transcripts, of which 192 were confirmed by CAGE and 159 by RAMPAGE. All TSSs detected by RAMPAGE were also identified by CAGE.

### Transcript validation using qRT-PCR

Fifteen transcripts of five viruses were further validated using qRT-PCR (Fig. 6). The Ct values clearly indicate that the long TSS isoforms of the TR genes overlapping the Ori and each other are expressed at low-levels.

### Transcription kinetics of three lncRNAs of PRV

In this work, we performed temporal sequence analyses on three PRV lncRNAs namely CTO-S, NOIR-1, and AZURE, along with three early transcripts (EP0, UL50, UL51) and four late-2 (L2) transcripts (UL19, UL25, UL47, UL48) utilizing both untreated (UT) and phosphoamino acid (PAA)-treated specimens (PAA acts as a blocker of DNA replication) (Fig. 7, Supplementary Table 3). Each specimen was generated in three biological replicates and assessed using dcDNA-Seq. We found that the PAA treatment suppressed the overall transcript levels; therefore, we adjusted the number of transcripts relative to the total viral RNAs. The effect of the treatment on the transcript abundancies was evaluated using the PAA/UT ratio. Logically, a value less than 1 indicates an L2 expression kinetic. Additionally, we evaluated these specimens at 12 h post-infection using qRT-PCR, yielding results that aligned closely with our findings using dcDNA-Seq analysis (Table 2). Based on our observations, it is evident that PAA treatment exerted the most pronounced effect on L2 genes, manifesting in a notable reduction in their relative abundance within the overall viral reads when compared to untreated specimens. Conversely, early transcripts manifested elevated ratios, especially during later stages, in the PAA-treated samples. While CTO-S exhibited an obvious early kinetic characteristic, the remaining two lncRNAs displayed complex expression trends. Notably, AZURE is characterized by low expression levels, which might explain its irregular expression dynamics.

### Discussion

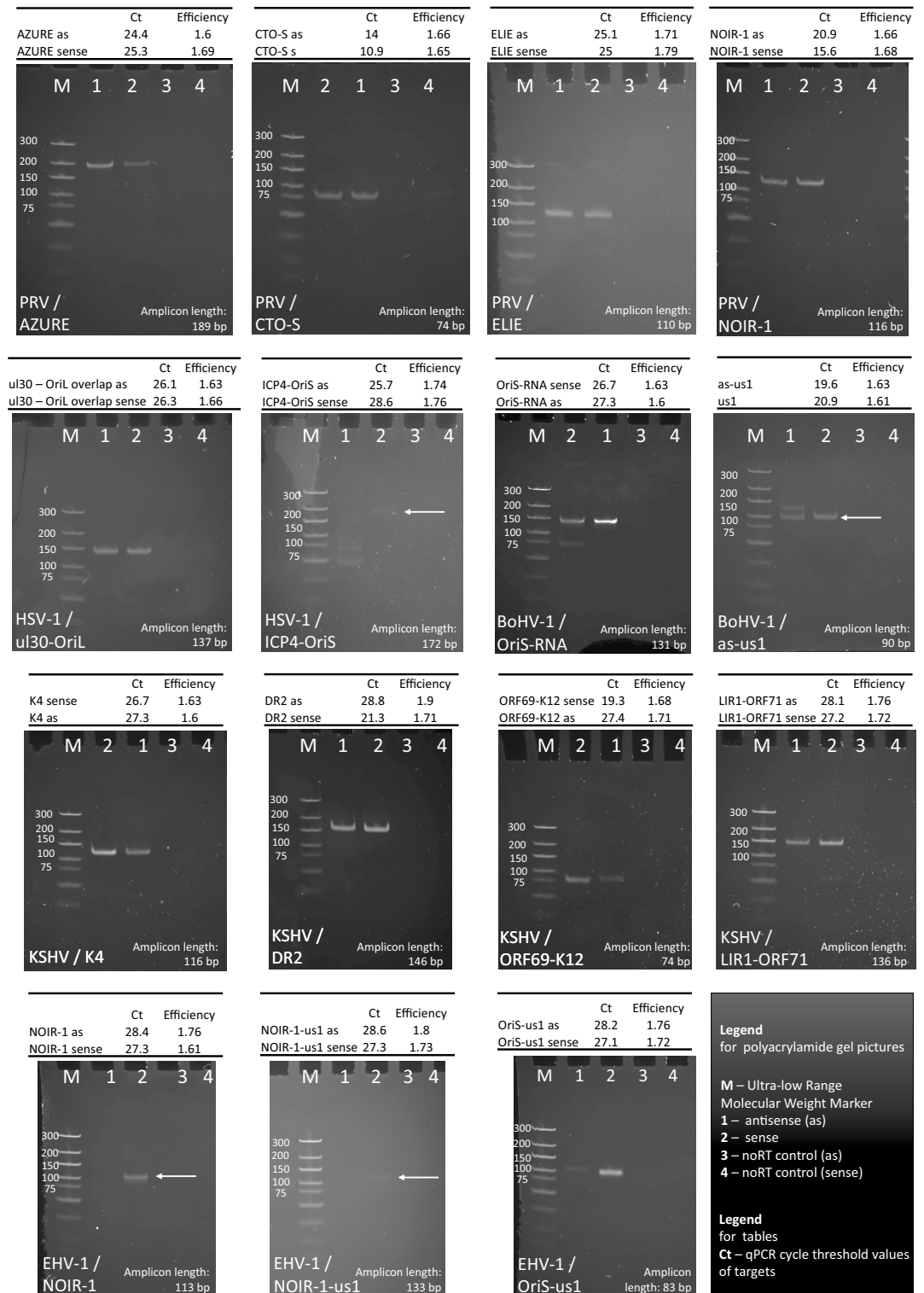
The advent of LRS technologies has made it possible to identify and precisely annotate transcripts and RNA isoforms, including length and splice variants. Technologies such as Single Molecule, Real-Time (SMRT; PacBio) sequencing<sup>77,78</sup>, nanopore sequencing (ONT)<sup>17</sup>, and LoopSeq synthetic LRS<sup>53</sup> (operating on Illumina platform; Loop Genomics) have been employed alone, or in conjunction with one other or with SRS to annotate viral transcripts in all three subfamilies of herpesviruses (HSV-1<sup>79,80</sup>; VZV<sup>75,81</sup>; PRV<sup>16</sup>; BoHV-1<sup>82</sup>; HCMV<sup>83</sup>; and EBV<sup>77,84</sup>).

In recent studies, we and others have demonstrated the existence of a previously hidden, highly complex, genome-wide network of transcriptional overlaps in different viral families<sup>23,75,77,80,82</sup>. It has been shown that the RNA molecules encoded by closely spaced genes overlap each other in a parallel, divergent, or convergent manner. This phenomenon implies an interaction between the transcription machineries at the TOs throughout the entire viral genome<sup>85</sup>. We and others have previously demonstrated that in several viruses, the replication origins overlap with specific lncRNAs and with long 5' or 3' UTR isoforms of mRNAs<sup>86</sup>. We would like to point out that numerous 5' UTR isoforms might be non-coding, as there is a considerable distance between their TSSs and the start codons. Exceptions may be those transcripts whose large parts of the 5' UTR are spliced out. Functional analyses have revealed the mechanistic details of how replication RNAs control the onset of DNA synthesis through the formation of RNA:DNA hybrids in several viruses<sup>43</sup>.

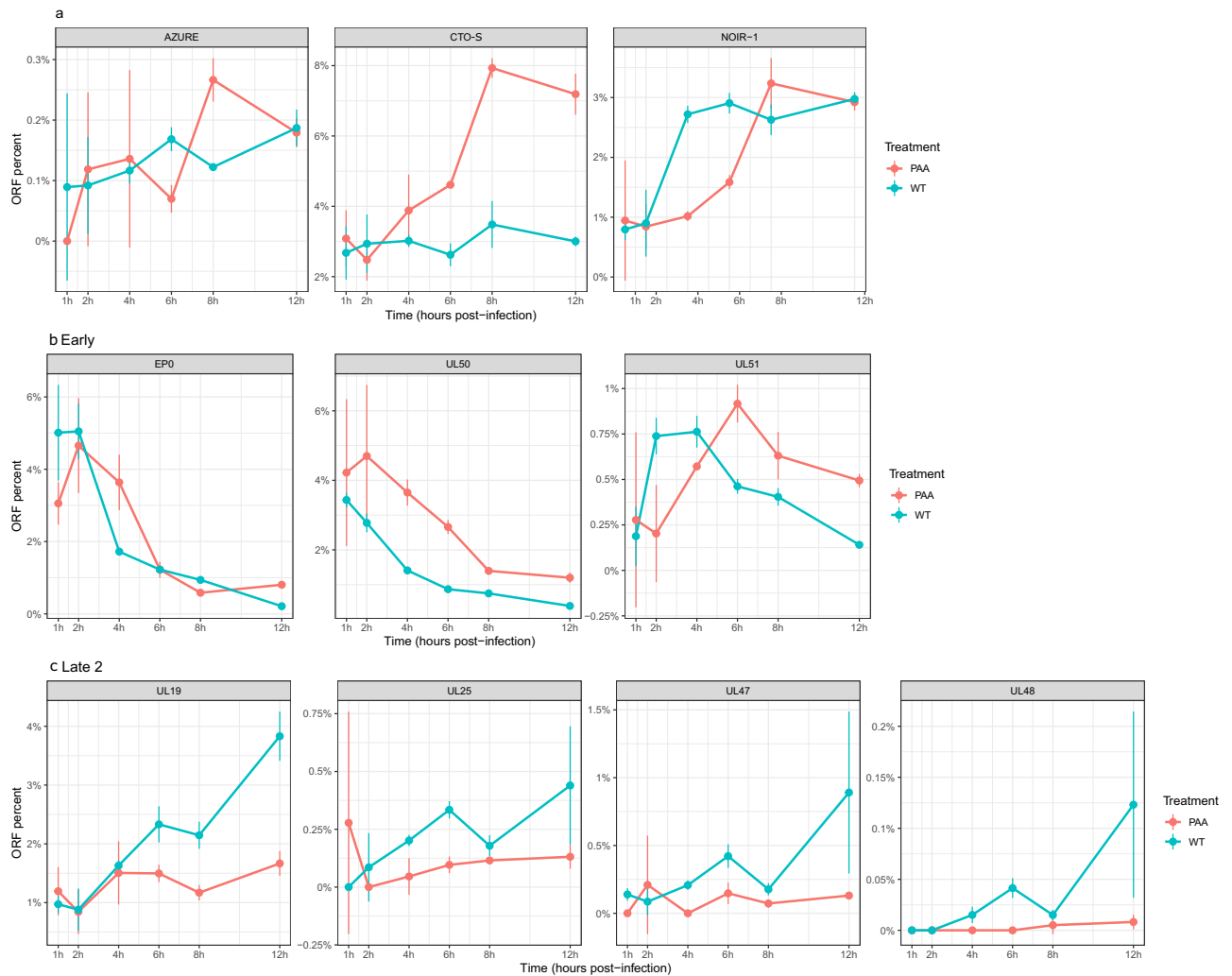
In this work, we present a state-of-the-art annotation of Ori-adjacent transcripts that potentially play key roles in regulating the replication and/or genome-wide transcription of herpesviruses. We put more emphasis on the examination of  $\alpha$ HVs because less information is available for such transcripts in these viruses. Using novel and previously published sequencing data, we discovered new transcripts and corrected the earlier annotations of already described RNAs. Furthermore, we identified an intricate meshwork of TOs between transcripts encoded by replication and transcription regulatory genes and also the specific lncRNAs around them. Promoter consensus elements within the Ori were also detected in all examined herpesviruses. Although herpesvirus OriS contain AT-rich sequences, which can be misidentified as TATA boxes, we detected the corresponding TSSs being proximal to these sequences in all cases. The terminology of the reported lncRNAs is somewhat arbitrary because their sequences and exact locations are in many cases poorly conserved. Therefore, it is possible that some of these lncRNAs with the same name have a polyphyletic origin. Nevertheless, the CTO and NOIR-1 transcript families are apparently orthologous in PRV and EHV-1.

