

REVIEW

OPEN ACCESS



Novel classes of replication-associated transcripts discovered in viruses

Zsolt Boldogkői , Zsolt Balázs, Norbert Moldován , István Prazsák , and Dóra Tombác 

Department of Medical Biology, Faculty of Medicine, University of Szeged, Szeged, Hungary

ABSTRACT

The role of RNA molecules in the priming of DNA replication and in providing a template for telomerase extension has been known for decades. Since then, several transcripts have been discovered, which play diverse roles in governing replication, including regulation of RNA primer formation, the recruitment of replication origin (Ori) recognition complex, and the assembly of replication fork. Recent studies on viral transcriptomes have revealed novel classes of replication-associated (ra)RNAs, which are expressed from the genomic locations in close vicinity to the Ori. Many of them overlap the Ori, whereas others are terminated close to the replication origin. These novel transcripts can be both protein-coding and non-coding RNAs. The Ori-overlapping part of the mRNAs is generally either the 5'-untranslated regions (UTRs), or the 3'-UTRs of the longer isoforms. Several raRNAs have been identified in various viral families using primarily third-generation long-read sequencing. Hyper-editing of these transcripts has also been described.

ARTICLE HISTORY

Received 5 November 2018
Revised 3 December 2018
Accepted 18 December 2018

KEYWORDS

Viruses; replication origin; replication fork; DNA polymerase; replication-associated transcript; non-coding RNA; long-read sequencing

DNA replication

The first step of eukaryotic DNA replication is the identification of replication origin (Ori) through the origin recognition complex (ORC), which is a multi-subunit structure composed of Orc1-6p [1]. ORC serves as a landing platform for the assembly of the replication forks. The unwinding of the double stranded DNA molecule into two single strands is initiated at the Ori. This process is carried out by the DNA helicase, an enzyme disrupting the hydrogen bonds between the base pairs of the complementary DNA strands. While the bacterial genome contains a single Ori, the replication of eukaryotic DNA molecule is initiated at hundreds of thousands points. Depending on the species, viruses have a single or a few Oris. DNA viruses usually code for proteins that are responsible for viral DNA replication [2–4]. However, in some cases, such as the Epstein-Barr virus (EBV) during latency, viruses can use the host replication proteins for viral DNA synthesis [5]. The replication fork is the site for the assembly of the replisome, a complex structure composed of the DNA polymerase (DNP), DNA helicase, topoisomerase, primase, DNA gyrase, DNA ligase, and telomerase among others. First, a short (11 ± 1 bp long) RNA sequence is synthesized by the primase enzyme, then this transcript is removed by an endonuclease, which is followed by the elongation with the DNP away from the origin of replication. Due to the 5' to 3' directionality of the DNA synthesis, the leading strand is continuously extended, whereas the lagging strand is synthesized discontinuously from multiple RNA primers. The resulting Okazaki fragments are joined together by the DNA ligase forming a single unified strand. In *E. coli* the termination of DNA replication occurs at specific consensus sequences and results in the disassembly of the replisome [6], while termination in eukaryotes is mostly not sequence specific [7]. Since

eukaryotes have linear DNA molecules, the DNP is unable to synthesize the very ends of the chromosomes (telomeres), and it leads to the shortening of telomeres in each replication cycle. Telomeres act as protective caps to prevent chromosomal integrity. In germ-line cells and in stem cells the telomerase enzyme catalyzes the repair of the telomere sequences by carrying a short complementary RNA molecule used for the priming of this process.

Replication-associated transcripts in prokaryotes and eukaryotes

The genomes of eukaryotic organisms are mostly comprised of non-coding DNA. Transcriptomic studies have revealed that the major parts of these genomic regions are transcriptionally active, producing non-protein coding RNAs (ncRNAs) [8]. These transcripts possess a wide range of functionality at practically all levels of the genetic regulation, including epigenetics, transcription, and post-transcriptional processes [9]. Evidence is also emerging that certain ncRNAs participate in the regulation of DNA replication. Besides the RNA primers and the telomerase RNA component – which were discovered decades ago, many replication-associated (ra) RNAs have been identified in the past few years. A recent genome-wide analysis has demonstrated that 72% of mammalian ORC1s are associated with active promoters, 46.5% of which controls the expression of protein coding genes, whereas 53.5% controls ncRNA genes [10].

Regulation of RNA primer synthesis by raRNAs

The vast majority of bacterial plasmids encode the Rep protein, the function of which is to separate the two DNA strands at the Ori region [11]. An alternative mechanism based on the

regulation of replication by ncRNAs has been demonstrated in a few cases. It has been shown that the Ori region of ColE1 plasmids of *Enterobacteriaceae* encodes two ncRNAs: RNA I and RNA II [12,13]. RNA II acts as a pre-primer by hybridizing with the DNA followed by being processed with RNase H. The resulted RNA fragments tend to initiate the synthesis of the leader DNA strand. RNA I is complementary to a short region of RNA II, and functions as a modulator of RNA II by binding to it, and thereby preventing the formation of RNA II/DNA hybrid that is needed for the initiation of replication. Besides the ColE1 family, a very similar RNA-mediated replication initiation mechanism has also been described in the marine RNA-based (MRB) plasmid family of *Vibrionaceae* [14]. Intriguingly, the sequences of MRB RNA I and RNA II transcripts are evolutionarily unrelated to those of ColE1-regulatory transcripts.

Inhibition of Rep synthesis by raRNAs

The expression of Rep protein itself is regulated in order to control the copy number of the plasmids. The R1 plasmids use, among others, an antisense RNA, termed CopA, for this control. CopA acts at the post-transcriptional level through binding RepA mRNA. The resultant double-stranded RNA is digested by RNase III, thereby preventing RepA synthesis [15]. Antisense RNAs also participate in the control of replication in Collb-p9 plasmids. For the efficient translation of Rep protein, a pseudoknot has to be formed on the Rep mRNA. The Inc RNA has a complementary structure to the Rep mRNA, and therefore, it can block the replication through hybridization to the Rep mRNA [16].

raRNAs mediate ORC recruitment to the Ori

The mechanism of ORC recruitment to the Ori in eukaryotes varies in the different species. The ORC lacks sequence-specific DNA binding motifs (except in yeast [17]), and so far, it has been unclear what factors control ORC binding to the DNA. One of the candidates are the non-coding raRNAs as they can provide sequence-specificity for the origin formation. In the protozoa *Tetrahymena thermophile* ribosomal DNA (rDNA) is amplified ~9,000 times during development. The 26T RNA has been shown to mediate the recruitment of ORC to the Ori region during the rDNA amplification through base pairing with the rDNA Ori [18]. It has also been demonstrated in mammalian systems that G-rich RNA mediates ORC recruitment to telomeres and to AT-rich heterochromatin [19,20]. Vertebrate genomes express the evolutionarily conserved Y RNAs, which are non-coding stem-loop transcripts, and they play a role in the initiation of replication [21,22]. The precise mechanism through which Y RNAs mediate their effects is unknown at present; however, it has been shown that these transcripts are recruited to the chromatin by the ORC.

Replication control by miRNAs

In addition to the above mechanisms, microRNAs (miRNAs) have also been described to control DNA replication through

fine-tuning this process. The miRNAs normally target the complementary mRNAs for translational repression or degradation. For example, in human cells, miR-29a targets the mRNA of Cdc7/Dbf4 kinase, which plays an essential role in the initiation of DNA replication. DNA damage leads to the up-regulation of Cdc7/Dbf4, which is accompanied by the repression of miR-29a in order to maximize the efficiency of repair processes [23].

