Transcription termination and readthrough in African swine fever virus

Gwenny Cackett[1]*, Michal Sýkora[1], Raquel Portugal[2],
Christopher Dulson[1], Linda Dixon[2] and Finn Werner[1]*

[1]Institute for Structural and Molecular Biology, University College London, London, United Kingdom,
[2]Pirbright Institute, Pirbright, Surrey, United Kingdom

Introduction: African swine fever virus (ASFV) is a nucleocytoplasmic large DNA virus (NCLDV) that encodes its own host-like RNA polymerase (RNAP) and factors required to produce mature mRNA. The formation of accurate mRNA 3′ ends by ASFV RNAP depends on transcription termination, likely enabled by a combination of sequence motifs and transcription factors, although these are poorly understood. The termination of any RNAP is rarely 100% efficient, and the transcriptional "readthrough" at terminators can generate long mRNAs which may interfere with the expression of downstream genes. ASFV transcriptome analyses reveal a landscape of heterogeneous mRNA 3′ termini, likely a combination of *bona fide* termination sites and the result of mRNA degradation and processing. While short-read sequencing (SRS) like 3′ RNA-seq indicates an accumulation of mRNA 3′ ends at specific sites, it cannot inform about which promoters and transcription start sites (TSSs) directed their synthesis, i.e., information about the complete and unprocessed mRNAs at nucleotide resolution.

Methods: Here, we report a rigorous analysis of full-length ASFV transcripts using long-read sequencing (LRS). We systematically compared transcription termination sites predicted from SRS 3′ RNA-seq with 3′ ends mapped by LRS during early and late infection.

Results: Using *in-vitro* transcription assays, we show that recombinant ASFV RNAP terminates transcription at polyT stretches in the non-template strand, similar to the archaeal RNAP or eukaryotic RNAPIII, unaided by secondary RNA structures or predicted viral termination factors. Our results cement this T-rich motif (U-rich in the RNA) as a universal transcription termination signal in ASFV. Many genes share the usage of the same terminators, while genes can also use a range of terminators to generate transcript isoforms varying enormously in length. A key factor in the latter phenomenon is the highly abundant terminator readthrough we observed, which is more prevalent during late compared with early infection.

Discussion: This indicates that ASFV mRNAs under the control of late gene promoters utilize different termination mechanisms and factors to early promoters and/or that cellular factors influence the viral transcriptome landscape differently during the late stages of infection.

Introduction

Evolutionary conserved double-psi beta-barrel (DPBB) RNA polymerases (RNAPs) transcribe the genomes of bacteria, archaea, and eukaryotes (1). Eukaryotic double-stranded DNA viruses including nucleocytoplasmic large DNA viruses (NCLDVs) like variola virus (smallpox), vaccinia virus (VACV), and African swine fever virus (ASFV) also employ a DPBB RNAP to transcribe the viral genome in the cytoplasm of the infected cell. In comparison to their cellular counterparts, the viral transcription systems are understudied and their RNAPs and associated factors are poorly understood, despite their importance as therapeutic targets in the treatment of viral disease. ASFV causes hemorrhagic fever in domestic and wild pigs with almost 100% fatality, and as there are no available antiviral drugs or vaccines, it presents a severe threat to global food security. As seen in other NCLDVs, ASFV particles include all components required for early virus transcription including RNAP, regulatory factors, capping enzyme, and polyadenylate polymerase (2, 3). The accurate formation of the mRNA 5′ end relies on events during initiation, i.e., transcription start site (TSS) selection, which is dependent on the RNAP, initiation factors, and the ASFV promoter sequences we previously identified (4). In comparison to initiation, transcription termination is poorly understood, despite being an essential aspect of gene expression control. However, most terminators across all RNAP transcription systems are "leaky" to some extent and will allow the "readthrough" of some RNAPs into regions downstream of a terminator. In eukaryotes, this phenomenon is often associated with cellular stress and viral infection (5–7). Importantly, premature termination in the upstream region of genes provides a potent means for regulating the transcription output in eukaryotes, archaea, and bacteria. Prokaryotic genomes can be organized into multicistronic operons, where several ORFs are under the control of the same promoter and transcribed as one mRNA, and premature termination is an effective means of modulating the stoichiometry of gene products. Though in the dense prokaryotic genomes of bacteria and archaea, there are both factor-dependent and intrinsic termination mechanisms employed to prevent disruptive readthrough into closely neighboring genes (8, 9).

In summary, transcription termination is not only facilitating precise mRNA 3′ end formation and polyadenylation but also a means for gene regulation.

Most cellular RNAPs from bacteria, archaea, and eukaryotes utilize "intrinsic" or factor-independent means of termination that involves a polyT stretch in the coding strand, equivalent to a polyU tract in the mRNA (10–12). Bacterial intrinsic terminators furthermore include an RNA hairpin secondary structure element upstream of the ~8-nt polyU tract. Termination in *Escherichia coli* can be enhanced or suppressed by accessory factors including NusA, though it is not essential for termination (13). PolyU tracts also facilitate termination by RNAPIII and the archaeal RNAP, but the latter is not dependent on any secondary structures. Phylogenetic analysis of the large RNAP subunits across NCLDVs, vRPB1 and vRPB2, revealed that *Asfarviridae* including ASFV RNAP are most closely related to RNAPI (14). This raises the possibility that ASFV termination also resembles that of RNAPI, which utilizes the Reb1p factor in conjunction with the recognition of a T-rich sequence motif required for transcript cleavage, release, and processing of the nascent transcript (15–18). Termination depends on the RNAPI subunit Rpa12p which is conserved in some NCLDVs like ASFV (vRPB9) but not VACV (15, 16, 19, 20). In contrast to prokaryotes, archaea, and bacteria, termination by eukaryotic RNAPII is coupled to polyadenylation, and cellular stress impairs termination, particularly on genes with a weaker polyadenylation signal (PAS) sequence (21, 22). Interestingly, both RNAPII and VACV utilize a PAS, AAUAAA/AUUAAA and UUUUUN, respectively, which are located upstream of the site at which the RNA terminates (10–30 nt and ~40 nt, respectively) (22–24). For RNAPII, this involves processing by endonucleolytic cleavage and polyadenylation. The process in poxviruses is far less understood than its eukaryotic hosts but is thought to involve a range of factors that change as the stage of infection progresses. Despite the different predicted termination signals in *Poxviridae* and *Asfarviridae*, several transcription termination factors are conserved between them (Table 1).

Improving the characterization of the mRNA 3′ landscape is a vital step toward understanding transcription termination and

TABLE 1 Summary of predicted transcription VACV termination factor homologs conserved in ASFV, their function, and their presence or absence in virus particles.

ASFV gene	In ASFV particles (3)	VACV factor name	VACV gene	In VACV particles (25)	Role in VACV transcription
NP868R (26)	Yes	VTF/CE	*D1R* and *D12L*	Yes	All—transcript capping and termination (27) cVACV–RNAP complex (20)
Q706L (28)	Yes	NPH-I	*D11L*	Yes	All—termination (29) cVACV–RNAP complex (20)
–	–	VLTF-4	*H5R*	Yes	Intermediate/late transcription elongation/termination (30)
–	–	G2	*G2R*	No	Intermediate/late transcription elongation/termination (31)
QP509L | *A859L* (32–35)	No | No	A18	*A18R*	Yes	DNA helicase, intermediate and late transcription termination factor (36, 37)
B962L (32)	Yes	NPH-II	*I8*	Yes	RNA-dependent NTPase, DE-H family (38)