The co-temporal activity of DNA replication and transcription within the same genomic region generates interference between the two processes along the entire genome<sup>87</sup>. Evidence suggests that the collisions have more dramatic consequences when they occur in convergent orientation rather than co-directionally<sup>88</sup>. Several molecular mechanisms have evolved to minimize the conflict between the RNP and the replication fork<sup>89</sup>. However, our current study suggests the existence of an in-built, intrinsic regulatory system based on an interaction between the two apparatuses for controlling the initiation of both replication and transcription cooperatively. This interaction is thought to be mediated through the clash and competition between the transcribing RNP and the replisome, as well as the assembly of pre-replication and transcription initiation complexes by the ongoing DNA and RNA syntheses (Fig. 8 and Supplementary Figs. 8 and 9).

In  $\alpha$ HVs, the position of OriS is highly conserved within the genome: it is located in the US repeats and surrounded by the major TF genes, *icp4* and *us1*. Here, we demonstrate that TF genes produce transcript isoforms that overlap not only the OriS but also the TF transcripts of at the other side of the replication origin, and/or specific proximal lncRNAs. The probable function of the TOs is to facilitate additional forms of interplay between these genes (asides from the TF/promoter interaction), which may occur through RNA:DNA and possibly by RNA:RNA hybridization, as well as interference between the transcription apparatuses. As a result, the US-IR



**Figure 6.** Validation of Ori-adjacent transcripts using qRT-PCR. This figure presents the validation of key Ori-adjacent transcripts using qRT-PCR and gel electrophoresis. The virus and the transcript names are indicated at the appropriate panels. We examined the transcriptional activity of both DNA strands, and used no-RT controls for each transcript. The lanes in every panel are as follows. M: molecular weight marker; 1: antisense transcripts; 2: sense transcripts (mRNA, or the canonical lncRNA); 3: no-RT for the antisense transcripts; 4: no-RT for the sense transcripts. We also indicated the amplicon lengths the Ct and efficiency values, which allow transcript quantity estimation.



**Figure 7.** Expression kinetics of PRV transcripts in untreated and in PAA-treated samples. **(a)** Three lncRNAs: AZURE, CTO-S, NOIR-1. **(b)** Three early transcripts: EP0, UL30, UL50, UL51. **(c)** Four late (L2) transcripts: UL19, UL25, UL47, UL48. The ratios of every time points were calculated by normalizing the read count of a specific transcript against the read counts of total viral transcripts in both the untreated and PAA-treated samples. The influence of PAA-treatment was most pronounced on the L2 genes, as their expression is contingent upon DNA replication, a process inhibited by PAA. CTO-S clearly displayed early expression kinetics, whereas other two lncRNAs demonstrated complex expression dynamics.

region of  $\alpha$ HVs appears to function as a ‘super regulatory center’, where the initiation of both DNA replication and global transcription is collectively controlled via a mutually interacting multilevel system. In addition, this genomic segment governs the transition between the lytic cycle and latency, as well as the maintenance of these processes. Hence, this region is functionally the most complex genomic locus of  $\alpha$ HVs, encoding lncRNAs such as intergenic transcripts and asRNAs, long TIs of mRNAs, and an intricate TO pattern of local transcripts. Besides the lytic transcripts, several latent lncRNAs (LAT, LLT, L/ST) and miRNAs are also expressed from this genomic region<sup>30</sup>.

The NOIR-1 transcript family, mapped to the US-IR region, is the evolutionary innovation of the Varicellovirus genus. The precise location of these lncRNA molecules exhibits significant variation. In PRV, the long TIs of NOIR-1 overlap the *icp4*, but no OriS overlapping NOIR-1 TES isoforms have been detected in this virus. A possible explanation for this is that in PRV this transcript is opposed to NOIR-2 whose expression might prevent transcriptional read-through by NOIR-1 toward the OriS. The PRV NOIR-1 is 3' co-terminal with the LLT transcripts expressed during latency. In EHV-1, the NOIR-1 long TI overlaps with *icp4*, while the long TSS isoform of US1 transcript, driven by the NOIR-1 promoter, overlaps the OriS. In VZV and SVV, the NOIR-1 promoters also control the transcription of the long TSS variants of US1. Additionally, in VZV, the upstream NOIR-1 transcripts also overlap the *icp4*. Intriguingly, in SVV, the NOIR-1 terminates at the OriS, which appears to be a unique solution for the interplay between the two machineries.

BoHV-1 lacks NOIR-1, but has a long TSS isoform of US1 transcripts that overlaps the OriS. Additionally, BoHV-1 expresses OriS-RNA, which has an orientation parallel to *icp4*. In HSV-1, the OriS-RNA1 has an opposite polarity to NOIR-1 and is situated on the other side of the OriS. The promoter of this transcript also controls

Sample			Ct value	Average Ct value	$\Delta$ Ct	$\Delta\Delta$ Ct	Expression fold change	
28S rRNA	Untreated	1	13	12.53	0.87			
		2	12.4					
		3	12.2					
	PAA	1	11.7	11.67				
		2	11.7					
		3	11.6					
ncRNAs	AZURE	Untreated	1	25.1	24.97	-0.63	-1.50	2.83
			2	25				
			3	24.8				
		PAA	1	25.4	25.6			
			2	26.2				
			3	25.2				
	CTO-S	Untreated	1	12.5	12.67	-2.47	-3.33	10.08
			2	12.9				
			3	12.6				
		PAA	1	15.4	15.13			
			2	15.2				
			3	14.8				
	NOIR-1	Untreated	1	18.3	17.77	-3.00	-3.87	14.59
			2	18.2				
			3	16.8				
		PAA	1	21.1	20.77			
			2	20.6				
			3	20.6				
Early genes	ep0	Untreated	1	19.2	18.97	-0.97	-1.83	3.56
			2	19.3				
			3	18.4				
		PAA	1	19.8	19.93			
			2	20.2				
			3	19.8				
	ul50	Untreated	1	20.7	20.53	-0.53	-1.40	2.64
			2	20.8				
			3	20.1				
		PAA	1	21.2	21.07			
			2	21.2				
			3	20.8				
	ul51	Untreated	1	25,3	25,3	-0.73	-1.60	3.03
			2	25,3				
			3	25,3				
		PAA	1	25.9	26.03			
			2	26.1				
			3	26.1				
Continued								



Sample			Ct value	Average Ct value	$\Delta$ Ct	$\Delta\Delta$ Ct	Expression fold change		
Late-2 genes	ul19	Untreated	1	15.3	14.97	-2.97	-3.83	14.25	
			2	15.1					
			3	14.5					
		PAA	1	18					17.93
			2	18.2					
			3	17.6					
	ul25	Untreated	1	17.3	17.03				
			2	17.3					
			3	16.5					
		PAA	1	19.8					19.67
			2	19.9					
			3	19.3					
	ul47	Untreated	1	21.8	21.67				
			2	22.2					
			3	21					
		PAA	1	25.3					25.23
			2	25.2					
			3	25.2					
	ul48	Untreated	1	16	15.87				
			2	15.9					
			3	15.7					
		PAA	1	20.6					20.53
			2	20.7					
			3	20.3					

**Table 2.** qRT-PCR and dcDNA-Seq analysis of PRV transcripts in PAA-treated and untreated samples.

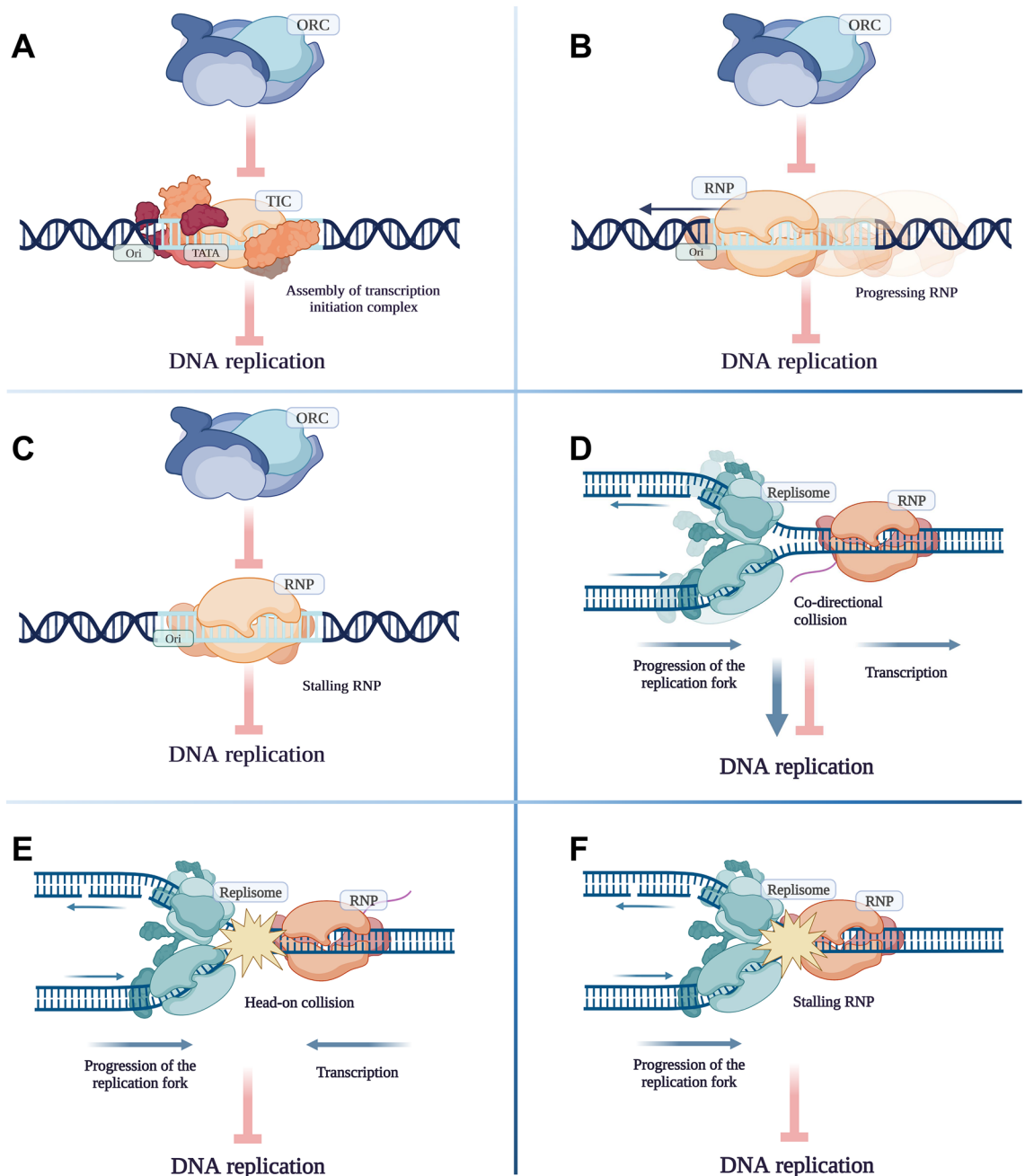
the long TSS isoform of the ICP4 transcript. NOIR-1 transcripts may be involved in controlling transcription through overlapping the *icp4* gene, and/or in controlling DNA replication through overlapping the OriS (in cases when they are the upstream part of the US1 TI), or in controlling both transcription and replication (in cases when they overlap both OriS and *icp4*). The same might also be true for the HSV-1 OriS-RNA1, as it also serves as the 5' UTR part of the long ICP4 TI.