Replication-associated RNAs in viruses

Replication-associated transcripts have also been identified in viruses. For example, a small replication-regulating (sr)RNA has been recently described in human BK polyomavirus (BKV) isolated from murine mammary tumor cells [24]. This transcript binds simultaneously to both sense and antisense DNA strands within the Ori region of the virus. The srRNA dramatically inhibits the replication of the poliovirus through interfering with the RNA primer synthesis, which changes the structure of the initiation complex even when the regulatory RNA is ectopically expressed in human cells [24]. BKV has an additional RNA-based mechanism for the control of replication: a miRNA targets the mRNA of the large T antigen during the early stage of infection that helps to establish persistence in the host cells [25]. Another example for the viral raRNAs is a highly structured GC-rich transcript of EBV, the function of which is to help for the viral proteins EBNA1 and HMGA1a proteins in ORC recruitment, and the origin formation at various chromosomal locations [26]. The miR-BART2 is a miRNA encoded by the EBV genome, and it binds to the mRNA of BALF5, which is the catalytic subunit of the EBV DNP. Repression of DNP blocks lytic replication, and it this leads to the establishment of latency in the human [27].

nroRNAs – novel replication-associated transcripts in herpesviruses

Next-generation short-read sequencing (SRS), and recently, third-generation long-read sequencing (LRS) have identified several novel raRNA molecules that are expressed from the genomic regions mapped in close vicinity to the replication origins in various viruses. These transcripts designated as near-replication origin (nro)RNAs in herpesviruses and further raRNAs in other viruses are supposed to be produced by a mechanism that regulates the replication initiation and the orientation of replisome progression, which is based on the collision of the replication and transcription apparatuses [28]. The herpesvirus family is subdivided into three subfamilies: alphaherpesviruses, including the human pathogenic Herpes simplex virus type 1 (HSV-1) and the Varicella-zoster virus (VZV) as well as the veterinary pathogen Pseudorabies virus (PRV); betaherpesviruses, such as the Human cytomegalovirus (HCMV); and gammaherpesviruses, including the Epstein-Barr virus (EBV) and Kaposi's sarcoma-associated virus (KSHV). Alphaherpesviruses have two genera: Varicellovirus (e.g. VZV and PRV) and Simplexvirus (e.g. HSV-1). The genome of alpha- and betaherpesviruses are composed of a unique long (UL) and a unique short (US)

region, which are either both bracketed by inverted repeats (IRs) such as in HSV-1, VZV and HCMV, or only the US region is surrounded by IRs such as in PRV (Fig. 1). Alphaherpesviruses express a variety of nroRNAs from the DNA segment around their OriS. PRV contains three OriS, one in its UL region (OriL), and two on the two copies of the IRs (OriS). CTO-S (Fig. 2A), a 286bp ncRNA, is first expressed at the onset of DNA replication (at 4 h post-infection) [28,29]. Despite being the most abundant PRV transcript, CTO-S is practically not expressed at all during the early stage of viral infection [30], which is very rare even among late viral transcripts. CTO-S has a transcription end site (TES) isoform (CTO-AT) which overlaps the convergently-oriented longer UL22 TES variant. CTO-L is a very long TES isoform of UL21 transcript. CTO-M is transcribed using the poly(A) signal of UL21 as a promoter. Except the low-abundance CTO-AS, the rest of the CTO transcripts have a common transcript start site (TSS). The CTO-M and CTO-L transcripts overlap the OriS. The PTO and the PTO-US1 transcripts are encoded by the genomic segments located near the PRV OriLs. The PTO is an ncRNA and it does not overlap the Ori, whereas the PTO-US1 (which can be considered as a very long TSS isoform of US1, or alternatively, a TES variant of PTO) overlaps the OriLs. PTO-US1 contain an intact open reading frame (ORF) of *us1* gene, but it is unknown whether this coding potential is realized in translation or not.

VZV contains two OriSs, one in each IR region, but it lacks OriL. The VZV NTO transcripts are similar to those of the PRV PTOs regarding their locations; however, the sequences of these transcripts are non-homologous (Fig. 2B) [31]: for example, the transcripts NTO1-3 overlap the ORF62 mRNA in an antiparallel manner, which is not the case for PTO; it does not overlap the ORF62-homologue IE180. The NTO2-4 RNAs are all non-coding. The NTO2-3 transcripts do not overlap the OriS, whereas the spliced NTO1 (expressed in two TES isoforms) does overlap it. Furthermore, the longer TSS variants of ORF63 (ORF63-L) and ORF63-64 (ORF63-L-64) are initiated very closely to the OriS (ORF63 is homologous to the US1 gene of PRV and HSV-1). Intriguingly, two closely-related organisms have evolved in parallel a distinct set of transcripts around their OriSs with presumably identical functions. The sequence of the nroRNAs appears to be irrelevant, only the Ori-proximal location seems to be important. A peculiar feature of the NTO3 transcript is that it is A to I hyper-edited, which has not been observed in any other VZV RNAs [31]. Hyper-editing has been shown to play a role in inhibiting RNA interference through making the mRNAs resistant to Dicer cleavage [32]. The function of this nucleotide modification in VZV and the importance of its proximity to the viral replication origin remain to be explored.

Unlike in PRV, the HSV-1 *us1* gene is located in the US region but it also produces OriS overlapping transcripts: the

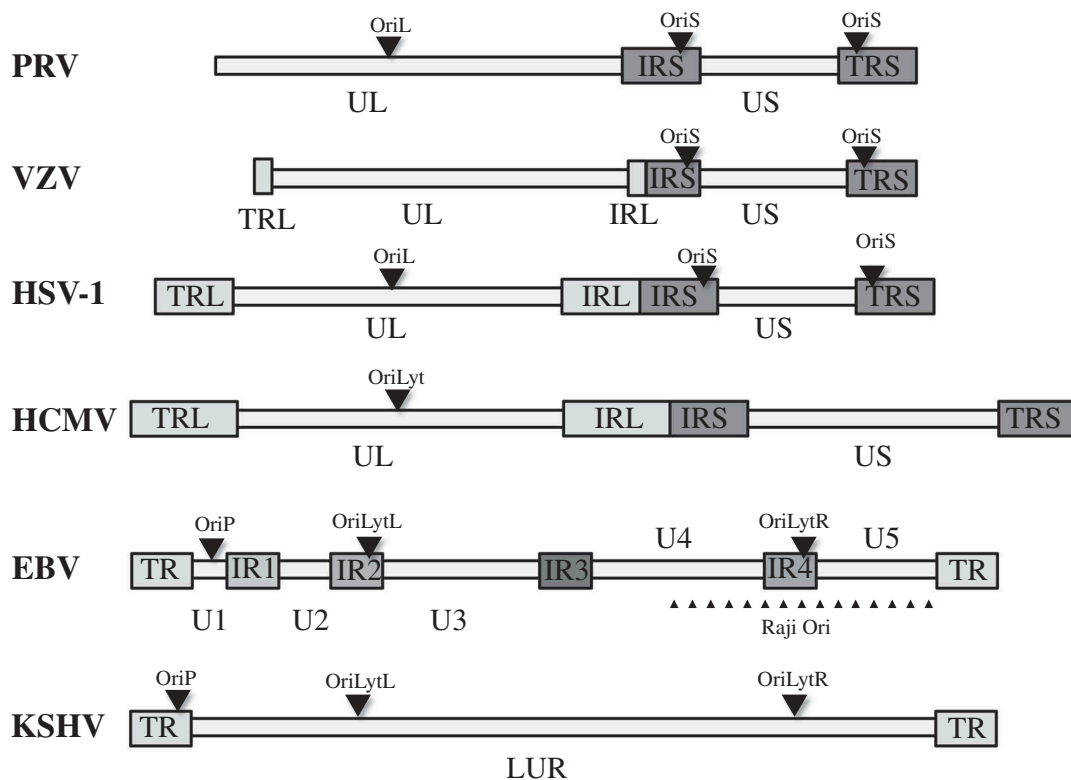


Figure 1. The genomic structure of various herpesviruses. The genomes of herpesviruses are composed of varying number of unique and repeat regions. The replication origins can be located either in the unique, or in the repeat regions, or in both.