mRNA processing in ASFV. We previously explored the ASFV transcriptome including short-read sequencing (SRS) Illumina-based 3′ RNA-seq in the first genome-wide analysis of transcription termination in ASFV (4). This method, however, had limitations in detecting 3′ end signals from the RNAs of late genes and crucially no way to distinguish signals arising from transcription readthrough. Considering the high variability of ASFV transcript lengths from previous individual gene studies (39–59), the best method for investigating ASFV RNA 3′ end formation is long-read sequencing (LRS), such as sequencing using Oxford Nanopore Technologies (60) as demonstrated in the seminal work by Olasz et al. (61) and Torma et al. (62). We, therefore, followed up the previous SRS analysis of transcription termination in ASFV-BA71V by applying an LRS strategy, to specifically investigate the utilization of termination sites during early and late stages of an ASFV infection time course, with a focus on sequence motif utilization and the frequency of terminator readthrough. We found that from early to late infection, there was a marked increase in readthrough at transcription terminators. Besides the stage of viral gene expression, the relative orientation of genes to their neighbors appears to influence termination. Terminator readthrough is common among ASFV genes, but reduced by longer polyT stretches, and for convergent gene pairs (genes oriented head-to-head), we observed disruptive premature 3′ end formation. Using a highly defined *in-vitro* transcription system, we show that a recombinant ASFV core RNAP is able to terminate faithfully at polyT terminators in the absence of termination factors like NPH-I, which plays an important role in termination for vaccinia virus RNAP (29). ASFV RNAP termination is furthermore independent of RNA secondary structures upstream of the polyT, a hallmark of bacterial intrinsic terminators (11). This emphasizes the conservation of termination mechanisms between ASFV RNAP, archaeal RNAP (10), and RNAPIII (12, 63, 64)—between viruses and cellular domains of life.

Results

Comparison of short- and long-read sequencing in ASFV

To probe for similarities and differences between the 5′ and 3′ ends of ASFV transcripts obtained by short- and long-read sequencing (SRS and LRS, respectively), we systematically compared the SRS techniques 5′ CAGE and 3′ RNA-seq, with LRS Oxford Nanopore Technologies (ONT) sequencing results. We carried out ASFV infection as described previously (4) and isolated total RNA at 5 and 16 h post-infection (hpi) representing early and late stages of infection. Libraries were prepared and sequenced according to the manufacturer's instructions for ONT native RNA sequencing. We first mapped LRS reads to the BA71V genome (Figure 1) and compared the 5′ and 3′ termini to the SRS-derived previously annotated TSSs and transcription termination sites (TTSs), respectively (4). This allowed us to follow each transcript from its originating TSS to its 3′ end, as well as allowed us to analyze where this occurred relative to a gene's ORF and SRS-annotated TTS. We defined each transcript as either terminating prematurely

(Pre), correctly (Corr), or reading through (RT) relative to the SRS-annotated TTS (Figure 2A). Similar to previous observations using ONT native RNA-seq, the mRNA 5′ ends derived from LRS were not well-resolved relative to the TSS annotated by SRS CAGE-seq (61, 62). In contrast, the 3′ mRNA ends derived from LRS showed a close proximity with TTS mapped by 3′ RNA-seq SRS, typically within 10 nt (Figure 2B). This outcome is overall consistent with native RNA sequencing where the 3′ end resolution is better than the 5′ end resolution (66). One of our aims was to create a connection between specific promoter and terminator utilization. Due to the variable 5′ read end locations, we chose a window of within a hundred base pairs of the CAGE-seq TSS to qualify as a cognate or "matched" 5′ end.

This article is focused on RNA 3′ end formation which can be the result of transcription termination and/or processing; we apply a technique (native RNA-seq) that is optimized to isolate and sequence the 3′-polyadenylated transcripts. Unlike 3′ RNA-seq, native RNA-seq is not subject to the transcript-internal mispriming at A-rich sequences as it specifically selects RNA 3′ ends to initiate the sequencing. However, it does require transcripts to be polyadenylated as previously reported for ASFV mRNAs (2, 4). Of the 41,265 reads originating at a mapped ASFV promoter, the 3′ ends from approximately half the mRNAs (20,189 reads) were matched to that same gene's primary (p)TTS (from 151 genes in total) mapped by 3′ RNA-seq SRS (Figure 2B). The pTTS was previously defined as the RNA 3′ with the largest number of reads downstream of that gene's ORF. The other half of the LRS reads terminated >100 nt either upstream or downstream of the SRS-defined pTTS (Figure 2C); these 3′ termini are potentially generated by premature termination, RNA processing, or terminator readthrough, respectively. The biological importance of terminator readthrough relates to the possibility of multicistronic mRNAs (61, 62). However, it is not certain if the additional ORFs downstream of the first ORF will be translated in the infected cell, especially as no internal ribosome binding sites have been identified in ASFV. The fact that approximately half of the 3′ termini are not associated with primary transcription terminators suggests a complex termination landscape, suggesting multiple transcription termination and additional RNA 3′ formation processes.

Detailed comparison of LRS TTS mapping with 3′ RNA-seq

A total of 10,885 LRS reads, from 151 BA71V (Vero-adapted ASFV strain) genes (4), matched both their LRS 5′ and 3′ ends to SRS pTSS and pTTS, respectively, corroborating our LRS approach. What makes a terminator strong, i.e., associated with low readthrough? From previous results, the SRS pTTSs were associated with either a polyT motif or no motif (4). This primary sequence motif consists of >4 T residues in the coding DNA strand, corresponding to >4 U residues in the mRNA. The number of U residues correlated with the proportion of correctly terminating transcripts ($R = 0.38$) and anticorrelated with the proportion of mRNA 3′ ends generated by terminator readthrough ($R = -0.55$), while we found no correlation with premature termination
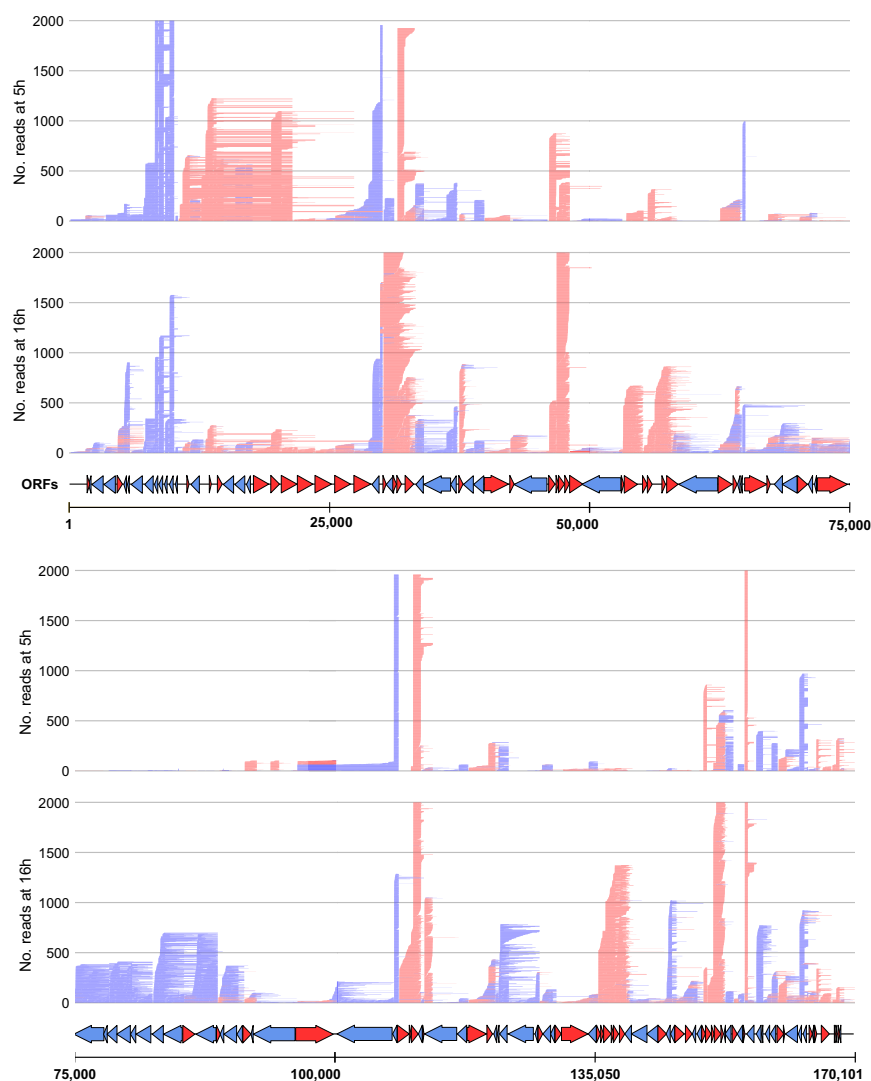
FIGURE 1
Mapping of the long-read sequencing (LRS) reads to the BA71V genome. Visualized in R using the gggenes (65) package. For visualization purposes, the genome was split in half, and each panel shows the long-read sequencing reads aligned from 5 hpi to 16 hpi (indicated by the left access). Arrows indicate BA71V ORFs oriented and colored according to their coding strand (red for plus, blue for minus).