It may not be a coincidence that the OriL of Simplexviruses is surrounded by the main replication genes (*ul29/30*). We hypothesize that these genes might interact with initiation process of replication not only through conventional TF/promoter binding but also at other levels, including the RNA:DNA and/or RNA:RNA hybridization, and through interference between transcription and replication machineries at the overlapping region. The CTO transcripts have been previously identified in PRV<sup>51</sup>. Here, we report the detection of the orthologous transcripts in EHV-1. The canonical CTO-S does not overlap the OriL, but it might interfere with replication by helping to separate the two DNA strands and thereby determining the orientation of replication, or in other ways. One of the long CTO-S TIs is controlled by a promoter within the OriL, and another one is the TES variant of *ul21* gene.

We also observed a potentially intriguing phenomenon: in  $\alpha$ HVs, a total of three 'hard' convergent TOs are formed, which may represent a strong confrontation between transcription apparatuses during the synthesis of RNA molecules. We noticed that one of the partners forming these TOs is always an auxiliary replication gene. The involvement of such genes in every 'hard' TOs may not be just a mere coincidence but have an unknown biological relevance. We can speculate that the 'hard' TO between the *ul30/ul31* partner genes of Simplexviruses allows *ul31* gene to generate an OriL-overlapping long TES isoform through transcriptional read-through. Indeed, we detected a high-frequency transcriptional read-through from this gene using RT<sup>2</sup>-PCR in PRV. Interestingly, although in PRV and EHV-1 the OriL is located between the *ul21* and *ul22* genes, the OriL is also overlapped by a long TES isoform of *ul21*. Since no long TES isoforms of *ul21* in Simplexviruses and of *ul31* in Varicelloviruses are produced, it is reasonable to assume that the role of these types of TIs is to interfere with the initiation of DNA replication.

The HCMV OriLyt is overlapped by lncRNAs, among which RNA4.9 has been shown to regulate DNA replication<sup>43</sup>. We have described novel Ori-overlapping lncRNAs of HCMV previously<sup>36</sup> and in this recent report. The OriLyt-L of EBV is adjacent to the non-coding BHLF1 gene, a latency regulator<sup>91</sup>, while the OriLyt-R is adjacent to LF2, an inhibitor of replication<sup>92</sup>. KSHV LANA plays an essential role as a latency regulator<sup>93</sup>. In this study, we describe additional Ori-overlapping transcripts, including lncRNAs and long 5' UTR isoforms of mono- and polycistronic mRNAs.

We quantified the amount of particularly long 5'UTR variants using qRT-PCR and found that their low abundance cannot be explained solely by the size-bias of the sequencing; these transcripts are indeed expressed at a low level. Although many of them carry complete ORFs, they are probably untranslated since their transcription



**Figure 8.** Potential effects of transcription on the DNA replication. (A) Recruitment of transcription initiation complex on promoters located within the replication origin inhibits assembly of ORC and replisome. (B) RNA polymerase II passage across the replication origin inhibits ORC and replisome assembly. (C) RNA polymerase II stalling on the replication origin inhibits ORC and replisome assembly. (D) Co-directional movement of replication and transcription machineries can slow down or speed up replication fork progression. Transcription can facilitate DNA replication by pre-opening the two DNA strands. (E) Head-on collision of replication and transcription machineries inhibits both processes. (F) RNA polymerase II stalling inhibits replication fork progression.

and translation start sites are far apart. Thus, they might be transcriptional by-products without functioning as protein-coding sequences, or even as RNA molecules.

In summary, even closely related herpesvirus species have developed unique strategies for establishing transcriptional overlaps (TOs) at the Oris and the TR genes. This underscores the significance of this phenomenon in regulating DNA replication and overall transcription. While the major TR genes appear to control each other and the initiation of DNA replication at the OriS region of  $\alpha$ HVs, the main replication genes at the OriL region of Simplexviruses, might rather regulate each other and DNA replication through mechanisms achieved via TOs. These putative mechanisms provide multiple regulatory layers in addition to conventional transcription factor/promoter interaction. ICP4 has been shown to promote the expression of *icp0*<sup>94</sup> genes by binding to its promoter.

ICP22 (*us1* product) inhibits the expression of both *icp4* and *icp0*<sup>95</sup>. Furthermore, ICP0 converts ICP4 from a repressor to an activator of mRNA synthesis in HSV-1<sup>96</sup>. The mutation of *us1* gene leads to a differential effect on the transcription kinetics of E and L genes<sup>97</sup>.

We believe that the implications of our results have a much broader perspective and represent a general way of how the regulation of herpesviral replication and transcription co-evolved with each other. Understanding these complex interactions between various genes and regulatory elements may provide valuable insights into the overall regulatory mechanisms governing herpesvirus replication and gene expression. Further investigation into these potential interactions and their functional significance may open the door for the creation of innovative therapeutic approaches to address herpesvirus infections.

## Methods

### Cells and viruses

In addition to our novel data, we also used several other datasets for the analyses in this study. The cell types used in this work are listed in Supplementary Table 4.

**PRV** For the generation of novel transcript data, we employed three immortalized cell lines to propagate the MdBio strain of pseudorabies virus (PRV-MdBio<sup>64</sup>): PK-15 porcine kidney epithelial cell line (ATCC® CCL-33™), C6 cell line derived from a rat glial tumor (ATCC® CCL-107™), and PC-12 (ATCC® CRL-1721™) derived from a pheochromocytoma of the rat adrenal medulla, which have an embryonic origin from the neural crest. Each experiment was conducted in three biological replicates. PK-15 cells were cultured in Dulbecco's modified Eagle medium (DMEM) (Gibco/Thermo Fisher Scientific), supplemented with 5% fetal bovine serum (FBS; Gibco/Thermo Fisher Scientific) and 80 µg of gentamycin per ml (Gibco/Thermo Fisher Scientific) at 37 °C in the presence of 5% CO<sub>2</sub>. C6 cells were cultivated in F-12 K medium (ATCC), complemented with 2.5% FBS and 15% horse serum (HS; Sigma-Aldrich) at 37 °C in the presence of 5% CO<sub>2</sub>. PC-12 cells were maintained in RPMI-1640 medium (ATCC) supplemented with 5% FBS and 10% HS at 37 °C in the presence of 5% CO<sub>2</sub>. Virus stock solution was prepared as follows: PK-15 cells were infected with 0.1 multiplicity of infection [MOI = plaque-forming units (pfu)/cell]. Viral infection was allowed to progress until complete cytopathic effect was observed, which was followed by three successive cycles of freezing and thawing of infected cells in order to release of viruses from the cells. Each cell type was infected with MOI = 1 of PRV-MdBio. Infected cells were incubated for 1 h at 37 °C followed by removal of the virus suspension and washing the cells with phosphate-buffered saline (PBS). Following the addition of new culture medium, the cells were incubated for 0, 1, 2, 4, 6, 8, or 12 h. Following the incubation, the culture medium was discarded and the infected cells were frozen at -80 °C until further use.

For the kinetic analysis, we used MOI = 1 of PRV-Ka for the infection of PK-15 cells. Cells were first incubated at 4 °C for 1 h for synchronization of infection, and then placed in a 5% CO<sub>2</sub> incubator at 37 °C. Infected cells were collected at every 30 min within an 0–8 h interval. Cells were washed with PBS, scraped from the culture plate, and centrifuged at 3000 RPM for 5 min at 4 °C.

**EHV-1** Equid alphaherpesvirus 1 was also employed in this study. A field isolate of the virus (EHV-1-MdBio) was used, which is originated from Marócpuszta (Hungary) from the organs of an aborted colt fetus in the 1980s. A confluent rabbit kidney (RK-13) epithelial cell line (ECACC 0,021,715) was used for viral propagation. The experiments were carried out in three technical replicates.

RK-13 cells were maintained in DMEM (Sigma). The media was supplemented with 10% fetal calf serum (FCS, Gibco). The culture conditions were as follows: 37 °C in the presence of 5% CO<sub>2</sub>. A virus stock solution was prepared by infecting the cells with EHV-1-MdBio at MOI = 0.1. Three freeze–thaw cycles were applied to release the viruses from the cells. For the EHV-1 long-read RNA-seq investigations, RK-13 cells were infected with MOI = 4 of the virus. Three technical replicates were carried out. Viral infected RK-13 cells were incubated at 4 °C for 1 h, then the virus suspension was removed and cells were washed with PBS. Next, new media was added to the cells, which were incubated for 1, 2, 4, 6, 8, 12, 18, 24 or 48 h. Finally, the culture medium was removed from the cells and they were stored at -80 °C until further use.

**KSHV**: The KSHV-positive primary effusion lymphoma cell line iBCBL1-3xFLAG-RTA<sup>98</sup> was maintained in RPMI 1640 medium, which was supplemented with 10% Tet System Approved FBS (TaKaRa), penicillin/streptomycin, and 20 µg/ml hygromycin B. KSHV lytic reactivation was induced by treating 1 million of iBCBL1-3xFLAG-RTA cells with 1 µg/ml doxycycline for 24 h. For measuring KSHV gene expression by qRT-PCR, cDNA was generated with iScript cDNA Synthesis kit (Bio-Rad) followed by SYBR green-based real-time quantitative PCR analysis using gene specific primers. The relative viral gene expression was calculated by the delta-delta Ct method where 18S was used for normalization. The sequences of the primers have been reported previously (Toth et al., 2013). The following antibodies were used for immunoblots: anti-FLAG (Sigma F1804), anti-ORF6 (from Dr. Gary S. Hayward, Johns Hopkins University), and anti-Tubulin (Sigma T5326).