Abbreviations: PRV: Pseudorabies virus; VZV: Varicella-zoster virus; HSV-1: Herpes simplex virus type 1; HCMV: Human cytomegalovirus; EBV: Epstein-Barr virus; KSHV: Kaposi's sarcoma herpesvirus; UL: unique long region; US: unique short region; TRL: terminal repeat of UL region; TRS: terminal repeat of US region; IR: internal repeat; LUR: long unique coding region; Ori: replication origin.

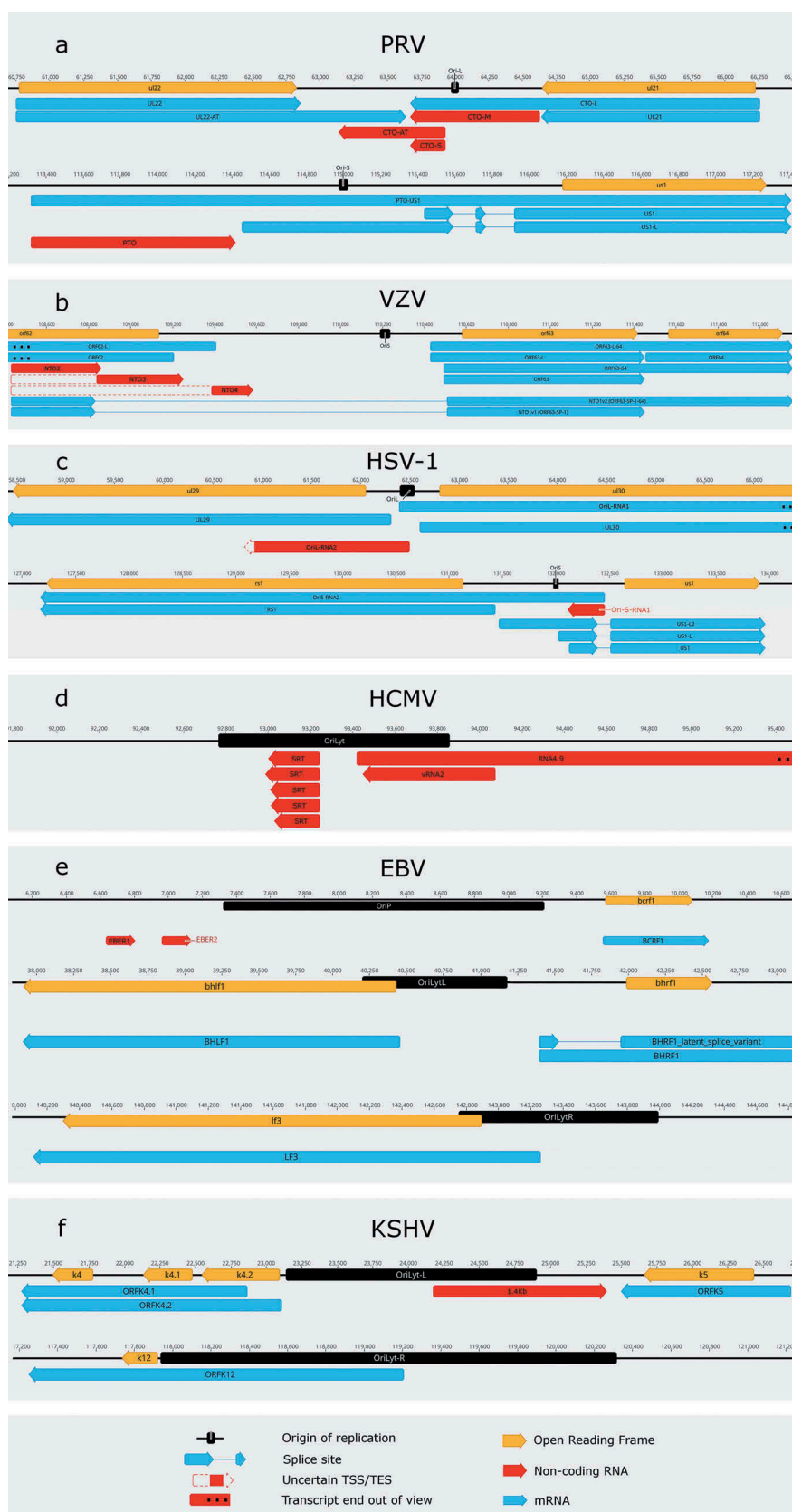


Figure 2. The locations of herpesvirus nroRNAs and replications origins. The replication-associated transcripts can overlap the Ori, or they can be in close vicinity to it. An Ori-overlapping nroRNA can be a non-coding transcript, or alternatively they can be the 5'-UTR, or 3'-UTR region of a longer TSS or TES isoform of a mRNA. Transcription initiation or termination in the proximity of the Ori can also affect the replication.

Coloring black rectangle: Ori; yellow arrow-rectangle: open reading frame (ORF); blue arrow-rectangle: mRNAs; red arrow-rectangle: non-coding RNA; red dashed rectangle: uncertain TSS and red rectangle with three black dots: transcript end terminating out of view. **Abbreviations:** PRV: Pseudorabies virus; VZV: Varicella-zoster virus; HSV-1: Herpes simplex virus type 1; HCMV: Human cytomegalovirus; EBV: Epstein-Barr virus; KSHV: Kaposi's sarcoma herpesvirus

5'-coterminal ncRNA pairs, the OriS-RNA₁ and the OriS-RNA₂ (Fig. 2C). The OriS-RNA₁ has been shown to be expressed with an early kinetics, whereas OriS-RNA₂ is a late transcript [33]. Moreover, the longer isoform of the US1 transcript partially overlaps the OriS with its 5'-UTR [33]. Similarly, the 5'-UTRs of the US12 transcript variants also overlap the OriS. The OriL is located at a different position than in the PRV, but this genomic location also expresses an nroRNA, the OriL-RNA. In HSV, some nroRNAs also overlap with each other in an antiparallel fashion.

The OriLyts of HCMV [34], EBV [35] and KSHV [36] contain binding sites for the transactivator proteins (IE2, Zta and Rta respectively). In HCMV, a bidirectional promoter in the OriLyt region has been shown to regulate replication (Fig. 2D). This promoter could be functionally substituted with an SV40 promoter. Further, IE2 activation has been found to be required both for promoter activity and for DNA replication [37]. The numerous 3'-isoforms of the small replicator transcript (SRT) overlap the essential pyrimidine-rich region of the OriLyt, and therefore, they have been implicated in the regulation of replication. SRT is expressed from 2 h post-infection, and its expression increases throughout the viral replication cycle. RNA4.9, one of the most abundant transcripts of HCMV, also originates in the OriLyt region [38]. It has recently been shown that RNA4.9 regulates lytic replication both in *cis* and in *trans* [39]. Two virion-associated RNA (vRNA) species, one 300bp and another 500bp long, have also been discovered in the OriLyt region forming RNA-DNA hybrids. Such molecules have been supposed to be essential for the initiation of viral replication [40,41].