($R=$ −0.018) (Figure 2D). In essence, longer T stretches reduce terminator readthrough. Given that ASFV is AT-rich (39% GC content in BA71V), it is common to find polyT and polyA sequences. There are 3,743 ≥4 T stretches across both strands of the BA71V genome, which may necessitate a means of control beyond the sequence context alone, e.g., by termination factors.

Motifs enriched at the 3′ ends of ASFV mRNAs

We scrutinized the genome-wide enrichment of any motifs at the RNA 3′ ends during early (Figures 3A, B) and late infection (Figures 3C, D). We observed a clear enrichment of polyT motifs at 3′ ends during both early and late infection, while the second most

common motif was a polyA. The polyT motif frequency was fivefold that of polyA at 5 hpi, decreasing to parity at 16 hpi (summarized in Figure 3E). This is consistent with our observations based on SRS 3′ RNA-seq, which identified polyT terminators as more prevalent among early compared with late genes. In our previous SRS 3′ RNA-seq approach, polyA signatures were filtered out due to the possibility of transcript-internal mispriming (4), but this independent verification by LRS native RNA-seq (free of any primed PCR step) demonstrates that there are indeed genome-templated polyA sequences at the 3′ ends of ASFV mRNAs.

Since a polyA stretch on one DNA strand corresponds to a polyT stretch on the other, it is possible that the head-on collision of transcription elongation complexes (TECs) on convergent gene pairs results in the termination of either or both RNAPs on such a site (67). The two genes *B646L* and *B385R* provide an example of a converging
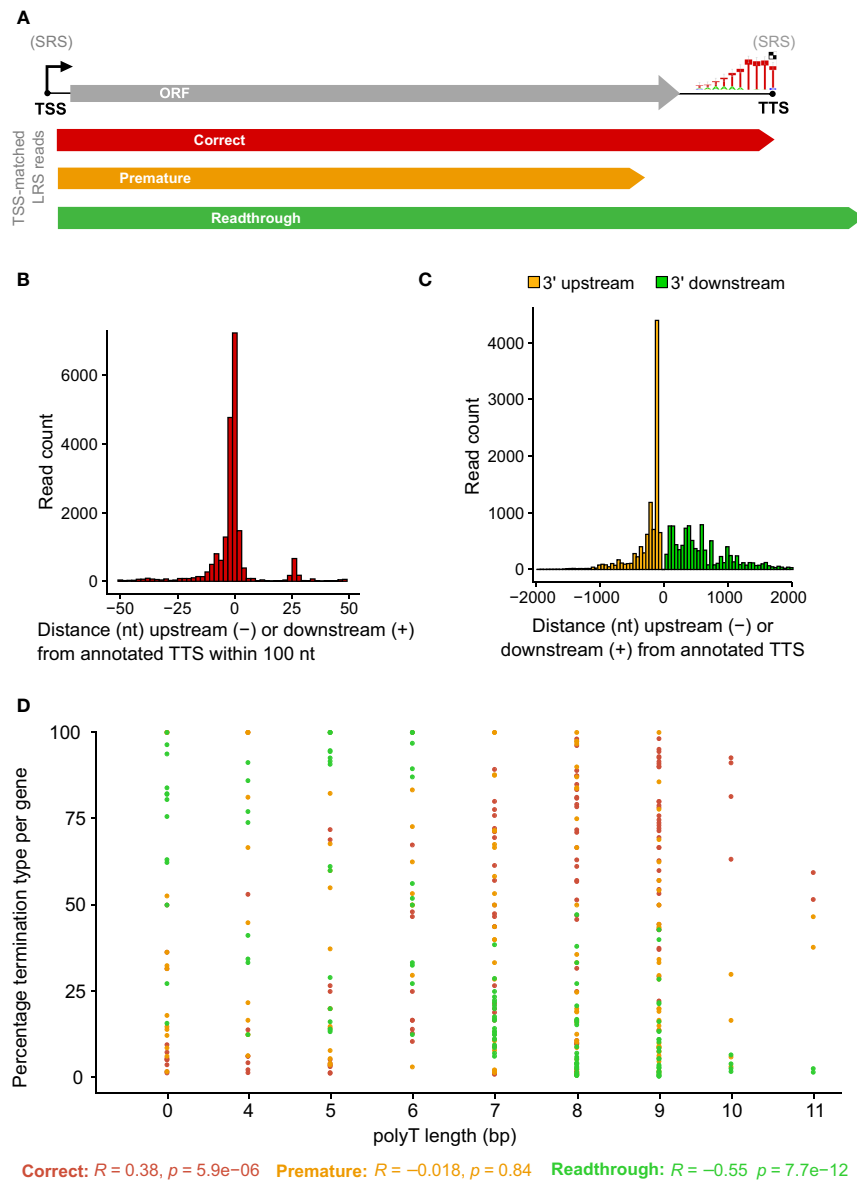
FIGURE 2

Comparison between LRS reads and SRS 3′ RNA-seq-based annotations. **(A)** Simplified schematic representation of how reads were classified according to their 3′ end location relative to the 3′ RNA-seq annotated pTTS. **(B)** Distribution of 3′ end (red) locations relative to that annotated 3′ RNA-seq TTS, from the 20,189 of 41,265 ONT reads which matched the 5′ end and the 3′ end—defined as within 100 bp of the TSS or TTS, respectively. **(C)** The locations of 3′ ends for the 21,076 reads which matched their 5′ end to the TSS location, but not the 3′ end with the pTTS location (bin width = 50 bp). **(D)** For each ASFV gene, the percentage of TSS-matched LRS reads that terminate prematurely, correctly, or readthrough relative to the SRS pTSS. Pearson correlation coefficients and *p*-values for percentage termination type versus polyT length at the SRS pTTS are shown underneath.

gene pair (Figure 3F). *B646L* is a well-characterized and highly expressed late gene that encodes the capsid protein p72. While its TSS determined by CAGE-seq was distinct and clear, the TTS signal was scattered, with multiple associated SRS 3′ RNA-seq peaks located downstream of its stop codon. LRS demonstrates that B646L transcription starts consistently at the *B646L* promoter with little or no readthrough originating from upstream genes B117L or B407L (Figure 3F). The *B646L* transcript 3′ ends, however, are located over a broad region downstream of its stop codon. There are 21 polyT stretches (≥4 T's) in the downstream region of *B646L*, with enrichment of polyT stretches on both strands in the intergenic

region with the converging *B385R* gene (Figure 3F, asterisk). However, rather than terminating transcription consistently at any of these clear terminator motifs, many *B646L* transcripts consistently read through into the downstream region of *B385R*. Given that the converging *B385R* gene is expressed at much lower levels, it is tempting to speculate that the high expression levels of *B646L* are connected to this poor termination behavior. For example, multiple TECs in tandem could "force" RNAP through termination signals and rarely collide with a TEC transcribing the *B385R* gene.