**VZV**: Human primary embryonic lung fibroblast cell line (MRC-5) obtained from the American Type Culture Collection (ATCC) was used for the propagation of the live attenuated OKA/Merck strain of varicella zoster virus. Cells were grown in DMEM supplemented with antibiotic/antimycotic solution and 10% fetal bovine serum (FBS) at 37 °C in a 5% CO<sub>2</sub> atmosphere. Infected cells were harvested by trypsinization when they displayed cytopathic effect (after 5 days).

**HSV-1**: An immortalized kidney epithelial cell line (Vero) was used for the propagation of HSV-1. The cell culture was grown in DMEM (Gibco/Thermo Fisher Scientific) with 10% Fetal Bovine Serum (Gibco/Thermo Fisher Scientific) and 100 µl/ml penicillin–streptomycin Mixture (Lonza), and in a 37 °C incubator with a humidified atmosphere of 5% CO<sub>2</sub> in air. Cells were infected with HSV-1 at an MOI = 1, then they were incubated for 1 h. The virus suspension was removed and cells were washed with PBS. This step was followed by the addition of fresh medium to the cells and they were incubated for 1, 2, 4, 6, 8, or 12 h. A mixture from the different time points were prepared for the further analysis.

**BoHV-1:** Madin–Darby Bovine Kidney (MDBK) cells were used for the infection using the Cooper isolate (GenBank Accession # JX898220.1) of Bovine alphaherpesvirus 1.1. Cells were incubated at 37 °C in a humidified incubator with 5% CO<sub>2</sub> and were cultured in DMEM supplemented with 5% (v/v) fetal bovine serum, 100 U/mL penicillin, and 100 µg/mL streptomycin. Cells were infected using MOI = 1 virus suspension. Infection was allowed to progress until complete cytopathic effect was observed. Infected cells were incubated for 1, 2, 4, 6, 8, and 12 h, then the supernatant was collected, and the cellular fraction was subjected to three successive cycles of freezing and thawing in order to release additional intracellular virions.

### Inhibition of DNA synthesis using phosphonoacetic acid

Before the infection, we treated the cells with 400 µg/ml PAA (phosphonoacetic acid) for 1 h at 37 °C in the presence of 5% CO<sub>2</sub> to inhibit replication and determine the kinetic class of transcripts. During the experiment, we examined the cells at six different time points (1, 2, 4, 6, 8, 12 h) using three biological replicates. For the treated samples, we also used an MOI of 10 of the virus for infection. RNA was isolated at 1, 2, 4, 6, 8, 12 h post-infection time points, and a dcDNA library was prepared according to the library preparation protocol of the direct cDNA Sequencing Kit (SQK-DCS109). We sequenced the samples on Oxford Nanopore MinION flow cells. Subsequently, the data were basecalled by Guppy, mapped using minimap2 software, and further analyzed using scripts deposited on GitHub.

([https://github.com/Balays/Rlyeh?fbclid=IwAR0HZJNXZjv9YUm\\_tsJ5J1eT2fKXnkhbJKf7WV0TxX9kvp7fjmdhWQILbjA](https://github.com/Balays/Rlyeh?fbclid=IwAR0HZJNXZjv9YUm_tsJ5J1eT2fKXnkhbJKf7WV0TxX9kvp7fjmdhWQILbjA)). Data have been deposited under the project accession number PRJEB64684 into the European Nucleotide Archive (ENA).

### RNA isolation

**Extraction of total RNA** Total RNA was isolated from the PRV, EHV-1, VZV and KSHV infected cells by using the NucleoSpin® RNA kit (Macherey–Nagel). The spin-column protocol was applied. In brief, cells were lysed by the addition of a chaotropic ion containing buffer solution (from the kit). Nucleic acids were then bound to a silica membrane. Samples were treated with DNase I to remove genomic DNA. Total RNAs were eluted with RNase-free water. To eliminate the potential remaining DNA from the samples, we used the TURBO DNA-free™ Kit. Samples were stored at -80 °C.

**Purification of polyadenylated RNA** The Qiagen Oligotex mRNA Mini Kit was used to enrich the mRNAs (and other RNAs with polyA-tail) from the PRV, EHV-1 and VZV samples, which were then used as templates for ONT and Illumina library preparations. The Spin Columns protocol of the manual was applied. In brief, the final volume of the RNA samples was set to 250 µL by adding RNase-free water. Then, 15 µL Oligotex suspension and 250 µL OBB buffer (both from the Oligotex kit) were added to the mixtures, which were first incubated at 70 °C for 3 min and then at 25 °C for 10 min. Samples were centrifuged at 14,000 × g for 2 min, and the supernatants were removed. Four-hundred µL Oligotex OW2 wash buffer was added to the samples, then they were spun down in Oligotex spin columns at 14,000 × g for 1 min. This step was repeated once, and finally, the poly(A) + RNA fraction was eluted from the membrane by adding 60 µL hot Oligotex elution buffer. To maximize the yield, a second elution step was also carried out. The Lexogen Poly(A) RNA Selection Kit V1.5 was used for the selection of polyadenylated RNAs from the KSHV total RNA samples. Briefly, 10 µL of total RNA (5 µg) was denatured at 60 °C for 1 min then it was held at 25 °C. The RNA samples were mixed with 10 µL washed bead (including in the kit) and the mixtures were incubated at 25 °C for 20 min with 1,250 rpm agitation. Afterwards, the tubes were placed in a magnetic rack for 5 min. Supernatant was discarded and the beads were resuspended in Bead Wash Buffer (supplied by the Lexogen kit) and were incubated 25 °C for 5 min with 1250 rpm agitation. Tubes were transferred onto the magnetic rack and supernatant was discarded after 5 min incubation. This washing step was carried out twice. After the second washing step, the beads were resuspended in Nuclease-free water (Lexogen kit) and then incubated at 70 °C for 1 min. Then, the tubes were transferred onto the magnet for 5 min. The supernatant containing the poly(A) + RNA fraction was transferred into a fresh tube.

**Removal of rRNA** Ribo-Zero Magnetic Kit H/M/R (Epicentre/Illumina) was used to remove ribosomal RNAs and to enrich mRNAs. Unlike poly(A) + purification, the rRNA depletion process also retains RNAs without polyA tails, except for rRNAs. For starting material, 5 µg of a mixture of EHV-1 total RNA was used. The sample was mixed with the Ribo-Zero Reaction Buffer and Ribo-Zero rRNA Removal Solution. The mixture was incubated at 68 °C for 10 min, then at room temperature for 5 min. Next, 225 µL washed Magnetic Bead was added to the sample and they were incubated at room temperature for 5 min, then at 50 °C for 5 min. Finally, the mixture was placed on a magnetic stand, then the supernatant containing the rRNA-depleted RNA was collected. The final purification was carried out by using the Agencourt RNAClean XP Beads (Beckman Coulter) as recommended by the Ribo-Zero manual.

**Enrichment of the 5' ends of RNAs** Terminator™ 5'-Phosphate-Dependent Exonuclease (Lucigen) was used to enrich the 5' ends of the transcripts. The process was carried out with a mixture of poly(A) + RNAs from the EHV-1 samples, which was mixed with Terminator 10X Reaction Buffer A, RiboGuard RNase Inhibitor and Terminator Exonuclease (1 Unit). The mixture was incubated at 30 °C for 60 min, then the reaction was stopped by the addition of 1 µL of 100 mM EDTA (pH 8.0). RNAClean XP beads (Beckman Coulter) was used for final purification.

### Measurement of nucleic acid quality and quantity

**RNA:** The Qubit 4. fluorometer and the Qubit Assay Kits (Invitrogen) were used for the measurement of RNA concentration. The RNA BR Assay was applied for the quantitation of total RNA samples whereas the RNA HS Assay was utilized for the poly(A) + and ribodepleted RNA fractions. The quality of the total RNA samples was checked by using the Agilent TapeStation 4150 device, RNA ScreenTape and reagents were used. RIN scores



above 9.6 were used for cDNA production. The RNA quality was assessed with the Agilent 2100 Bioanalyzer (for PacBio sequencing) or Agilent 4150 TapeStation System (for MinION sequencing) and RIN scores above 9.6 were used for cDNA production.

**cDNA:** The Qubit dsDNA HS Assay Kit (Invitrogen) was used to quantify the cDNA samples. For the analysis of Illumina library quality, the Agilent High Sensitivity D1000 ScreenTape was used.

### Direct cDNA sequencing

Libraries were created without PCR amplification from the poly(A) + fractions of RNAs from PRV, EHV-1 and KSHV as well as from the Terminator-treated EHV-1 samples. To achieve this, the ONT's Direct cDNA Sequencing Kit (SQK-DCS109) was utilized according to the ONT's manual. Briefly, the RNAs were combined with ONT VN primer and 10 mM dNTPs, and incubated at 65 °C for 5 min. Subsequently, RNaseOUT (Thermo Fisher Scientific), 5 × RT Buffer (Thermo Fisher Scientific), ONT Strand-Switching Primer were added to the mixtures, which were then incubated at 42 °C for 2 min. Maxima H Minus Reverse Transcriptase enzyme (Thermo Fisher Scientific) was added to the samples to produce the first cDNA strands. The reaction was conducted at 42 °C for 90 min, and the reactions were halted by heating the samples at 85 °C for 5 min. The RNAs from the RNA:cDNA hybrids were removed using RNase Cocktail Enzyme Mix (Thermo Fisher Scientific). This reaction took place at 37 °C for 10 min. The LongAmp Taq Master Mix [New England Biolabs (NEB)] and ONT PR2 Primer were used to synthesize the second strand of cDNAs. The following PCR condition was applied: 1 min at 94 °C, 1 min 50 °C, then 15 min at 65 °C. As a next step, the end-repair and dA-tailing were carried out with the NEBNext End repair /dA-tailing Module (NEB) reagents. These reactions were conducted at 20 °C for 5 min, followed by heating the samples at 65 °C for 5 min. Adapter ligation was then carried out with the NEB Blunt /TA Ligase Master Mix (NEB) at room temperature for 10 min. The ONT Native Barcoding (12) Kit was utilized to label the libraries, then they were loaded to ONT R9.4.1 SpotON Flow Cells (200 fmol mixture of libraries was loaded to one flow cell. AMPure XP Beads were used after each enzymatic step, and samples were eluted in UltraPure™ nuclease-free water (Invitrogen).

### Amplified Nanopore cDNA sequencing

For more accurate mapping of the 5'-end of VZV transcripts, random-primer based RT reactions were carried out as a first step of library preparation. For this, poly(A)-selected and Terminator-treated RNA samples and SuperScript IV enzyme (Life Technologies) were used. From the first-strand cDNAs, libraries were prepared according to the modified 1D strand switching cDNA by ligation protocol Ligation Sequencing kit (SQK-LSK108; Oxford Nanopore Technologies). KAPA HiFi DNA Polymerase (Kapa Biosystems) and Ligation Sequencing Kit Primer Mix (part of the 1D Kit) were used to amplify the cDNAs. Next, samples were end-repaired using NEBNext End repair /dA-tailing Module (New England Biolabs), then adapter ligation was carried out using the sequencing adapters supplied with the kit and NEB Blunt/TA Ligase Master Mix (New England Biolabs).