Lytic gammaherpesvirus DNA replication is dependent on the transcription in the viral Ori. The binding of KSHV Rta to the Rta responsive element has been shown to be essential for viral replication [36]. The produced OriLyt (T1.5) transcript, an early polyadenylated transcript [42], is indispensable for DNA replication [43]. The replication of EBV has also been reported to be dependent on Zta-induced transcription in the OriLyt [35]. The promoters and the transcription start sites of both LF3 and BHLF1 are found in the right and left lytic origins, respectively; however, no study has examined whether the transcription of these specific transcripts are required for the DNA replication.

Replication-associated transcripts in other viruses

Baculoviruses

The baculovirus OriS reside in the AT-rich repeat sequences, which are homologous within the members of *Baculoviridae* family [44,45] (Fig. 3A). The number and position of these homologous regions (*hrs*) varies greatly between the taxa [46]. The most studied baculovirus, *Autographa californica multiple nucleopolyhedrovirus* (AcMNPV) carries nine *hrs* [44,45]. Although the *hrs* have been identified as the main initiators of viral replication, other studies suggest that *non-hr* sequences [47,48] and promoters [49] can also act as origins of the DNA synthesis. It has been previously shown that the whole

AcMNPV genome is transcriptionally active [50,51], resulting in considerable read-through activity across all the *hrs*. LRS studies have revealed that the main source of overlaps is the very long transcripts spanning multiple oppositely-oriented ORFs (complex transcripts): 15 out of the 29 overlaps are formed by complex transcripts, seven by the most abundant isoforms, four by polycistronic transcripts, two by longer 5'-UTR isoforms, while only a single overlap is formed by a longer 3'-UTR isoform [50,51] (Fig. 3). Despite being the most numerous among the rRNAs by sort, complex RNAs are represented in low abundance [51], which indicates a much lower frequency of read-through events than in core and polycistronic transcripts as well as in UTR isoforms. Fourteen of the transcripts overlap the *hrs* with their 3'-UTRs, eleven with their 5'-UTRs, whereas in five transcripts (ORF114, PIF3-ORF114, ORF57-C, FP5K-ORF60-59-58 and FP25-K), the first protein-coding ORF overlaps the *hr* [50,51].

Poxviruses

The DNA primase-helicase encoded by Vaccinia virus (VACV) has been suggested to use RNA primers for the initiation of lagging-strand synthesis [52,53], although former studies proposed a model of rolling hairpin mechanism for viral replication [54,55]. The VACV DNA synthesis is initiated near the genome termini, where a repeat region forms a hairpin end acts as a conserved replication initiation site [56,57]. SRS analysis have revealed multiple potential replication start sites. The start positions of the leading strand are located around the Apex of the junction of concatenated VACV telomeric region within a near 400bp region, and putative Okazaki fragments have been mapped throughout the entire genome [58]. It has been demonstrated that VACV telomeres contain promoters for late RNA production, which may involve in concatemer resolution or in replication [59,60]. Some of these late telomeric transcripts (lateRNAs) bridge the entire Ori region. The 3' ends of the lateRNAs vary in length, and some of them encompass the entire telomeric hairpin, which contains the Ori (Fig. 3B). It has been proposed that the late transcription within the concatemer junction might play a role in opening up the DNA duplex during the initiation of the replication [58].

Circoviruses

A stem-loop structure in the intergenic region serves as the origin of replication for the small circular genome of circoviruses [61]. In the porcine circovirus type 1 (PCV-1), the 5'-UTR region of the core CAP transcript and of its two length isoforms form a full overlap with the Ori. Similarly, the 5'-UTR region of the three REP length isoforms form a full overlap, whereas the eight REP isoforms and the core REP transcript overlap the Ori with 5bps. Additionally, the 3'-UTR of the Atr transcript, encoding the putative ORF3 protein, overlaps the Ori with 18bps, while the two very long non-coding CTR and CTR' transcripts fully overlap the Ori [62-65] (Fig. 3C).

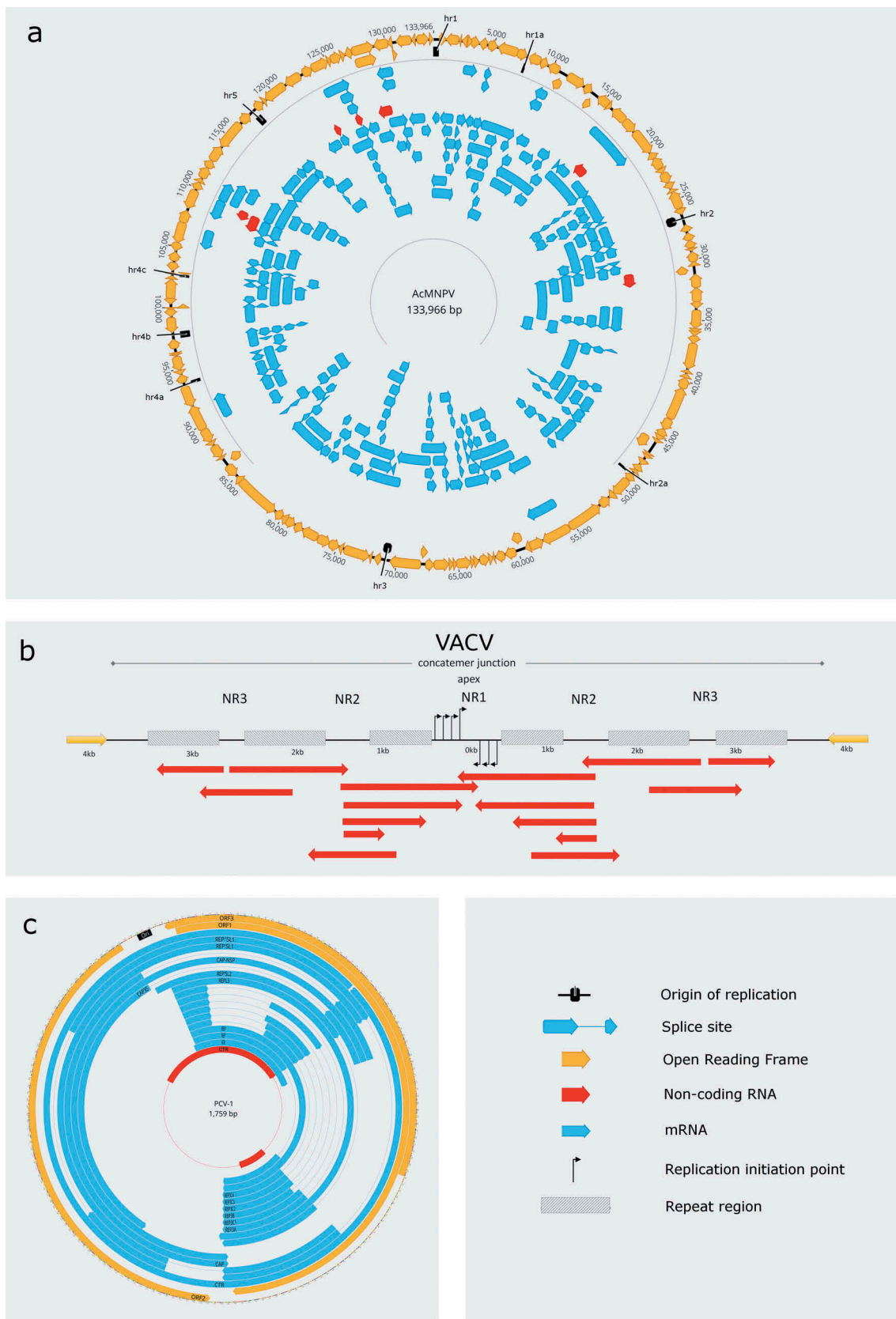


Figure 3. Replication-associated transcripts in a baculovirus, a poxvirus and a circovirus. These transcripts, similarly to the *nroRNA*-s of herpesviruses, are located in the close proximity of the replication origin, or they can overlap the Ori. a. Both ncRNAs and mRNAs have been described to start or terminate relatively close to the Ori of AcMNPV and VACV, while many of these overlap the Ori. b. All transcript isoforms of the PCV-1, except the CTR, overlap the Ori with their 5'-UTRs. c. The concatemer junction region of the Vaccinia virus encloses multiple replication origins [58], which are overlapped by many non-coding transcripts expressed in low abundance. The TSS and TES of these ncRNAs are highly variable.