The pattern observed with readthrough from *B646L* is not a universal rule, however, as can be observed with the highly expressed
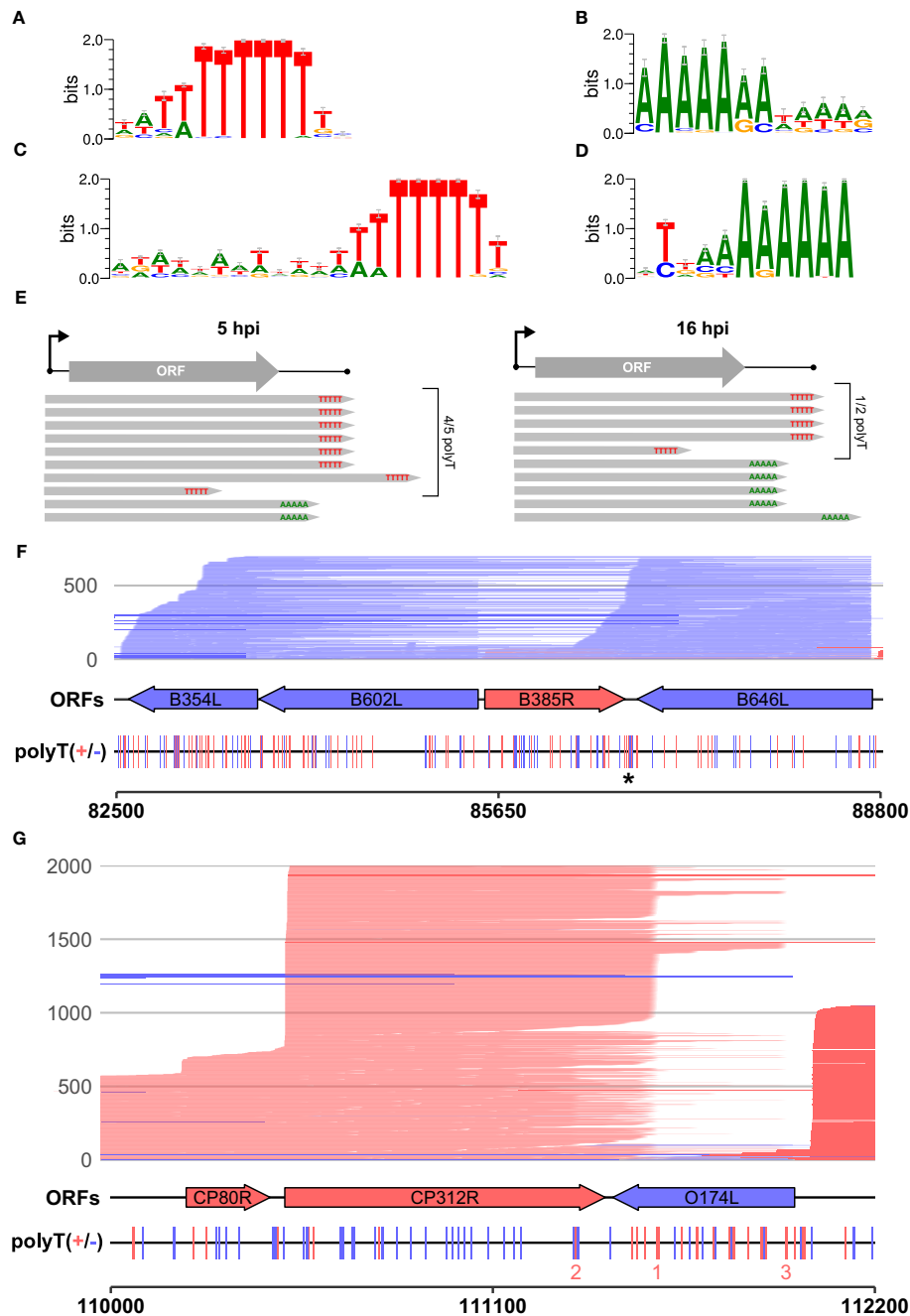
FIGURE 3

The 3′ ends of reads: enriched motifs and gene examples. MEME motif searches were carried out on all reads whose 5′ ends matched the CAGE-seq data for an annotated gene. The 3′ end nt of each read plus and minus 20 nt on either side were extracted for 3,823 and 9,216 total reads from 5 h to 16 h, respectively. **(A, B)** The first and second most significant motifs, respectively, detected from reads at 5 h. **(A)** This was found at 958 sites (E-value 3.2e−188) and **(B)** at 177 sites (E-value 1.1e-028). **(C, D)** The first (685 sites, E-value: 8.3e−050) and second motifs (580 sites, E-value 2.4e−005) found at the 3′ ends of reads from 16 h, respectively. **(E)** Summary schematic of analysis and results from panels **(A−D)**, i.e., after matching reads from 5 hpi to 16 hpi to their respective TSSs; the significant motifs found at 3′ ends of reads were mostly polyT (polyU) during early infection and an almost equal mix of polyT and polyA during late infection. **(F)** Full-length transcript landscape including, and downstream of, the gene *B646L*, representing non-discrete termination sites. The asterisk (*) indicates a polyT-rich region that could facilitate termination for either of the genes *B385R* and *B646L* but shows no clear enrichment of 3′ ends. **(G)** Full-length transcript landscape surrounding the gene *CP312R*, representing discrete termination sites. Reads are capped at 2,000 total reads for visualization. Total reads from 16 h are shown for the region of the BA71V genome indicated with the bottom scale for both **(F, G)**. Blue (minus) and red (plus) indicate strandedness of ORFs, polyT stretches of ≥4, and reads.

early gene *CP312R* that converges with the less well-expressed *O174L* (Figure 3G). *CP312R* mRNAs are initiated by the *CP312R* promoter, or alternatively by the utilization of the upstream *CP80R* and *CP530R* promoters, due to readthrough. Regardless of the promoter utilization,

CP312R, *CP80R*, and *CP530R* transcripts consistently terminate at discrete sites: one pTTS (1 in Figure 3G) and two secondary termination sites (2 and 3 in Figure 3G). Given these contrasting patterns for genes with seemingly similar local organization, it raises

the question of how much a gene's local environment affects its termination patterns, and if this differs between early and late ASFV genes. Further comparison between converging gene pairs with more similar gene expression levels shows reads stopping just short of one another at the two pTTS in Supplementary Figure 1A or overlapping and generating a "clash region" of ~100 bp between *A276R* and *A238L* in Supplementary Figure 1B. This latter gene pair also shows an example of the same terminator being used by two genes, but on opposing strands: the pTTS of *A276R* and an npTTS of *A238L* are located on the same terminator, but on opposing strands, generating a polyT and polyA motif, respectively.

The role of gene organization for transcription termination

The ~170-kb ASFV BA71V genome is densely packed with genes on both strands: 73 on the plus strand and 80 on the minus strand in the genome U18466.2. Gene pairs can be oriented in convergent (head-to-head) or tandem (contiguous genes) arrangement (as illustrated in Figure 4A), but one gene can be assigned to both categories when it is in tandem relative to the upstream gene and convergent to the downstream gene, or vice versa (Figure 4B). To systematically assess any dependence on the orientation of the downstream gene, we only considered genes according to their relative orientation to the closest downstream gene. Genes were classified as tandem or converging as summarized in Table 2 (detailed in Supplementary Table 1). Overall, ASFV genes are distributed equally on either strand and tend to be non-overlapping, with some exceptions (4). Our analysis shows that tandem gene pairs are more prominent than convergent ones genome-wide and that there are less convergent early compared with late genes (~38% and ~51%, respectively). A key factor in this is the layout of predominantly tandem-oriented multigene family (MGF) members toward the genome termini. The evolutionary selection pressures that have resulted in this genome organization are not fully understood but may have been shaped by optimizing or facilitating correct gene expression levels including transcription termination.