### Short-read sequencing

For the SRS transcriptomic analysis of EHV-1 virus, libraries were generated from a mixture of poly(A) + enriched and rRNA-depleted samples using the NEXTflex® Rapid Directional qRNA-Seq Kit (PerkinElmer). Fragmentation of the RNAs was carried out enzymatically at 95 °C for 10 min via the addition of NEXTflex® RNA Fragmentation Buffer. Next, the first strand cDNA was synthesized. First, the NEXTflex® First Strand Synthesis Primer was mixed with the RNA sample, which were incubated at 65 °C for 5 min, then the mixture was subsequently placed on ice. RT reaction was performed by using the NEXTflex® Directional First Strand Synthesis Buffer and Rapid Reverse Transcriptase according the following protocol: incubation at 25 °C for 10 min, then at 50 °C for 50 min, and termination at 72 °C for 15 min. Generation of the second cDNA strand was carried out at 16 °C for 60 min with the addition of NEXTflex® Directional Second Strand Synthesis Mix (with dUTPs). Adenylation of cDNAs were performed at 37 °C for 30 min using the NEXTflex® Adenylation Mix. The reaction was terminated by incubating the samples at 70 °C for 5 min. As a next step, Molecular Index Adapters (part of the Kit) were ligated to the sample using the NEXTflex® Ligation Mix. Ligation was carried out at 30 °C for 10 min. Samples were amplified by PCR. As a first step of this, the NEXTflex® Uracil DNA Glycosylase was added to the samples, and then they were incubated at 37 °C for 30 min, which was followed by heating at 98 °C for 2 min, and they were subsequently placed on ice. Then, the following components were added to the samples: PCR Master Mix, qRNA-Seq Universal forward primer, and qRNA-Seq Barcoded Primer (sequence: AACGCCAT; all from the PerkinElmer kit). The following settings were applied for the PCR: 2 min at 98 °C, followed by 15 cycles of 98 °C for 30 s, 65 °C for 30 s and 72 °C for 60 s, followed by a final extension at 72 °C for 4 min. AMPure XP Bead was used after each enzymatic reaction. The NEXTflex® Resuspension buffer was used for the final elution of the sequencing library from which 10 pM was loaded to the Illumina MiSeq reagent cassette. Paired-end transcriptome sequencing was performed on an Illumina MiSeq sequencer by using the MiSeq Reagent Kit v2 (300 cycles).

### Direct RNA sequencing

Native RNA sequencing approach was also applied for library preparation from the EHV-1 and KSHV samples with the aim to avoid potential biases associated with reverse transcription (RT) and PCR reactions to detect and validate novel splice variants as well as 3' UTR isoforms. For the dRNA-seq experiments, we used an RNA mixture containing RNA from each time points of infection from the Poly(A) + samples and from the Poly(A) + and Terminator-treated RNAs. The oligo dT-containing T10 adapter for RT priming and the RNA CS for monitoring the sequencing quality (both from the ONT kit) were added to the RNA mix along with NEBNext Quick Ligation Reaction Buffer, and T4 DNA ligase (both from NEB). The reaction was incubated for 10 min at room temperature. Then, 5 × first-strand buffer, DTT (both from Invitrogen), dNTPs (NEB) and UltraPure™ DNase/RNase-Free

water (Invitrogen) were added to the samples. Finally, SuperScript III enzyme (Thermo Fisher Scientific) was mixed with the sample and RT reaction was performed at 50 °C for 50 min, and subsequently stopped by heat inactivation of the enzyme at 70 °C for 10 min. RNA adapter (from the ONT kit) was ligated to the RNA:cDNA hybrid sample using the NEBNext Quick Ligation Reaction Buffer and T4 DNA ligase at room temperature for 10 min. The RNAClean XP Beads were used after each additional enzymatic step. Two flow cells were used for dRNA-seq, 100 fmol from the sample was loaded onto each of them.

### Real-time RT-PCR

Quantitative qRT-PCR was used for transcript validation and kinetic analysis, and was carried out as previously described<sup>99</sup> using the primers listed in Supplementary Table 5. Briefly, first strand cDNAs from the total RNAs of PRV, EHV-1, VZV, HSV-1, BoHV-1 and KSHV were synthesized by using the SuperScript III reverse transcriptase and gene-specific primers (ordered from Integrated DNA Technologies). The reactions were carried out in 5 µl of final volume containing 0.1 µg of total RNA, 2 pmol of the primer, dNTP mix (Invitrogen, 10 µM final concentration), 5× First-Strand Buffer, and the RT enzyme (Invitrogen). The reaction was performed at 55 °C for 60 min and then terminated by heating the samples to 70 °C for 15 min. Ten-fold dilutions from the RT products were used as template for real-time PCR amplification. PCR reactions were carried out using Absolute QPCR SYBR Green Mix (Thermo Fisher Scientific) in a Rotor-Gene Q (Qiagen) cyclor. The running conditions are shown in Supplementary Table 6. The following controls were used: loading control (28S rRNA as a reference gene), no template (checking the potential primer-dimer formation) and no RT control (to detect the potential DNA contamination). The relative expression levels (R) were calculated using the mathematical framework proposed by Soong and colleagues<sup>100</sup>. However, we made some adjustments: the R values were computed using the mean ECt value from the 6 h samples for every gene as a benchmark. This was then normalized against the mean value of the corresponding 28S values.

We used the Comparative Quantitation module of the Rotor-Gene Q software package that automatically calculates the qRT-PCR efficiency sample-by-sample and set the cycling thresholds values automatically.

$$R = \frac{(E_{\text{sample6h}})^{Ct_{\text{sample6h}}}}{(E_{\text{sample}})^{Ct_{\text{sample}}}} : \frac{(E_{\text{sample6h}})^{Ct_{\text{ref6h}}}}{(E_{\text{sample6h}})^{Ct_{\text{ref}}}}$$

We used the  $2^{-\Delta\Delta Ct}$  method for the comparison of gene expression values of the PAA-treated samples with the untreated samples<sup>101</sup>.

### Cap Analysis of Gene Expression

We employed CAGE-Seq to investigate the TSS distribution patterns at the examined genomic regions in EHV-1 and KSHV using three biological replicates. The CAGE™ Preparation Kit (DNAFORM, Japan) was utilized. CAGE libraries were prepared from 5 µg of total RNA following the manufacturer's recommendations. Briefly, the RNA and the RT primer (random primer mixture from CAGE™ Prep Kit) was denatured at 65 °C for 5 min. First cDNA strands were synthesized by using SuperScript III Reverse Transcriptase (Invitrogen). To enhance the activity and specificity of RT enzyme, trehalose/sorbitol mixture (CAGE™ Prep Kit) was also added. The reactions were incubated for 30 s at 25 °C, then the RT reaction was performed at 50 °C for 60 min. Diol groups in the Cap at the 5'-end (and ribose at 3'-end) were oxidized with NaIO<sub>4</sub> and Biotin (long arm) hydrazine was bound to it. First, the oxidation was carried out with the addition of NaOAc (1 M, pH 4.5, CAGE™ Prep Kit) and NaIO<sub>4</sub> (250 mM, CAGE™ Prep Kit) to the samples and incubation on ice for 45 min in the dark. After this step, 40% glycerol and Tris-HCl (1 M, pH 8.5, CAGE™ Prep Kit) were added to the samples. NaOAc (1 M, pH 6.0) and Biotin Hydrazine (10 mM, CAGE™ Prep Kit) were mixed with the samples and the oxidized diol residues were biotinylated at 23 °C for 2 h. After this step, single strand RNA was digested by applying RNase I (CAGE™ Prep Kit) treatment (37 °C 30 min). Biotinylated, Capped RNA samples were mixed and bound (Cap-trapping) to the pretreated Streptavidin beads (pretreatment details below) at 37 °C 30 min. Then they were incubated on a magnetic rack. The beads were washed with Wash Buffer 1 (twice), then with Wash Buffer 2, and finally with Wash Buffer 3 (both from the CAGE™ Prep Kit). Next, cDNAs were released from the beads: Releasing Buffer was added to the samples and they were incubated at 95 °C for 5 min. After a short incubation on a magnetic rack, the supernatant (containing the Capped cDNAs) was transferred to new tubes. RNase I buffer (CAGE™ Prep Kit) was added to the tRNA-Streptavidin bead and they were placed on a magnetic rack. The supernatant was transferred to the tubes containing the cDNAs and they were stored on ice. The samples were treated with an RNase mixture (RNase H and RNase I, both from the CAGE™ Prep Kit) and incubated at 37 °C for 15 min. The potential remaining RNA was digested with RNase I. The reaction was performed at 37 °C for 30 min.

Streptavidin beads were coated with tRNA (CAGE™ Prep Kit), followed by mixing and incubation on ice for 30 min, and then incubated on a magnetic stand for 3 min. The supernatant was removed, and the beads were washed with Wash Buffer 1 (CAGE™ Prep Kit) twice. Finally, the beads were eluted in Wash Buffer 1 and tRNA was also added. The volume of the samples was reduced by using the miVac DUO Centrifugal Concentrator (Genevac), then single strand 5' linkers (with barcodes) were ligated to the samples at 16 °C for 16 h using the DNA ligation mixture CAGE™ Prep Kit). After a purification step, the miVac DUO was used again in order to concentrate the samples. This was followed by the ligation of the 3' linker using the DNA ligation mixture. It was carried out at 16 °C for 16 h. (The 5' and 3' linkers were pre-heated at 55 °C and the cDNA samples at 95 °C before the ligation steps). The samples were mixed with Shrimp Alkaline Phosphatase (SAP, CAGE™ Prep Kit) to remove the phosphate group of the linkers. The reaction was conducted at 37 °C for 30 min and terminated at 65 °C for 15 min. Then USER enzyme was added to the sample, which digests the dUTP from 3' linker up strand. The USER treatment was carried out at 37 °C for 30 min and it was stopped at 95 °C for 5 min. After this



step, the barcoded samples were mixed and concentrated with miVac DUO. Finally, the second cDNA strands were synthesized with the 2nd primer, DNA polymerase, buffer and 10 mM dNTP (all from the CAGE™ Prep Kit). The denaturation step was set to 95 °C for 5 min, the annealing to 55 °C for 5 min, and the elongation to 72 °C for 30 min. The sample mixture was treated with Exonuclease I enzyme (37 °C for 30 min). The vacuum concentrator was used to completely dry the sample and finally it was dispensed in 10 µl of nuclease-free H<sub>2</sub>O. The amount of single-stranded cDNAs was measured using Qubit 2.0 and the Qubit ssDNA HS Assay Kit. The RNAClean XP Beads were used after RT, oxidation and biotinylation. AmpureXP Beads were used to purify the samples after Cap-trapping and Releasing, RNase I treatment, 5' and 3' linker ligation, SAP and USER treatments, 2nd strand cDNA synthesis and Exonuclease I treatment. Libraries with different barcodes were pooled and applied on the same flow cells. The libraries were sequenced on a MiSeq instrument with v3 (150 cycles) and v2 (300 cycles) chemistries (Illumina). Qubit 4.0 and 1X dsDNA High Sensitivity (HS) Assay was used to measure the concentration of the sample. The quality of the library was tested by TapeStation.