Coloring: black rectangle: Ori; yellow arrow-rectangle: open reading frame (ORF); blue arrow-rectangle connected with a line: intron; blue arrow-rectangle: mRNAs; red arrow-rectangle: non-coding RNA. The highly variable ncRNAs of VACV are shown by red arrow-rectangles.

Abbreviations: AcMNPV: Autographa californica Multiple Nucleopolyhedrosis Virus, VACV: Vaccinia virus, PCV-1: Porcine circovirus type 1; NR: nonrepeated sequences

Interactions between the transcription and replication machineries

It is debated whether eukaryotic cells efficiently separate transcription and replication [66–68]. Concurrent transcription and replication can lead to the confrontation of the replication fork and the RNA polymerase. Using an *in vitro* system, Liu et al. [69] have demonstrated that RNA polymerase can continue elongation meanwhile the replication fork progresses on the other strand, parallel to it. *In vivo* experiments on a replicating plasmid in *E. coli* have shown that head-to-tail collision does not impede the progression of the replication fork, whereas head on collision does [70]. It has been demonstrated that crashes of the replication fork with the transcription machinery cause genetic instability in yeast and bacteria [71,72]. It has been described that RNA:DNA hybrids formed at sites of transcription-replication clashes, and that RNase H1 acts to suppress the instability of DNA breakage hot spots known as common fragile sites (CFSs) [73]. The authors suggest that replication and transcription are spatially and temporally separated in eukaryotic cells in order to avoid the confrontation of the polymerase molecules. DNA instability has been shown to be dependent on the length of the genes at a given genomic location because transcription of the longer genes takes more time. Helmrich et al. [73] have demonstrated that interference between the replication fork and the transcription complex is inevitable in the longest genes because of the inability of the two processes to be separated at large transcription units during the replication. This results in the formation of CFS, which is prone to genomic instability. Altogether, the authors argue that the

ongoing replication negatively regulates the transcription initiation in the long genes. The co-directional arrangement of the replication and transcription in prokaryotic [74] and eukaryotic [75] genomes may also serve to preclude the collision of these machineries. An alternative explanation for the separation of the transcription and replication is that the interaction between these processes causes this phenomenon, that is, the two processes may be under a common control.

Transcription replication interference network (TRIN)

It has been observed that the overall transcription rate of the individual PRV genes declines following the onset of DNA replication [76]. Similar to other organisms, the expression kinetics of herpesvirus genes is mainly governed by transcription factors acting on the promoters of these genes. Nevertheless, it is also possible that the process of replication affects the gene expression, and *vice versa*, the transcription exerts an effect on the progress of the replication fork. The discoveries of rRNAs in herpesviruses, baculoviruses and circoviruses [28,29,33,50,51,62,77,78], along with the general repressive effect of DNA synthesis on the transcription suggest that the two processes interact with one another on both the replication origins and across the entire genome. We present the proposed mechanism of the effects of nroRNAs on the determination of the orientation of the replication fork progression as well as the facilitatory effect of these transcripts on the DNA synthesis (Fig. 4). It is assumed that the RNP transcribing the CTO-M and CTO-L collides with the replisome (at the OriL) progressing in two

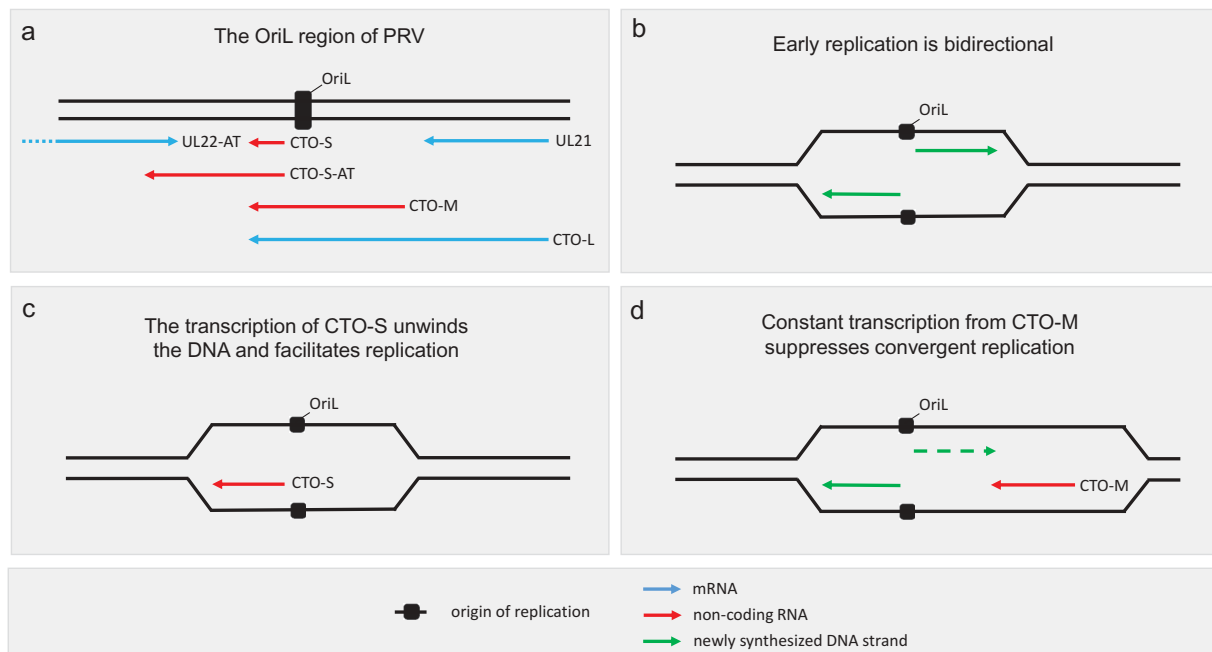


Figure 4. Collision of the RNA polymerase transcribing the nroRNAs and the DNA polymerase during the replication of PRV DNA. A: schematic representation of the transcripts expressed in the region around the PRV OriL (marked by a black box). B: First the PRV DNA is replicated through a bidirectional (theta) replication. C: Transcription initiated at the promoter of the CTO-S transcript furthers replication by unwinding the DNA. D: The CTO-M and CTO-L transcripts exert a repressive effect on DNA replication in one direction and the transcription of CTO-S facilitates it in the other direction by unwinding the DNA strands. In this hypothetical model, the existence of theta-type replication is not necessary, since the proposed collision-based mechanism can prevent its initiation.

directions (θ -type manner), and as a result, it renders the replication to be unidirectional (σ -type). In the meantime, CTO-S transcription unwinds the DNA strands, thereby helping the unidirectionality of the DNA replication. The transcription of CTO-AT is supposed to reduce the expression of the convergent *ul22* gene, thereby eliminating a potential inhibitory effect of transcription on the replication and also advancing the DNA synthesis in a σ -type manner. All in all, the CTO transcripts are thought to have a role in the switch from the θ -type to the σ -type of replication. Alternatively, this mechanism may impede the initiation of bidirectional replication. In this scenario, no θ -type replication occurs in herpesviruses at all. We assume that the interactions between the transcriptional machinery and the replication fork form a transcription and replication interference network (TRIN), which regulates the gene expression globally and determines the rate of the DNA synthesis in a mutually interdependent manner. It is very likely that the *nroRNAs* and the related *raRNAs* are not mere byproducts of this interference mechanism, but they also have functions as transcripts. This hypothesis is supported by the polyadenylation of these RNA molecules, the general function of which is to protect the RNA integrity. The viral *raRNAs* may direct the replication by forming R-loops [79].