We compared the proportion of mRNAs associated with i) correct, ii) premature termination, or iii) terminator readthrough for the tandem or convergent gene orientation (Figure 4C). The results show that premature termination is dominant among convergent genes (Figure 4D). Figures 4E–H illustrate the quantification of 3′ ends from four genes with different expression and genomic context patterns: early tandem Y118L (Figure 4E), early convergent *CP312R* (Figure 4F), late tandem *A224L* (Figure 4G), and late convergent *B646L* (Figure 4H). These results indicate that both early genes and the late tandem gene *A224L* have relatively consistent 3′ end formation patterns, albeit the latter with more premature termination. In contrast, late convergent *B646L* is characterized by an abundance of both premature and readthrough transcripts (see also Figure 3F).

De-novo definition of TTSs using LRS

Thus far, we have considered each full LRS-mapped transcript solely in the context of termination sites defined using SRS, but only approximately half of the LRS reads matched the previously mapped pTTS. Our analysis suggests that SRS had correctly picked up an accumulation of 3′ ends for these genes and, therefore, a putative termination site (previously defined as a pTTS). However, many of the reads originating from these gene promoters were not terminating at the SRS TTS. Supplementary Figure 2 shows examples of early and late genes whose mRNA 3′ ends were predominantly not located at the SRS-annotated pTTS.

The power of LRS is to capture the mRNA transcript in its entirety and to unequivocally assign which promoters (or TSSs) are associated with which gene terminators (or TTSs). We identified locations with an accumulation of RNA 3′ ends, newly defining these LRS TTSs (Figure 5A). For each of these 376 LRS TTSs, we searched upstream to identify cognate TSSs. We detected TTSs originating from 115 gene TSSs in total and subsequently defined the strongest TTS downstream of any ORF within that transcription unit as the pTTS (TTS, from here onwards refers to LRS TTS). For the >40,000 reads whose 5′ ends matched a TSS, the vast majority of their 3′ ends coincide with the new LRS-defined pTTSs (Figure 5B), indicating LRS TTSs captured the 3′ end landscape in a more comprehensive fashion, compared with those defined via SRS (as illustrated in Supplementary Figure 3A versus Supplementary Figure 3B). Supplementary Table 2 shows matched LRS and SRS TTSs, demonstrating 80 being perfectly matched, i.e., both methods defined them as either primary or non-primary and correctly identified them from which gene promoter transcripts they originated.

ASFV genes commonly share and utilize multiple termination sites

We categorized the novel 376 sites into four different "TTS types" (illustrated in Supplementary Figure 3B) representing the location of each TTS with respect to their originating gene, or lack thereof in the case of "intergenic" TTSs (Figure 5C; Supplementary Table 2). TTSs were defined as primary or non-primary TTSs (pTTS or npTTS) according to their prevalence (number of reads). TTSs within the gene's ORF were classified as "intra-ORF." This approach clarified and highlighted that many TTSs did indeed originate from a single TSS. However, many TTSs (primary and non-primary) were being used by several genes. The sharing of pTTS and npTTS is schematically illustrated in Supplementary Figure 3C, and two example genes (*D79L* and *D339L*) that share both primary and non-primary termination sites are shown in Figure 5D. We furthermore found that early genes and highly expressed genes on average use a greater number of TTSs although we cannot rule out that the detection limit of the method contributes to this effect (Supplementary Figures 3D, E, respectively). Three genes (*A151R*, *A224L*, and *A104R*) annotated in Supplementary Figure 3D had an unusually large number of TTSs and were all found in close proximity to one another (Figure 6). This region shows high levels of readthrough but consistent usage of distinct TTSs by both early and late transcripts (Figures 6A, B, respectively). Genes in this region are examples of both extensive sharing of the same TTSs, as well as genes that use a high number of
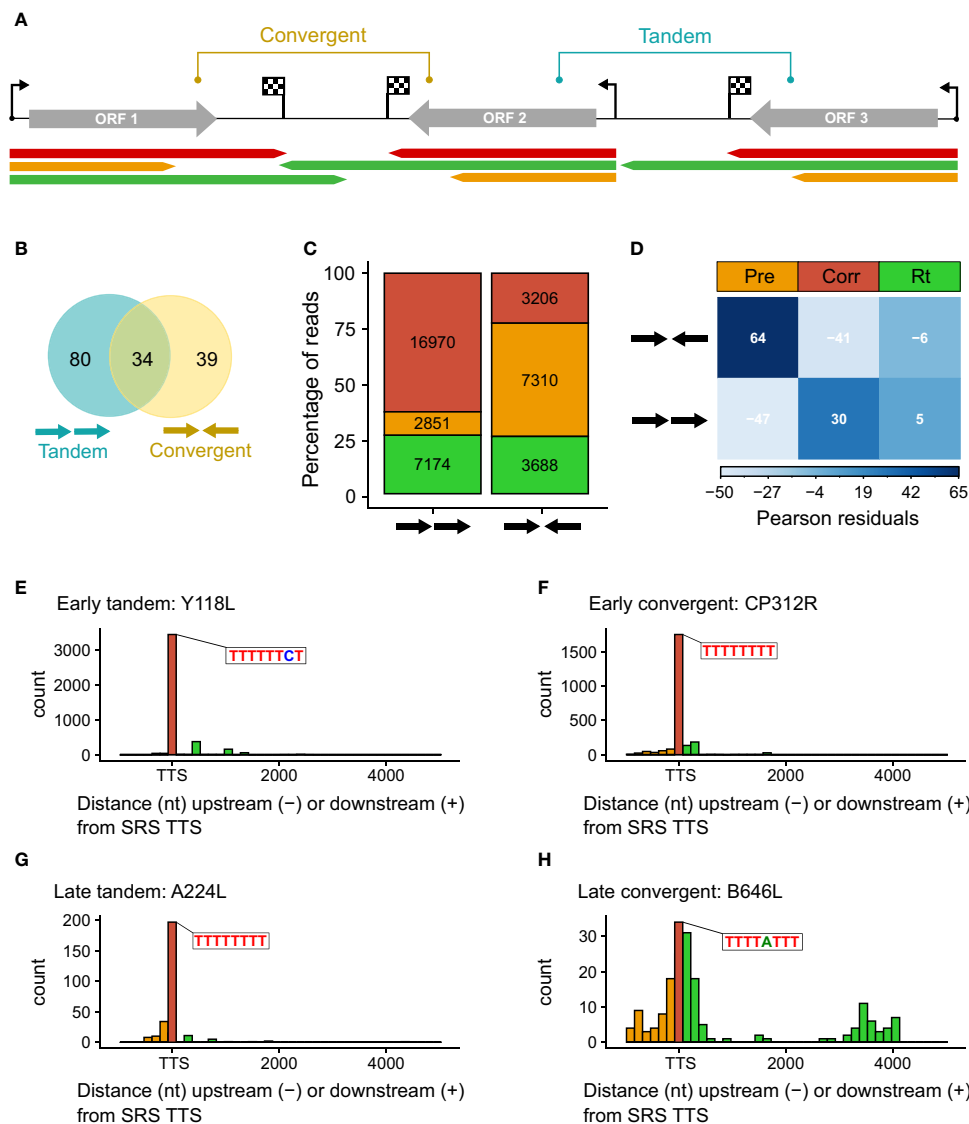
FIGURE 4