## Bioinformatics analyses

### Illumina CAGE sequencing data analysis

Read quality was checked with **fastqc** (<https://www.bioinformatics.babraham.ac.uk/projects/fastqc>). The reads were trimmed with **TrimGalore** (<https://github.com/FelixKrueger/TrimGalore>) using the following settings: `-length 151 -q`. The **STAR** aligner<sup>102</sup>, version 2.7.3.a was used to map the reads to the KSHV strain TREx reference genome (*GQ994935.1*) using `-genomeSaiindexNbases 8`, otherwise default parameters. The **CAGE-fightR** R package<sup>103</sup> was used to identify TSSs and TSS clusters using a minimum pooled value cutoff of 0.1 (`pooledcutoff=0.1`).

### ONT sequencing data analysis

**Guppy** software (v3.4.5) was used to basecall the ONT-MinION sequencing reads. Reads passing the quality filter of 8 (default), were mapped to the reference genome using **minimap2**<sup>104</sup> with the following settings: `-ax splice -Y -C5 -cs`. **ReadStatistics** script from **Seqtools** (<https://github.com/moldovannorberr/seqtools>) was used to compute mapping statistics. The **LoRTIA** toolkit (alpha version, accessed on 20 August 2019, <https://github.com/zsolt-balazs/LoRTIA>) was used to identify TESS, TSSs and introns and to subsequently reconstruct transcripts using these features.

The LoRTIA workflow with default settings was as follows: 1.) for dRNA and dcDNA sequencing: `-5 TGC CATTAGGCCGGG—five_score 16—check_in_soft 15—3 AAAAAAAAAAAAAAAAAA—three_score 16 s Poisson—f true`; and 2.) for o(dT)-primed cDNA reads: `-5 GCTGATATTGCTGGG—five_score 16—check_in_soft 15—3 AAAAAAAAAAAAAAAAAA—three_score 16 s Poisson—f true`.

In case of EHV-1, the searching for adapter sequences was performed by the following command: `samprocessor.py—five_adapter GCTGATATTGCTGGG—five_score 14—check_in_soft 15—three_adapter AAAAAAAAAA AAAAAAAAAA—three_score 14 input output`. The next step in the workflow was the annotation of TSS and TES. For the TES positions, the wobble has been increased to 20, the TSS wobble value remains the default. Then following parameters were used on the 'sam' files: `Stats.py -r genome -f r5 -b 10` and `Stats.py -r genome -f l5 -b 10` for the TSS detection; while `Stats.py -r genome -f r3 -b 20` and `Stats.py -r genome -f l3 -b 20` for TES detection; and `Stats.py -r genome -f in` for intron detection.

Several sequencing techniques were used in the analysis of BoHV-1, the LoRTIA workflow was set up differently depending on sequencing techniques. The following parameters were used for dRNA sequencing: `LoRTIA -5 AGAGTACATGGG—five_score 16—check_in_soft 15 -3 AAAAAAAAAAAAAAAAAAAAAAAAAA—three_score 12 -s poisson -f True`. For oligo d(T) cDNA sequencing: `LoRTIA -5 TGCCATTAGGCCGGGGG—five_score 14—check_in_soft 15—3 AAAAAAAAAAAAAAAAAAAAAAAAAA—three_score 14-s poisson-f True`. During random primer cDNA sequencing: `LoRTIA -5 TGCCATTAGGCCGGGGG—five_score 14—check_in_soft 15 -3 GAAGATAGA GAGCGACA—three_score 14-s poisson -f True`. Also for dcDNA sequencing: `LoRTIA -5 GCTGATATTATTGCT GGG—five_score 16—check_in_soft 15—3 AAAAAAAAAAAAAAAAAAAAAAAAAA—three_score 14 -s poisson -f True`.

A TSS was accepted if the adapters were correct while TESs were accepted if polyA tails were present and there were no false priming events detected by LoRTIA. Further, in the case of KSHV, EBV, EHV-1, TSS were accepted only if at least one dcDNA read, and one dRNA or CAGE read validated it. In the case of CMV, BoHV-1, PRV and HSV-1, TSSs were accepted only if they were present in at least three different samples. In case of introns, we accepted only those that were present in the dRNA sequencing, as this method is considered to be the 'Gold Standard' in identifying alternative splicing events. Several transcripts were manually included if they were a long TSS variant of already accepted TSSs. **MotifFinder** from **Seqtools** was used to find promoter elements around the accepted TSSs.

### Downstream data analysis and visualization

Downstream data analysis was carried out within the R environment, using **GenomicRanges**<sup>105</sup>, **tidygenomics**<sup>106</sup> and packages from the **tidyverse**<sup>107</sup>. **Gviz**<sup>108</sup> was used to create Figs. 2, 3 and 4— (<https://github.com/ivanek/Gviz>). Figure 5 was created using a custom R workflow. Briefly, the 'bam' files were imported into R using **Rsamtools** (<https://bioconductor.org/packages/Rsamtools>). The coverage was calculated and the 5' ends were summed per genomic position in each sample and finally an area-plot was generated using **ggplot2**'s `geom_area` `geom` function. The density along with the genome annotation was plotted using a custom plotting function, utilizing **gggenes** (<https://github.com/wilkox/gggenes>). These scripts are used to import other alignments into R and to generate similar plots from their 3' or 5' distribution or coverage on the reference genomes as well (<https://github.com/Balays/Rlyeh>).

## Data availability

Our data used in this study has been deposited to in European Nucleotide Archive (ENA) under the following accessions: PRJEB24593 (<https://www.ebi.ac.uk/ena/browser/view/PRJEB24593>), ERP106430 (<https://www.ebi.ac.uk/ena/browser/view/PRJEB24593>), PRJEB33511 (<https://www.ebi.ac.uk/ena/browser/view/PRJEB33511>), PRJEB38992 (<https://www.ebi.ac.uk/ena/browser/view/PRJEB38992>), PRJEB22072 (<https://www.ebi.ac.uk/ena/browser/view/PRJEB22072>), PRJEB25680 (<https://www.ebi.ac.uk/ena/browser/view/PRJEB25680>), ERP019579 (<https://www.ebi.ac.uk/ena/browser/view/PRJEB17709>), PRJEB25401 (<https://www.ebi.ac.uk/ena/browser/view/PRJEB25401>), PRJEB25433 (<https://www.ebi.ac.uk/ena/browser/view/PRJEB25433>), PRJEB9526 (<https://www.ebi.ac.uk/ena/browser/view/PRJEB9526>), PRJEB12867 (<https://www.ebi.ac.uk/ena/browser/view/PRJEB12867>) and to Gene Expression Omnibus (GEO) under the accession number: GSE97785 (<https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE97785>). We also acquired datasets from other groups, which were downloaded from GEO: GSE79337 (<https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE79337>), GSE59717 (<https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE59717>), GSE128324 (<https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE128324>), from Sequence Read Archive: PRJNA505045 (<https://www.ncbi.nlm.nih.gov/bioproject/PRJNA505045/>), PRJNA482043 (<https://www.ncbi.nlm.nih.gov/bioproject/?term=PRJNA482043>), PRJNA483305 (<https://www.ncbi.nlm.nih.gov/bioproject/?term=PRJNA483305>), PRJNA533478 (<https://www.ncbi.nlm.nih.gov/bioproject/?term=PRJNA533478>), and from ENA: PRJEB27861 (<https://www.ebi.ac.uk/ena/browser/view/PRJEB27861>), PRJEB42868 (<https://www.ebi.ac.uk/ena/browser/view/PRJEB42868>), PRJEB38829 (<https://www.ebi.ac.uk/ena/browser/view/PRJEB38829>). [https://www.ebi.ac.uk/ena/browser/view/PRJEB64684?fbclid=IwAR2U3rT3i0MTVwb2xYpi9I92DOuyuvLA8aAtWD48Qip6tw11cxpczt\\_hLg](https://www.ebi.ac.uk/ena/browser/view/PRJEB64684?fbclid=IwAR2U3rT3i0MTVwb2xYpi9I92DOuyuvLA8aAtWD48Qip6tw11cxpczt_hLg), PRJEB64684.

## Code availability

LoRTIA: <https://github.com/zsolt-balazs/LoRTIA>. R scripts: <https://github.com/Balays/Rlyeh>. R workflow: [https://github.com/Balays/KSHV\\_RNASeq](https://github.com/Balays/KSHV_RNASeq).