Conclusions

It has become evident by now that all of the examined viruses express transcripts or transcript isoforms, which overlap the replication origin, or alternatively they start or end in close vicinity of the Ori. Altogether, we can distinguish four types of *raRNAs* on the basis of their coding potency and position to the Ori: (1) non-coding transcripts that do not overlap the Ori (such as CTO-S, CTO-S-AT, and PTO of PRV; NTO2, NTO3 and NTO4 of VZV; as well as SRT and RNA4.9 of HCMV); (2) non-coding transcripts that do overlap the Ori (such as CTO-M of PRV); (3) mRNA isoforms with very long 3'-UTR (such as CTO-L of PRV); and (4) mRNA isoforms with very long 5'-UTR variant [such as PTO-US1 of PRV, OriS RNA1 of HSV-1, as well as NTO1v1 and NTO1v2 of VZV]. The baculovirus *raRNAs* are all coding, two of them exhibit ORFs-Ori overlapping. The VACV lateRNAs are all non-coding. The difference between the size, location, expression characteristic, and posttranscriptional modification of the various viral *raRNAs* do not necessarily means that they would influence viral replication through distinct mechanisms. On the contrary, the expression of non-homologous transcripts near the Ori suggests that the same function can be solved in different ways, which are the result of convergent evolution. The mechanistic details of how these replication-associated transcripts exert their effects on the DNA synthesis have not yet been ascertained. We hypothesize that these transcripts are produced as byproducts of a mechanism regulating the DNA synthesis through the interaction between the replication and transcription machineries, but these transcripts very likely also function as RNA molecules. The inhibition of replication initiation through intervening in *raRNA* synthesis can be a useful approach for developing effective antiviral therapies. Even though the proteins responsible for

replication differ in viruses and eukaryotes, the mechanistic interactions between the replisome and the transcriptosome may be universal.

Abbreviations

AcMNPV	autographa californica multiple nucleopolyhedrovirus
BKV	BK polyomavirus
CFS	common fragile site
DNP	DNA polymerase
EBV	Epstein-Barr virus
HCMV	Human cytomegalovirus
HSV-1	herpes simplex virus type 1
lateRNA	late telomeric RNA
LRS	long-read sequencing
miRNA	microRNA
ncRNA	non-coding RNA
<i>nroRNA</i>	near-replication-origin RNA
ORC	origin recognition complex
ORF	open reading frame
PRV	Pseudorabies virus
<i>raRNA</i>	replication-associated RNAs
RNP	RNA polymerase
<i>srRNA</i>	small replication-regulating (sr)RNA
SRS	short-read sequencing
TES	transcript end site
TSS	transcript start site
UTR	untranslated region
VACV	Vaccinia virus
VZV	Varicella-zoster virus

Disclosure statement

No potential conflict of interest was reported by the authors.

Funding

This project was supported by NKFIH OTKA K 128247 grant to ZBo and by NKFIH OTKA FK 128252 to DT.

ORCID

Zsolt Boldogkői  <http://orcid.org/0000-0003-1184-7293>

Norbert Moldován  <http://orcid.org/0000-0003-4138-586X>

István Prazsák  <http://orcid.org/0000-0003-3195-503X>

Dóra Tombác  <http://orcid.org/0000-0001-5520-2978>

References

- [1] Santocanale C, Diffley JF. ORC- and Cdc6-dependent complexes at active and inactive chromosomal replication origins in *Saccharomyces cerevisiae*. *Embo J* [Internet]. 1996 Dec 2 [cited 2018 Nov 27];15(23):6671–6679. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/8978693>
- [2] Moss B. Poxvirus DNA Replication. *Cold Spring Harb Perspect Biol*. 2013 Sep;5(9):a010199–a010199.
- [3] Cheung AK. Porcine circovirus: transcription and DNA replication. *Virus Res*. 2012 Mar;164(1–2):46–53.
- [4] Weller SK, Coen DM. Herpes simplex viruses: mechanisms of DNA replication. *Cold Spring Harb Perspect Biol*. 2012 Sep;4(9):a013011.
- [5] Dhar SK, Yoshida K, Machida Y, et al. Replication from oriP of Epstein-Barr virus requires human ORC and is inhibited by geminin. *Cell*. 2001 Aug;106(3):287–296.
- [6] Neylon C, Kralicek AV, Hill TM, et al. Replication termination in *Escherichia coli*: structure and antihelicase activity of the Tus-Ter