(A) Schematic representing the categories of tandem and convergent gene layout, along with examples of LRS read mapping, colored according to termination type. **(B)** Summary Venn diagram representing the positioning of genes relative to one another across the BA71V genome. BEDTools was used to classify each of the 153 BA71V genes [as annotated in Cackett et al. (4)], according to the next genes up- and downstream. **(C)** Bar chart showing the proportion of reads showing each termination type versus their gene organization (tandem or convergent). Termination type is colored as before (red, amber, and green represent correct, premature, and readthrough, respectively). The bar height represents the percentage of termination type per gene layout type, annotated with the number of reads per termination type. These reads were extracted from the 41,265 which matched the 5′ ends from CAGE-seq; 24 reads were excluded due to no annotated gene downstream (at the genome termini). **(D)** Correlation matrix plot following a chi-squared test of independence on read frequency per termination type against each gene layout. Pearson's chi-squared test of independence: χ^2 = 4214.7, p-value < 0.001. The scale indicates Pearson residuals, with navy indicating a strong positive association (e.g., between converging genes and reads prematurely terminating or between contiguous genes and readthrough) and white indicating a strong negative association (e.g., between converging genes and reads prematurely reading through or between contiguous genes and reads prematurely terminating). **(E–H)** Distribution of distances between the 3′ read ends from LRS versus the 3′ RNA-seq TTSs. Shown as histograms with a bin width of 150 nt for every graph. There are two examples each for early tandem and convergent genes (*Y118L* and *CP312R*) and late tandem and convergent genes (*A224L* and *B646L*). Color scheme as before: amber, red, and green represent premature, correct, and readthrough termination relative to the 3′ RNA-seq TTS (or ORF stop codon in the case of A104R, shown in blue).

them, such as *A104R* whose transcripts extend kilobases beyond its stop codon, utilizing the same terminators of *A118R*, *A151R*, and *A276R*. *A104R* is a highly expressed late gene, and analysis of the 3′ end formation between *A104R* and similarly expressed *B646L*, *K78R*, and *E184L* shows high levels of heterogeneity, with premature and readthrough transcripts being highly abundant (Supplementary Figure 4).

LRS termination site motifs and tandem terminators

After defining TTSs based on LRS data, we investigated the sequence motifs at the different TTS types including 111 pTTSs (Figures 7A, B), 179 npTTSs (Figures 7C, D), and 87 intra-ORF-TTSs (Figure 7E). Supplementary Table 2 lists these TTSs and their
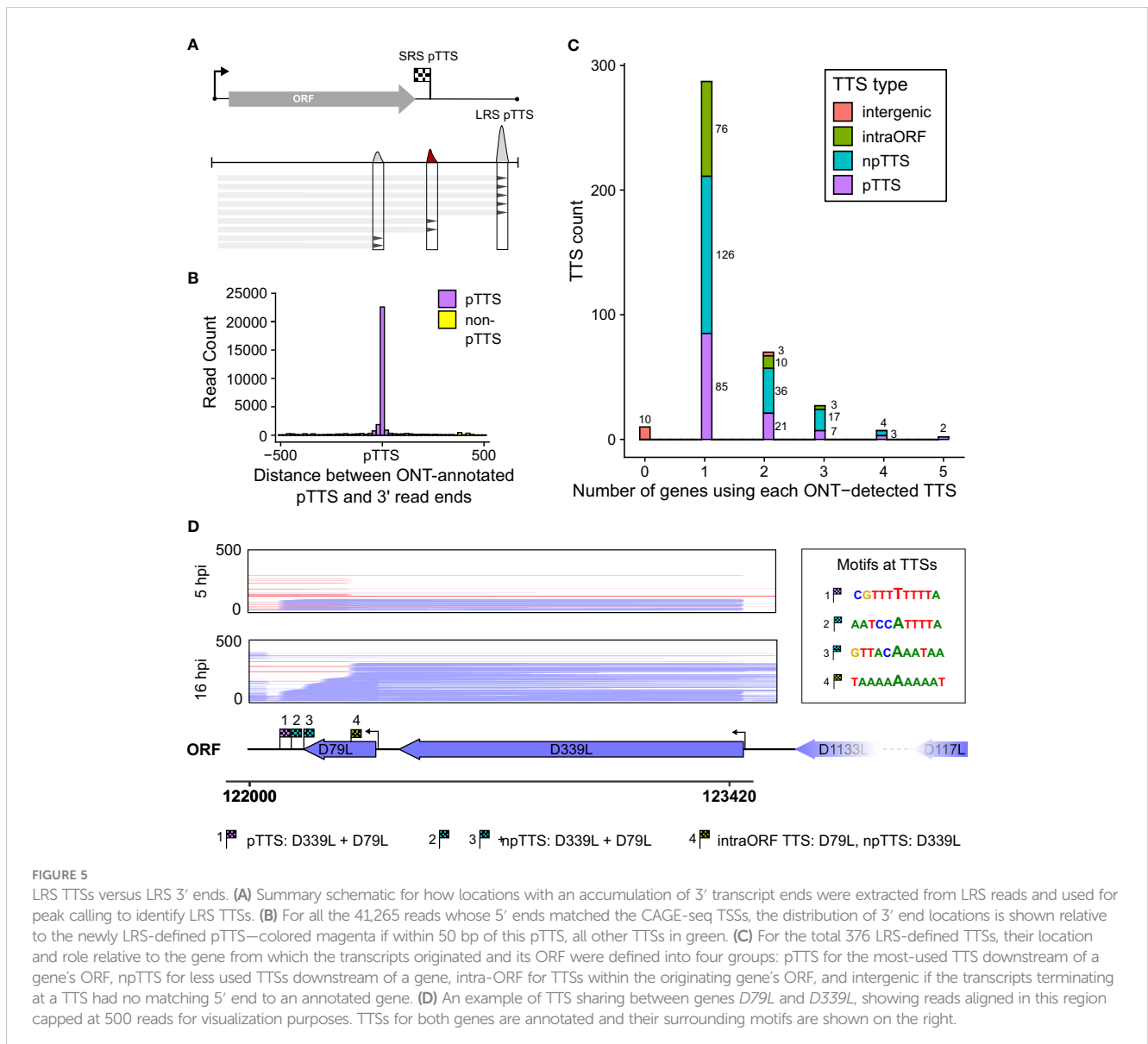
TABLE 2 Comparison between the BA71V genes classified as "late" or "early" according to their differential expression between 5 h and 16 h from CAGE-seq, as well as each gene's relationship to the next gene downstream.

Gene type	Relationship to the down-stream gene	Number of genes	Proportion of genes per gene type (%)
Early	⇨⇨ (tandem)	37	61.7
Early	⇨⇦ (converging)	23	38.3
Late	⇨⇨ (tandem)	41	49.4
Late	⇨⇦ (converging)	42	50.6
NC	⇨⇨ (tandem)	1	16.7
NC	⇨⇦ (converging)	5	83.3

There were a total of 149 genes for which we had CAGE-seq data and had a gene downstream, and a detailed list of these is shown in Supplementary Table 1.

Arrows represent relative gene orientation of each gene to its closest neighbor downstream, with head-to-head arrows illustrating converging genes (on opposing strands), and head-to-tail arrows representing genes in tandem orientation (on the same strand).