Received: 16 May 2023; Accepted: 22 September 2023

Published online: 29 September 2023

## References

- Roizman, B. Herpes Simplex Viruses. In *Fields Virology* (Lippincott Williams & Wilkins, 2013).
- Deluca, N. A., Mccarthy, A. M. & Schaffer, P. A. Isolation and characterization of deletion mutants of herpes simplex virus type 1 in the gene encoding immediate-early regulatory protein ICP4. *J. Virol.* **56**, 558–570 (1985).
- Fox, H. L., Dembowski, J. A. & DeLuca, N. A. A herpesviral immediate early protein promotes transcription elongation of viral transcripts. *MBio* **8**, 1–16 (2017).
- Hagglund, R. & Roizman, B. Role of ICP0 in the strategy of conquest of the host cell by herpes simplex virus 1. *J. Virol.* **78**, 2169–2178 (2004).
- Zhou, C. & Knipe, D. M. Association of herpes simplex virus type 1 ICP8 and ICP27 proteins with cellular RNA polymerase II holoenzyme. *J. Virol.* **76**, 5893–5904 (2002).
- Aubert, M. & Blaho, J. A. The herpes simplex virus type 1 regulatory protein ICP27 is required for the prevention of apoptosis in infected human cells. *J. Virol.* **73**, 2803–2813 (1999).
- Stinski, M. F. & Meier, J. L. *Human Herpesviruses: Biology, Therapy, and Immunoprophylaxis* (Cambridge University Press, 2007).
- Liu, P. & Speck, S. H. Synergistic autoactivation of the Epstein-Barr virus immediate-early BRLF1 promoter by Rta and Zta. *Virology* **310**, 199–206 (2003).
- Perng, G.-C. *et al.* The latency-associated transcript gene of herpes simplex virus type 1 (HSV-1) is required for efficient in vivo spontaneous reactivation of HSV-1 from latency. *J. Virol.* **68**(12), 8045–8055 (1994).
- Chen, S.-H.H., Kramer, M. F., Schaffer, P. A. & Coen, D. M. A viral function represses accumulation of transcripts from productive-cycle genes in mouse ganglia latently infected with herpes simplex virus. *J. Virol.* **71**, 5878–5884 (1997).
- Wang, Q.-Y.Y. *et al.* Herpesviral latency-associated transcript gene promotes assembly of heterochromatin on viral lytic-gene promoters in latent infection. *Proc. Natl. Acad. Sci. USA* **102**, 16055–16059 (2005).
- Boldogkői, Z., Murvai, J. & Fodor, I. G and C accumulation at silent positions of codons produces additional ORFs. *Trends Genet.* **11**, 125–126 (1995).
- Tormanen, K., Allen, S., Mott, K. R. & Ghiasi, H. The latency-associated transcript inhibits apoptosis via downregulation of components of the type I interferon pathway during latent herpes simplex virus 1 ocular infection. *J. Virol.* **93**, 1–16 (2019).
- Wesley, R. D. & Cheung, A. K. A pseudorabies virus mutant with deletions in the latency and early protein O genes: Replication, virulence, and immunity in neonatal piglets. *J. Vet. Diagn. Investig.* **8**, 21–24 (1996).
- Lee, L. Y. *et al.* A virus with a mutation in the ICP4-binding site in the L/ST promoter of herpes simplex virus type 1, but Not a virus with a mutation in open reading frame P, exhibits cell-type-specific expression of  $\gamma$  1 345 transcripts and latency-associated transcrip. *J. Virol.* **72**, 4250–4264 (1998).
- Tombácz, D. *et al.* Full-length isoform sequencing reveals novel transcripts and substantial transcriptional overlaps in a herpesvirus. *PLoS One* **11**, 1–29 (2016).
- Prazsák, I. *et al.* Long-read sequencing uncovers a complex transcriptome topology in varicella zoster virus. *BMC Genom.* **19**, 873 (2018).
- Tombácz, D. *et al.* Hybrid sequencing reveals novel features in the transcriptomic organization of equid alphaherpesvirus. *SSRN Electron. J.* <https://doi.org/10.2139/ssrn.4141334> (2022).
- Moldován, N. *et al.* Multi-platform sequencing approach reveals a novel transcriptome profile in pseudorabies virus. *Front. Microbiol.* **8**, 1–13 (2018).
- Mackiewicz, P., Zakrzewska-Czerwińska, J., Zawilak, A., Dudek, M. R. & Cebrat, S. Where does bacterial replication start? Rules for predicting the oriC region. *Nucl. Acids Res.* **32**, 3781–3791 (2004).
- Vashee, S. *et al.* Sequence-independent DNA binding and replication initiation by the human origin recognition complex. *Genes Dev.* **17**, 1894–1908 (2003).
- Dorn, E. S. & Cook, J. G. Nucleosomes in the neighborhood. *Epigenetics* **6**, 552–549 (2011).
- Boldogkői, Z., Tombácz, D. & Balázs, Z. Interactions between the transcription and replication machineries regulate the RNA and DNA synthesis in the herpesviruses. *Virus Genes* **55**, 274–279 (2019).
- Santocanale, C. & Diffley, J. F. ORC- and Cdc6-dependent complexes at active and inactive chromosomal replication origins in *Saccharomyces cerevisiae*. *EMBO J.* **15**(23), 6671–6679 (1996).

25. Elias, P. & Lehman, I. R. Interaction of origin binding protein with an origin of replication of herpes simplex virus 1. *Proc. Natl. Acad. Sci. USA* **85**, 2959–2963 (1988).
26. Weller, S. K. & Coen, D. M. Herpes simplex viruses: Mechanisms of DNA replication. *Cold Spring Harb. Perspect. Biol.* **4**, 1–14 (2012).
27. Packard, J. E. & Dembowski, J. A. HSV-1 DNA replication—coordinated regulation by viral and cellular factors. *Viruses* **13**(10), 2015 (2021).
28. Hammerschmidt, W. & Sugden, B. Identification and characterization of oriLyt, a lytic origin of DNA replication of Epstein-Barr virus. *Cell* **55**(3), 427–433 (1988).
29. Ballestas, M. E., Chatis, P. A. & Kaye, K. M. Efficient persistence of extrachromosomal KSHV DNA mediated by latency-associated nuclear antigen. *Science* **284**, 641–644 (1999).
30. 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).
31. AuCoin, D. P., Colletti, K. S., Xu, Y., Cei, S. A. & Pari, G. S. Kaposi's sarcoma-associated herpesvirus (human herpesvirus 8) contains two functional lytic origins of DNA replication. *J. Virol.* **76**, 7890–7896 (2002).
32. Dheekollu, J. *et al.* Article promotes replication termination at oriP and viral episome maintenance || Article Cell-cycle-dependent EBNA1-DNA crosslinking promotes replication termination at oriP and viral episome maintenance. *Cell* **184**, 643–654.e13 (2021).
33. Hammerschmidt, W. & Sugden, B. Replication of Epstein-Barr viral DNA. *Cold Spring Harb. Perspect. Biol.* **5**, 1–14 (2013).
34. Boldogkői, Z., Moldován, N., Balázs, Z., Snyder, M. & Tombácz, D. Long-read sequencing—a powerful tool in viral transcriptome research. *Trends Microbiol.* **27**, 578–592 (2019).
35. Ma, Y. *et al.* Human CMV transcripts: an overview. *Future Microbiol.* **7**(5), 577–593 (2012).
36. Kakuk, B. *et al.* Combined nanopore and single-molecule real-time sequencing survey of human betaherpesvirus 5 transcriptome. *Sci. Rep.* **11**(1), 14487 (2021).
37. Tombácz, D. *et al.* Transcriptomewide survey of pseudorabies virus using next-and third-generation sequencing platforms. *Sci. Data* **5**, 1–13 (2018).
38. Barkley, L. R. & Santocanale, C. MicroRNA-29a regulates the benzo[a]pyrene dihydrodiol epoxide-induced DNA damage response through Cdc7 kinase in lung cancer cells. *Oncogenesis* **2**, e57–e67 (2013).
39. Marchese, F. P. & Huarte, M. A long noncoding RNA in DNA replication and chromosome dynamics. *Cell Cycle* **16**, 151–152 (2017).
40. Dellino, G. I. *et al.* Genome-wide mapping of human DNA-replication origins: Levels of transcription at ORC1 sites regulate origin selection and replication timing. *Genome Res.* **23**, 1–11 (2013).
41. Tikhonovich, I., Liang, B., Seoighe, C., Folk, W. R. & Nasheuer, H. P. Inhibition of human BK polyomavirus replication by small noncoding RNAs. *J. Virol.* **85**, 6930–6940 (2011).
42. Gatherer, D. *et al.* High-resolution human cytomegalovirus transcriptome. *Proc. Natl. Acad. Sci. USA* **108**, 19755–19760 (2011).
43. Tai-Schmiedel, J. *et al.* Human cytomegalovirus long noncoding RNA49 regulates viral DNA replication. *PLoS Pathog.* **16**(4), e1008390 (2020).
44. Rossetto, C. C., Tarrant-Elorza, M. & Pari, G. S. Cis and trans acting factors involved in human cytomegalovirus experimental and natural latent infection of CD14 (+) monocytes and CD34 (+) Cells. *PLoS Pathog.* **9**(5), e1003366 (2013).
45. Huang, L., Zhu, Y. & Anders, D. G. The variable 3' ends of a human cytomegalovirus oriLyt transcript (SRT) overlap an essential, conserved replicator element. *J. Virol.* **70**, 5272–5281 (1996).
46. Prichard, M. N. *et al.* Identification of persistent RNA-DNA hybrid structures within the origin of replication of human cytomegalovirus. *J. Virol.* **72**, 6997–7004 (1998).
47. Rennekamp, A. J. & Lieberman, P. M. Initiation of Epstein-Barr Virus lytic replication requires transcription and the formation of a stable RNA-DNA hybrid molecule at oriLyt. *J. Virol.* **85**, 2837–2850 (2011).
48. Xu, Y., Cei, S. A., Huete, R. A., Colletti, K. S. & Pari, G. S. Human cytomegalovirus DNA replication requires transcriptional activation via an IIE2- and UL84-responsive bidirectional promoter element within oriLyt. *J. Virol.* **78**, 11664–11677 (2004).
49. Norseen, J. *et al.* RNA-dependent recruitment of the origin recognition complex. *EMBO J.* **27**, 3024–3035 (2008).
50. Hubenthal, J., Roizman, B., Voss, J. H. & Roizman, B. Properties of two 5'-coterminal RNAs transcribed part way and across the S component origin of DNA synthesis of the herpes simplex virus 1 genome. *Proc. Natl. Acad. Sci. USA* **85**, 8454–8458 (1988).
51. Tombácz, D. *et al.* Characterization of novel transcripts in pseudorabies virus. *Viruses* **7**, 2727–2744 (2015).
52. Tombácz, D. *et al.* Characterization of the dynamic transcriptome of a herpesvirus with long-read single molecule real-time sequencing. *Sci. Rep.* **7**, 1–13 (2017).
53. Moldován, N. *et al.* Time-course profiling of bovine alphaherpesvirus 1.1 transcriptome using multiplatform sequencing. *Sci. Rep.* **10**, 1–14 (2020).
54. Tombácz, D. *et al.* Multiple long-read sequencing survey of herpes simplex virus dynamic transcriptome. *Front. Genet.* **10**, 1–20 (2019).
55. Torma, G. *et al.* An integrated sequencing approach for updating the pseudorabies virus transcriptome. *Pathogens* **10**, 1–22 (2021).
56. Rutkowski, A. J. *et al.* Widespread disruption of host transcription termination in HSV-1 infection. *Nat. Commun.* **6**, 7126 (2015).
57. Pheasant, K. *et al.* Nuclear-cytoplasmic compartmentalization of the herpes simplex virus 1 infected cell transcriptome is co-ordinated by the viral endoribonuclease vhs and cofactors to facilitate the translation of late proteins. *PLoS Pathog.* **14**(11), e1007331 (2018).
58. Tang, S., Patel, A. & Krause, P. R. Hidden regulation of herpes simplex virus 1 pre-mRNA splicing and polyadenylation by virally encoded immediate early gene ICP27. *PLoS Pathog.* **15**, 1–30 (2019).
59. Whisnant, A. W. *et al.* Integrative functional genomics decodes herpes simplex virus 1. *Nat. Commun.* **11**, 1–14 (2020).
60. O'Grady, T. *et al.* Global transcript structure resolution of high gene density genomes through multi-platform data integration. *Nucl. Acids Res.* **44**(18), e145–e145 (2016).
61. Braspenning, S. E. *et al.* The architecture of the simian varicella virus transcriptome. *PLoS Pathog.* **17**, 1–26 (2021).
62. Oláh, P. *et al.* Characterization of pseudorabies virus transcriptome by Illumina sequencing. *BMC Microbiol.* **15**(1), 1–9 (2015).
63. Tombácz, D. *et al.* Strain Kaplan of pseudorabies virus genome sequenced by PacBio single-molecule real-time sequencing technology. *Genome Announc.* **2**, 6–7 (2014).
64. Csabai, Z., Tombácz, D., Deim, Z., Snyder, M. & Boldogkoi, Z. Analysis of the complete genome sequence of a novel, pseudorabies virus strain isolated in Southeast Europe. *Can. J. Infect. Dis. Med. Microbiol.* <https://doi.org/10.1155/2019/1806842> (2019).
65. Hein, M. Y. & Weissman, J. S. Functional single-cell genomics of human cytomegalovirus infection. *Nat. Biotechnol.* **40**, 391–401 (2022).
66. Koons, M. D., Van Scoy, S. & Hearing, J. The replicator of the Epstein-Barr Virus latent cycle origin of DNA replication, oriP, is composed of multiple functional elements. *J. Virol.* **75**(22), 10582–10592 (2001).
67. Purushothaman, P., Uppal, T. & Verma, S. C. Molecular biology of KSHV lytic reactivation. *Viruses* **7**, 116–153 (2015).
68. Zhu, F. X., Cusano, T. & Yuan, Y. Identification of the immediate-early transcripts of Kaposi's sarcoma-associated herpesvirus. *J. Virol.* **73**, 5556–5567 (1999).