- complex. *Microbiol Mol Biol Rev* [Internet]. 2005 Sep [cited 2018 Nov 27];69(3):501–526. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/16148308>
- [7] McGuffee SR, Smith DJ, Whitehouse I. Quantitative, genome-wide analysis of eukaryotic replication initiation and termination. *Mol Cell* [Internet]. 2013 Apr 11 [cited 2018 Nov 27];50(1):123–135. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/23562327>
- [8] Hangauer MJ, Vaughn IW, McManus MT. Pervasive transcription of the human genome produces thousands of previously unidentified long intergenic noncoding RNAs. *PLoS Genet*. 2013 Jun;9(6):e1003569.
- [9] Palazzo AF, Lee ES. Non-coding RNA: what is functional and what is junk? *Front Genet*. 2015 Jan;6:2.
- [10] Dellino GI, Cittaro D, Piccioni R, et al. Genome-wide mapping of human DNA-replication origins: levels of transcription at ORC1 sites regulate origin selection and replication timing. *Genome Res*. 2013 Jan;23(1):1–11.
- [11] Wegrzyn K, Fuentes-Perez ME, Bury K, et al. Sequence-specific interactions of Rep proteins with ssDNA in the AT-rich region of the plasmid replication origin. *Nucleic Acids Res*. 2014 Jul;42(12):7807–7818.
- [12] Tomizawa J, Itoh T, Selzer G, et al. Inhibition of ColE1 RNA primer formation by a plasmid-specified small RNA. *Proc Natl Acad Sci U S A*. 1981 Mar;78(3):1421–1425.
- [13] Masukata H, Tomizawa J. Control of primer formation for ColE1 plasmid replication: conformational change of the primer transcript. *Cell*. 1986 Jan;44(1):125–136.
- [14] Le Roux F, Davis BM, Waldor MK. Conserved small RNAs govern replication and incompatibility of a diverse new plasmid family from marine bacteria. *Nucleic Acids Res*. 2011 Feb;39(3):1004–1013.
- [15] Blomberg P, Nordström K, Wagner EG. Replication control of plasmid R1: repA synthesis is regulated by CopA RNA through inhibition of leader peptide translation. *Embo J*. 1992 Jul;11(7):2675–2683.
- [16] Asano K, Mizobuchi K. Copy number control of IncIalpha plasmid ColIb-P9 by competition between pseudoknot formation and antisense RNA binding at a specific RNA site. *Embo J*. 1998 Sep;17(17):5201–5213.
- [17] Bell SP, Stillman B. ATP-dependent recognition of eukaryotic origins of DNA replication by a multiprotein complex. *Nature* [Internet]. 1992 May 14 [cited 2018 Nov 29];357(6374):128–134. Available from: <http://www.nature.com/doi/10.1038/357128a0>
- [18] Mohammad MM, Donti TR, Sebastian Yakisich J, et al. Tetrahymena ORC contains a ribosomal RNA fragment that participates in rDNA origin recognition. *Embo J*. 2007 Dec;26(24):5048–5060.
- [19] Hoshina S, Yura K, Teranishi H, et al. Human origin recognition complex binds preferentially to G-quadruplex-preferable RNA and single-stranded DNA. *J Biol Chem*. 2013 Oct;288(42):30161–30171.
- [20] Deng Z, Norseen J, Wiedmer A, et al. TERRA RNA Binding to TRF2 Facilitates Heterochromatin Formation and ORC Recruitment at Telomeres. *Mol Cell*. 2009 Aug;35(4):403–413.
- [21] Christov CP, Gardiner TJ, Szuts D, et al. Functional Requirement of Noncoding Y RNAs for Human Chromosomal DNA Replication. *Mol Cell Biol*. 2006 Sep;26(18):6993–7004.
- [22] Krude T, Christov CP, Hyrien O, et al. Y RNA functions at the initiation step of mammalian chromosomal DNA replication. *J Cell Sci*. 2009 Aug;122(Pt 16):2836–2845.
- [23] Barkley LR, Santocanale C. MicroRNA-29a regulates the benzo[a]pyrene dihydrodiol epoxide-induced DNA damage response through Cdc7 kinase in lung cancer cells. *Oncogenesis*. 2013 Jul;2(7):e57–e57.
- [24] Tikhanovich I, Liang B, Seoighe C, et al. Inhibition of human BK polyomavirus replication by small noncoding RNAs. *J Virol*. 2011 Jul;85(14):6930–6940.
- [25] Broekema NM, Imperiale MJ. miRNA regulation of BK polyomavirus replication during early infection. *Proc Natl Acad Sci*. 2013 May;110(20):8200–8205.
- [26] Norseen J, Thomae A, Sridharan V, et al. RNA-dependent recruitment of the origin recognition complex. *Embo J*. 2008 Nov;27(22):3024–3035.
- [27] Barth S, Pfuhl T, Mamiani A, et al. Epstein-Barr virus-encoded microRNA miR-BART2 down-regulates the viral DNA polymerase BALF5. *Nucleic Acids Res* [Internet]. 2007 Nov 26 [cited 2018 Oct 16];36(2):666–675. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/18073197>
- [28] Tombácz D, Csabai Z, Oláh P, et al. Characterization of novel transcripts in pseudorabies virus. *Viruses* [Internet]. 2015 May 22 [cited 2017 Nov 14];7(5):2727–2744. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/26008709>
- [29] Tombácz D, Csabai Z, Oláh P, et al. Full-Length Isoform Sequencing Reveals Novel Transcripts and Substantial Transcriptional Overlaps in a Herpesvirus. *PLoS One* [Internet]. 2016 Sep 29 [cited 2016 Oct 6];11(9):e0162868. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/27685795>
- [30] Tombácz D, Balázs Z, Csabai Z, et al. Characterization of the Dynamic Transcriptome of a Herpesvirus with Long-read Single Molecule Real-Time Sequencing. *Sci Rep* [Internet]. 2017 Mar 3 [cited 2017 Mar 29];7:43751. Available from: <http://www.nature.com/articles/srep43751>
- [31] Prazsák I, Moldován N, Balázs Z, Tombácz D, Megyeri K, Szűcs A, et al. Long-read sequencing uncovers a complex transcriptome topology in varicella zoster virus. *BMC Genomics*. 2018 Dec;19(1):873.
- [32] Scadden AD, Smith CW. RNAi is antagonized by A-to-I hyper-editing. *EMBO Rep*. 2001 Dec;2(12):1107–1111.
- [33] Voss JH, 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 U S A* [Internet]. 1988 Nov 1 [cited 2018 Aug 21];85(22):8454–8458. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/2847162>
- [34] Zhu Y, Huang L, Anders DG. Human cytomegalovirus oriLyt sequence requirements. *J Virol*. 1998 Jun;72(6):4989–4996.
- [35] Fixman ED, Hayward GS, Hayward SD. Replication of Epstein-Barr virus oriLyt: lack of a dedicated virally encoded origin-binding protein and dependence on Zta in cotransfection assays. *J Virol*. 1995 May;69(5):2998–3006.
- [36] Wang Y, Li H, Chan MY, et al. Kaposi's sarcoma-associated herpesvirus ori-Lyt-dependent DNA replication: cis-acting requirements for replication and ori-Lyt-associated RNA transcription. *J Virol*. 2004 Aug;78(16):8615–8629.
- [37] Xu Y, Cei SA, Rodriguez Huete A, et al. Human cytomegalovirus DNA replication requires transcriptional activation via an IE2- and UL84-responsive bidirectional promoter element within oriLyt. *J Virol*. 2004 Nov;78(21):11664–11677.
- [38] Gatherer D, Seirafian S, Cunningham C, et al. High-resolution human cytomegalovirus transcriptome. *Proc Natl Acad Sci U S A*. 2011 Dec;108(49):19755–19760.
- [39] Tai-Schmiedel J, Karniely S, Ezra A, et al. The virally encoded long non-coding RNA4.9 is controlling viral DNA replication. In: *International Herpesvirus Workshop 2018*. Vancouver: University of British Columbia; 2018. p. 2.32.
- [40] Prichard MN, Jairath S, Penfold ME, et al. Identification of persistent RNA-DNA hybrid structures within the origin of replication of human cytomegalovirus. *J Virol*. 1998 Sep;72(9):6997–7004.
- [41] Rennekamp AJ, Lieberman PM. Initiation of Epstein-Barr Virus Lytic Replication Requires Transcription and the Formation of a Stable RNA-DNA Hybrid Molecule at OriLyt. *J Virol*. 2011;85(6):2837–2850.
- [42] Purushothaman P, Thakker S, Verma SC. Transcriptome analysis of Kaposi's sarcoma-associated herpesvirus during de novo primary infection of human B and endothelial cells. *J Virol*. 2015 Mar;89(6):3093–3111.
- [43] Wang Y, Tang Q, Maul GG, et al. Kaposi's sarcoma-associated herpesvirus ori-Lyt-dependent DNA replication: dual role of replication and transcription activator. *J Virol*. 2006 Dec;80(24):12171–12186.