main gene users, matches them to their SRS counterparts, and describes their motifs, while Supplementary Table 3 lists their location in bed format. In good agreement with the SRS results, the most common pTTS, npTTS, and intra-ORF terminator signature was a polyT tract (Figures 7A, C, E). The second most common motif of pTTSs and npTTSs was a polyA tract (Figures 7B, D, respectively). Furthermore, we found no significant evidence of secondary structure formation to be more likely in the 50 nt of RNA upstream of the TTSs detected, after calculating their minimal folding energies (MFEs) and comparing them to the genomic background (Supplementary Figure 5). For the TTSs lacking discernable terminator sequence motifs, we examined the sequence downstream and found that the majority of non-polyT TTSs were within 100 bp of a putative polyT terminator sequence further downstream. Based on our results, we cannot rule out that transcription terminated at these downstream polyT motifs and that the observed mRNA 3′ ends were generated by cleavage or trimming, i.e., the outcome of co- or posttranscriptional endo- or
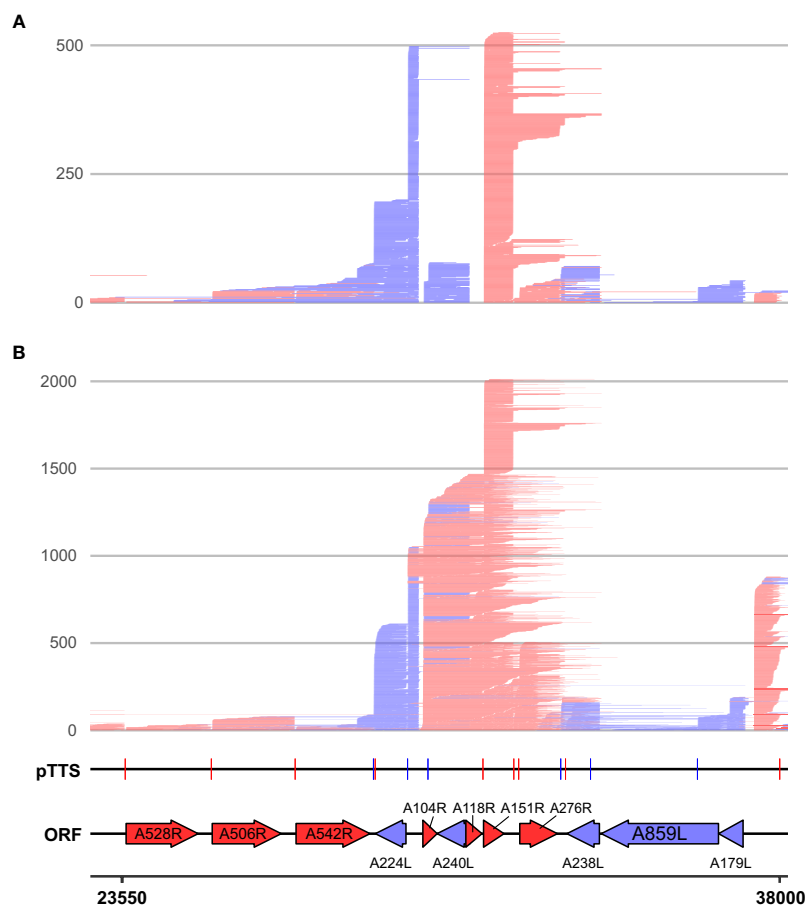

FIGURE 5
LRS TTSs versus LRS 3′ ends. **(A)** Summary schematic for how locations with an accumulation of 3′ transcript ends were extracted from LRS reads and used for peak calling to identify LRS TTSs. **(B)** For all the 41,265 reads whose 5′ ends matched the CAGE-seq TSSs, the distribution of 3′ end locations is shown relative to the newly LRS-defined pTTS—colored magenta if within 50 bp of this pTTS, all other TTSs in green. **(C)** For the total 376 LRS-defined TTSs, their location and role relative to the gene from which the transcripts originated and its ORF were defined into four groups: pTTS for the most-used TTS downstream of a gene's ORF, npTTS for less used TTSs downstream of a gene, intra-ORF for TTSs within the originating gene's ORF, and intergenic if the transcripts terminating at a TTS had no matching 5′ end to an annotated gene. **(D)** An example of TTS sharing between genes *D79L* and *D339L*, showing reads aligned in this region capped at 500 reads for visualization purposes. TTSs for both genes are annotated and their surrounding motifs are shown on the right.

FIGURE 6

LRS reads aligning between 23,550 and 38,000 on the BA71V genome at **(A)** 5 hpi and **(B)** 16 hpi. Novel LRS-annotated pTTSs and known ORFs are labeled, while all strands, ORFs, and TTSs are colored red or blue according to the strand (plus and minus, respectively). *A528R*, *A506R*, and *A542R* are also known as MGF505-7R, MGF505-8R, and MGF505-10R, respectively.

exonucleolytic events. Lastly, we compared the different TTS types with the differential gene expression of their associated mRNAs. In good agreement with our previously published CAGE-seq data, early genes had a higher proportion of polyT terminators compared with late genes (Figure 7G), while TTSs with a polyA or non-discernable motifs were more commonly associated with late genes (Figure 7H). However, the most significant and discernable motif was still the polyT across TTS types.

The ASFV core RNAP is able to recognize polyT terminators independently of termination factors

Our genome-wide analysis of transcription termination sites highlighted the importance of polyT signature motifs in ASFV. While the host RNAPII depends on termination and polyadenylation factors (68), archaeal RNAP (10, 69) and RNAPIII (12, 63, 64) are able to faithfully terminate transcription at polyT motifs, without the requirement for upstream RNA secondary structures that characterize canonical bacterial intrinsic terminators abundant in bacteria (11, 70). VACV RNAP

transcription termination depends on factors, some of which are conserved between VACV and ASFV (Table 1), despite the ASFV enzyme being structurally closer to its host RNAPII counterpart (19). To probe whether the ASFV RNAP conforms to the factor-dependent paradigm of RNAPII and VACV RNAP or is more like intrinsic termination found in RNAPIII and archaeal RNAP, we tested whether a recombinant ASFV RNAP made of the eight core subunits was able to recognize a range of terminators identified in our sequencing data. We recently reported the production of catalytically active, wholly recombinant ASFV RNAP expressed in insect cells that is suitable for a rigorous functional analysis *in vitro* (19).