69. Majerciak, V., Alvarado-Hernandez, B., Lobanov, A., Cam, M. & Zheng, Z. M. Genome-wide regulation of KSHV RNA splicing by viral RNA-binding protein ORF57. *PLoS Pathog.* **18**(7), e1010311 (2022).
70. Arias, C. *et al.* KSHV 2.0: A comprehensive annotation of the Kaposi's sarcoma-associated herpesvirus genome using next-generation sequencing reveals novel genomic and functional features. *PLoS Pathog.* **10**(1), e1003847 (2014).
71. Bruce, G. A. *et al.* Quantitative analysis of the KSHV transcriptome following primary infection of blood and lymphatic endothelial cells. *Pathogens* **6**(1), 11 (2017).
72. Schifano, J. M., Corcoran, K., Kelkar, H. & Dittmer, D. P. Expression of the antisense-to-latency transcript long noncoding RNA in Kaposi's sarcoma-associated herpesvirus. *J. Virol.* **91**(4), 10–1128 (2017).
73. Pearce, M., Matsumura, S. & Wilson, A. C. Transcripts encoding K12, v-FLIP, v-cyclin, and the microRNA cluster of Kaposi's sarcoma-associated herpesvirus originate from a common promoter. *J. Virol.* **79**, 14457–14464 (2005).
74. Tombácz, D. *et al.* In-depth temporal transcriptome profiling of an alpha herpesvirus using nanopore sequencing. *Viruses* **14**(6), 1289 (2022).
75. Braspenning, S. E. *et al.* Decoding the architecture of the varicella-zoster virus transcriptome. *MBio* **11**, 1–19 (2020).
76. Ye, X., Zhaoid, Y. & Karijolich, J. The landscape of transcription initiation across latent and lytic KSHV genomes. *PLoS Pathog.* **15**, 1–26 (2019).
77. O'Grady, T. *et al.* Global transcript structure resolution of high gene density genomes through multi-platform data integration. *Nucleic Acids Res.* **44**, 1–17 (2016).
78. Balázs, Z., Tombácz, D., Szűcs, A., Snyder, M. & Boldogkői, Z. Long-read sequencing of the human cytomegalovirus transcriptome with the Pacific biosciences RSII platform. *Sci. Data* **4**, 170194 (2017).
79. Boldogkői, Z. *et al.* Transcriptomic study of herpes simplex virus type-1 using full-length sequencing techniques. *Sci. Data* **5**, 1–14 (2018).
80. Depledge, D. P. *et al.* Direct RNA sequencing on nanopore arrays redefines the transcriptional complexity of a viral pathogen. *Nat. Commun.* **10**, 754 (2019).
81. Tombácz, D., Prazsák, I., Moldován, N., Szucs, A. & Boldogkoi, Z. Lytic transcriptome dataset of varicella zoster virus generated by long-read sequencing. *Front. Genet.* **9**, 1–5 (2018).
82. Tombácz, D. *et al.* Dynamic transcriptome sequencing of bovine alphaherpesvirus type 1 and host cells carried out by a multi-technique approach. *Front. Genet.* **12**, 1–8 (2021).
83. Balázs, Z., Tombácz, D., Szűcs, A., Snyder, M. & Boldogkői, Z. Dual platform long-read RNA-sequencing dataset of the human cytomegalovirus lytic transcriptome. *Front. Genet.* **9**, 432 (2018).
84. Fülöp, Á. *et al.* Integrative profiling of Epstein-Barr virus transcriptome using a multiplatform approach. *Virol. J.* **19**, 1–17 (2022).
85. Boldogkői, Z. Transcriptional interference networks coordinate the expression of functionally related genes clustered in the same genomic loci. *Front. Genet.* **3**, 1–17 (2012).
86. Boldogkői, Z., Balázs, Z., Moldován, N., Prazsák, I. & Tombácz, D. Novel classes of replication-associated transcripts discovered in viruses. *RNA Biol.* **16**, 166–175 (2019).
87. García-Muse, T. & Aguilera, A. Transcription–replication conflicts: how they occur and how they are resolved. *Nat. Rev. Mol. Cell Biol.* **17**, 553–563 (2016).
88. Srivatsan, A., Tehranchi, A., MacAlpine, D. M. & Wang, J. D. Co-orientation of replication and transcription preserves genome integrity. *PLoS Genet.* **6**, e1000810 (2010).
89. Brambati, A., Colosio, A., Zardoni, L., Galanti, L. & Liberi, G. Replication and transcription on a collision course: Eukaryotic regulation mechanisms and implications for DNA stability. *Front. Genet.* **6**, 1–8 (2015).
90. Umbach, J. L. *et al.* MicroRNAs expressed by herpes simplex virus 1 during latent infection regulate viral mRNAs. *Nature* **454**, 780–783 (2008).
91. Yetming, K. D. *et al.* The BHLF1 locus of epstein-barr virus contributes to viral latency and B-Cell immortalization. *J. Virol.* **94**, 10–1128 (2020).
92. Calderwood, M. A., Holthaus, A. M. & Johannsen, E. The Epstein-Barr virus LF2 protein inhibits viral replication. *J. Virol.* **82**, 8509–8519 (2008).
93. Uppal, T., Banerjee, S., Sun, Z., Verma, S. C. & Robertson, E. S. KSHV LANA—the master regulator of KSHV latency. *Viruses* **6**, 4691–4998 (2014).
94. Sampath, P. & DeLuca, N. A. Binding of ICP4, TATA-binding protein, and RNA polymerase II to herpes simplex virus type 1 immediate-early, early, and late promoters in virus-infected cells. *J. Virol.* **82**, 2339–2349 (2008).
95. Guo, L. *et al.* Herpes simplex virus 1 ICP22 inhibits the transcription of viral gene promoters by binding to and blocking the recruitment of P-TEFb. *PLoS One* **7**, 1–11 (2012).
96. Liu, M. *et al.* ICP0 antagonizes ICP4-dependent silencing of the herpes simplex virus ICP0 gene. *PLoS One* **5**, e8837 (2010).
97. Takács, I. F. *et al.* The ICP22 protein selectively modifies the transcription of different kinetic classes of pseudorabies virus genes. *BMC Mol. Biol.* **14**, 1–12 (2013).
98. Papp, B. *et al.* Genome-wide identification of direct RTA targets reveals key host factors for Kaposi's sarcoma-associated herpesvirus lytic reactivation. *J. Virol.* **93**, 1–22 (2019).
99. Tombácz, D., Tóth, J. S., Petrovski, P. & Boldogkoi, Z. Whole-genome analysis of pseudorabies virus gene expression by real-time quantitative RT-PCR assay. *BMC Genom.* **10**, 491 (2009).
100. Soong, R., Ruschoff, J. & Tabiti, K. *Detection of Colorectal Micrometastasis by Quantitative RT-PCR of Cytokeratin 20 mRNA* (Publ, 2000).
101. Livak, K. J. & Schmittgen, T. D. Analysis of relative gene expression data using real-time quantitative PCR and the 2- $\Delta\Delta$ CT method. *Methods* **25**, 402–408 (2001).
102. Dobin, A. *et al.* STAR: Ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15–21 (2013).
103. Thodberg, M., Thieffry, A., Vitting-Seerup, K., Andersson, R. & Sandelin, A. CAGEfightR: Analysis of 5'-end data using R/Bioconductor. *BMC Bioinf.* **20**, 1–13 (2019).
104. Li, H. Minimap2: Pairwise alignment for nucleotide sequences. *Bioinformatics* **34**, 3094–3100 (2018).
105. Aboyoun P, Pages H, L. M. GenomicRanges: Representation and manipulation of genomic intervals. *R Packag. version* (2010).
106. Ahlmann-Eltze, C. Tidygenomics: Tidy verbs for dealing with genomic data frames. *Sci. Rep.* **11**, 14487 (2019).
107. Wickham, H. *et al.* Welcome to the tidyverse. *J. Open Sour. Softw.* **4**, 1686 (2019).
108. Hahne, F. & Ivanek, R. *Visualizing Genomic Data Using Gviz and Bioconductor* (Springer, 2016).

## Acknowledgements

This research was supported by National Research, Development and Innovation Office (NRDIO), Researcher-initiated research projects (Grant Numbers: K 128247 and K 142674) to ZB and by the NRDIO Research projects initiated by young researchers (Grant Number: FK 142676) to DT. The work was also supported by NIH Grant R01AI132554 to ZT. The APC fee was covered by the University of Szeged, Open Access Fund: 5928. ZC was also by the UNKP-21-4-SZTE-126. The work was also supported by the MTA Lendület (Momentum) Programme LP2020-8/2020 to DT.



## Author contributions

G.T.: carried out bioinformatics, data analysis, interpretation and integration of data. D.T.: carried out library preparation, long-read – and CAGE sequencing, participated in interpretation of data and drafted the manuscript. A.I.A.A.: carried out RNA purification and long-read sequencing, participated in analysis. Z.C.: carried out viral infection, propagation of cells, RNA purification, long-read sequencing and qRT-PCR. G.Á.N.: participated in bioinformatics and interpretation of data. B.K.: participated in bioinformatics and interpretation of data. G.G.: participated in sequencing and bioinformatics. L.M.S.: participated in propagation of cells, carried out RNA purification. I.G.: participated in data analysis. Á.H.: participated in data analysis. Á.F.: participated in bioinformatics and interpretation of data. Á.D.: participated in RNA purification, long-read sequencing. I.P.: participated in RNA purification and data analysis. M.M.: carried out propagation of cell cultures, long-read participated in sequencing. V.É.D.: participated in propagation of cells and interpretation of data. V.C.: participated in propagation of cells. Z.Z.: propagated cells and viruses, participated in RNA purification. Z.T.: participated in the design of the experiments and drafted the manuscript. Z.B.: conceived and designed the experiments, supervised the study, wrote the manuscript. All authors read and approved the final paper.

## Funding

Open access funding provided by University of Szeged.

## Competing interests

The authors declare no competing interests.

## Additional information

**Supplementary Information** The online version contains supplementary material available at <https://doi.org/10.1038/s41598-023-43344-y>.

**Correspondence** and requests for materials should be addressed to Z.B.

**Reprints and permissions information** is available at [www.nature.com/reprints](http://www.nature.com/reprints).

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2023