- [44] Kool M, van Den Berg P, MMM, Tramper J, et al. Location of Two Putative Origins of DNA Replication of Autographa californica Nuclear Polyhedrosis Virus. *Virology* [Internet]. 1993 Jan [cited 2018 Sep 6];192(1):94–101. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/8517035>
- [45] Pearson M, Bjornson R, Pearson G, et al. The Autographa californica baculovirus genome: evidence for multiple replication origins. *Science* [Internet]. 1992 Sep 4 [cited 2018 Sep 6];257(5075):1382–1384. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/1529337>
- [46] van Oers MM, Vlak JM. Baculovirus genomics. *Curr Drug Targets* [Internet]. 2007 Oct [cited 2018 Sep 6];8(10):1051–1068. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/17979665>
- [47] Lee HY, Krell PJ. Generation and analysis of defective genomes of Autographa californica nuclear polyhedrosis virus. *J Virol* [Internet]. 1992 Jul [cited 2018 Sep 6];66(7):4339–4347. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/1602548>
- [48] Pearson MN, Bjornson RM, Ahrens C, et al. Identification and Characterization of a Putative Origin of DNA Replication in the Genome of a Baculovirus Pathogenic for *Orgyia pseudotsugata*. *Virology* [Internet]. 1993 Dec [cited 2018 Sep 6];197(2):715–725. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/8249294>
- [49] Wu Y, Liu G, Carstens EB. Replication, integration, and packaging of plasmid DNA following cotransfection with baculovirus viral DNA. *J Virol* [Internet]. 1999 Jul [cited 2018 Sep 6];73(7):5473–5480. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/10364295>
- [50] Chen Y-R, Zhong S, Fei Z, et al. The Transcriptome of the Baculovirus Autographa californica Multiple Nucleopolyhedrovirus in *Trichoplusia ni* Cells. *J Virol* [Internet]. 2013 Jun 1 [cited 2017 Nov 13];87(11):6391–6405. Available from: <http://jvi.asm.org/cgi/doi/10.1128/JVI.00194-13>
- [51] Moldován N, Tombácz D, Szűcs A, et al. Third-generation Sequencing Reveals Extensive Polycistronism and Transcriptional Overlapping in a Baculovirus. *Sci Rep* [Internet]. 2018 Dec 5 [cited 2018 Jun 5];8(1):8604. Available from: <http://www.nature.com/articles/s41598-018-26955-8>
- [52] De Silva FS, Lewis W, Berglund P, et al. Poxvirus DNA primase. *Proc Natl Acad Sci*. 2007;104(47):18724–18729.
- [53] Sèle C, Gabel F, Gutsche I, et al. Low-Resolution Structure of Vaccinia Virus DNA Replication Machinery. *J Virol*. 2013;87(3):1679–1689.
- [54] Baroudy BM, Venkatesan S, Moss B. Structure and replication of vaccinia virus telomeres. *Cold Spring Harb Symp Quant Biol*. 1982;47(2):723–729.
- [55] Baroudy BM, Venkatesan S, Moss B. Incompletely base-paired flip-flop terminal loops link the two DNA strands of the vaccinia virus genome into one uninterrupted polynucleotide chain. *Cell*. 1982;28(2):315–324.
- [56] Traktman P. 27 Poxvirus DNA replication. In: *DNA replication in eukaryotic cells*. Cold Spring Harbor Monograph Archive. New York (NY): Woodbury; 1996. p. 775–798.
- [57] Moss B. Poxvirus DNA replication. *Cold Spring Harb Perspect Biol*. 2013;5(9).
- [58] Senkevich TG, Bruno D, Martens C, et al. Mapping vaccinia virus DNA replication origins at nucleotide level by deep sequencing. *Proc Natl Acad Sci U S A* [Internet]. 2015 Sep 1 [cited 2018 Oct 29];112(35):10908–10913. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/26286988>
- [59] Parsons BL, Pickup DJ. Transcription of orthopoxvirus telomeres at late times during infection. *Virology*. 1990;175(1):69–80.
- [60] Hu FQ, Pickup DJ. Transcription of the terminal loop region of vaccinia virus DNA is initiated from the telomere sequences directing DNA resolution. *Virology*. 1991;181(2):716–720.
- [61] Mankertz A, Persson F, Mankertz J, et al. Mapping and characterization of the origin of DNA replication of porcine circovirus. *J Virol* [Internet]. 1997 Mar 15 [cited 2018 Sep 10];71(3):2562–2566. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/9032401>
- [62] Moldován N, Balázs Z, Tombácz D, et al. Multi-platform analysis reveals a complex transcriptome architecture of a circovirus. *Virus Res* [Internet]. 2017 Jun 2 [cited 2018 Jan 15];237:37–46. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/28549855>
- [63] Mankertz J, Buhk H-J, Blaess G, et al. Transcription Analysis of Porcine Circovirus (PCV). *Virus Genes*. 1998;16(3):267–276.
- [64] Mankertz A, Hillenbrand B. Analysis of transcription of Porcine circovirus type 1. *J Gen Virol* [Internet]. 2002 Nov [cited 2016 Oct 11];83(Pt 11):2743–2751. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/12388810>
- [65] Cheung AK. Comparative analysis of the transcriptional patterns of pathogenic and nonpathogenic porcine circoviruses. *Virology* [Internet]. 2003 May 25 [cited 2016 Oct 11];310(1):41–49. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/12788629>
- [66] Hassan AB, Errington RJ, White NS, et al. Replication and transcription sites are colocalized in human cells. *J Cell Sci*. 1994;107(2):425–434.
- [67] Wei X, Samarabandu J, Devdhar RS, et al. Segregation of transcription and replication sites into higher order domains. *Science* [Internet]. 1998 Sep 4 [cited 2018 Nov 29];281(5382):1502–1506. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/9727975>
- [68] MacAlpine DM. Coordination of replication and transcription along a Drosophila chromosome. *Genes Dev* [Internet]. 2004 Dec 15 [cited 2018 Nov 29];18(24):3094–3105. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/15601823>
- [69] Liu B, Wong ML, Alberts B. A transcribing RNA polymerase molecule survives DNA replication without aborting its growing RNA chain. *Proc Natl Acad Sci U S A*. 1994 Oct;91(22):10660–10664.
- [70] Mirkin EV, Mirkin SM. Mechanisms of transcription-replication collisions in bacteria. *Mol Cell Biol*. 2005 Feb;25(3):888–895.
- [71] Prado F, Aguilera A. Impairment of replication fork progression mediates RNA polII transcription-associated recombination. *Embo J*. 2005 Mar;24(6):1267–1276.
- [72] Torres JZ, Schnakenberg SL, Zakian VA. Saccharomyces cerevisiae Rrm3p DNA helicase promotes genome integrity by preventing replication fork stalling: viability of rrm3 cells requires the intra-S-phase checkpoint and fork restart activities. *Mol Cell Biol*. 2004 Apr;24(8):3198–3212.
- [73] Helmrich A, Ballarino M, Tora L. Collisions between replication and transcription complexes cause common fragile site instability at the longest human genes. *Mol Cell*. 2011 Dec;44(6):966–977.
- [74] Rocha EPC. The Organization of the Bacterial Genome. *Annu Rev Genet*. 2008 Dec;42(1):211–233.
- [75] Huvet M, Nicolay S, Touchon M, et al. Human gene organization driven by the coordination of replication and transcription. *Genome Res* [Internet]. 2007 Jul 25 [cited 2017 Nov 11];17(9):1278–1285. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/17675363>
- [76] Takács IF, Tombácz D, Berta B, et al. The ICP22 protein selectively modifies the transcription of different kinetic classes of pseudorabies virus genes. *BMC Mol Biol*. 2013 Jan;14:2.
- [77] Huang L, Zhu Y, Anders DG. The variable 3' ends of a human cytomegalovirus oriLyt transcript (SRT) overlap an essential, conserved replicator element. *J Virol* [Internet]. 1996 Aug 1 [cited 2018 Aug 21];70(8):5272–5281. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/8764037>
- [78] Tombácz D, Csabai Z, Szűcs A, et al. Long-Read Isoform Sequencing Reveals a Hidden Complexity of the Transcriptional Landscape of Herpes Simplex Virus Type 1. *Front Microbiol* [Internet]. 2017 Jun 20 [cited 2017 Dec 18];8:1079. Available from: <http://journal.frontiersin.org/article/10.3389/fmicb.2017.01079/full>
- [79] Lombrana R, Almeida R, Álvarez A, et al. R-loops and initiation of DNA replication in human cells: a missing link? *Front Genet*. 2015;6:158.