Based on a protocol we previously developed for archaeal RNAP (69), we assembled transcription elongation complexes (TECs) with ASFV RNAP and a nucleic acid scaffold consisting of an RNA primer, template, and non-template DNA strand (Figure 8A). Following preincubation to allow for TEC assembly, we challenged the reaction with heparin to reduce primer-independent transcription. In the presence of NTP substrates, RNAP will extend the ^{32}P-labeled RNA primer and carry out transcription elongation independent of promoter sequences or transcription initiation factors (Figure 8B). The reaction products
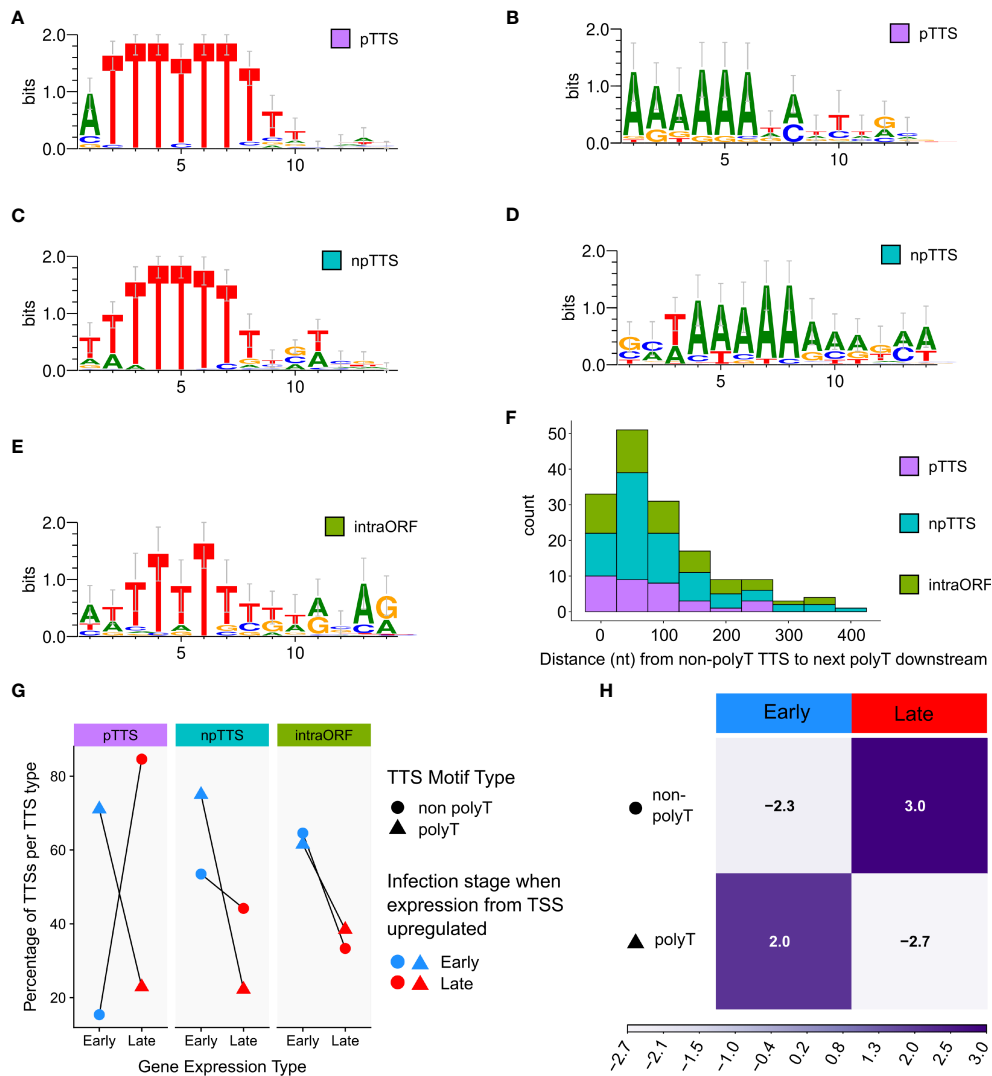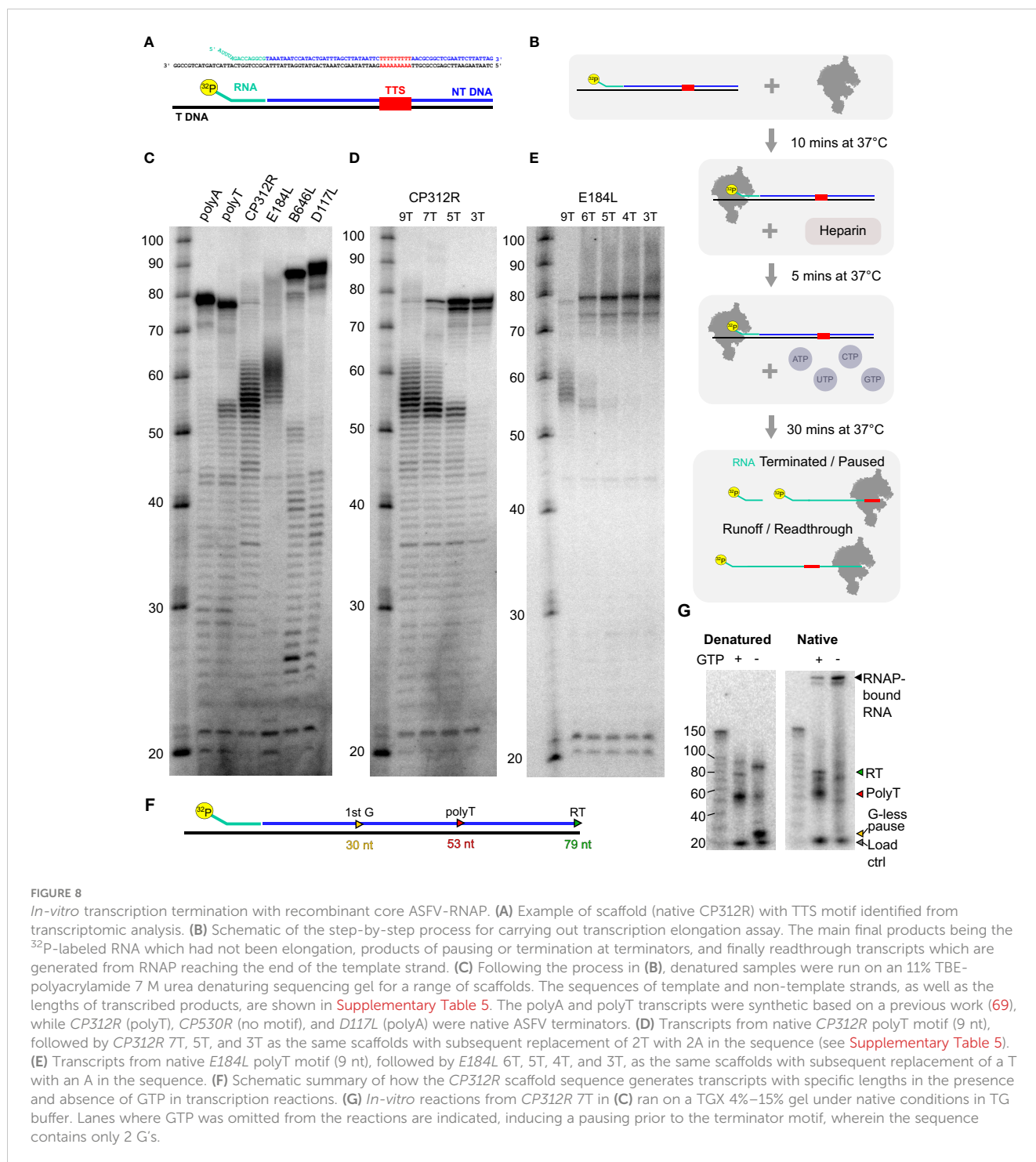
FIGURE 7

Significantly enriched DNA motifs detected via MEME, searching the 10 bp up- and downstream of the TTS, separated according to type (pTTS, npTTS, and intra-ORF), ordered according to abundance. **(A, B)** The only two significant motifs detected at 71 (E-values 3.2e−056) and 21 (3.5e−007) sites, respectively, from a total of 111 pTTSs. **(C)** The most significant motif detected from 179 npTTSs, which was found in 65 sites (E-value 8.9e −036). **(D)** The second most common motif detected among npTTSs was detected at 27 sites (E-value 9.9e−005). **(E)** This was the only significant motif found at 22 of the 87 intra-ORF TTSs (E-value 2.4e−002). WebLogo was used to create these motifs from the MEME fasta output. **(F)** The distances in nt from each of the 158 lacking any polyT TTSs to the next polyT downstream. One non-polyT TTS was omitted as it had no polyT downstream—being at the genome terminus. **(G)** A summary of TTS types according to their classification as primary, non-primary, or intra-ORF, whether their sequence contains a polyT or not and if the TSS from which their reads predominantly originate was defined as an early or late gene TSS according to previous CAGE-seq data. **(H)** Correlation matrix plot following a chi-squared test of independence, on the number of early and late gene terminators per motif category. Pearson's chi-squared test of independence: χ^2 = 24.9, *p*-value < 0.001. The scale indicates Pearson residuals, with dark purple indicating a positive association and white indicating a negative association.

are separated on denaturing polyacrylamide ("sequencing") gels to characterize RNA products at single-nucleotide resolution or on native gels to probe for the association/dissociation of the RNA from the TEC (69). We designed a range of templates encoding terminators mapped using SRS and LRS methods including those associated with the genes *CP312R* (polyT), *E184L* (polyT), *D117L* (polyA), and *B646L* (polyA). In addition, we included synthetic templates used in the archaeal termination study containing embedded 7T or 7A motifs as positive and negative controls, respectively. The corresponding read alignments for the sequences used *in vitro* are shown in Supplementary Figure 5

(*CP312R* and *E184L*) and Supplementary Figure 6 (*D117L* and *B646L*). As is the case with the archaeal RNAP, the synthetic polyT template resulted in termination along with runoff transcripts, and the polyA template exclusively produced the latter. All terminator constructs generated multiple bands, suggesting that the mRNA 3′ formation of RNAP *in vitro* can accommodate a degree of flexibility. Factor-independent (i.e., intrinsic) transcription termination is frequently accompanied by the formation of several termination products in the closely related RNAPs of archaea (71) and eukaryotic RNAPIII (12, 63, 64). In comparison, intrinsic termination by bacterial RNAP tends to be more precise, likely

FIGURE 8

In-vitro transcription termination with recombinant core ASFV-RNAP. **(A)** Example of scaffold (native CP312R) with TTS motif identified from transcriptomic analysis. **(B)** Schematic of the step-by-step process for carrying out transcription elongation assay. The main final products being the ^{32}P-labeled RNA which had not been elongation, products of pausing or termination at terminators, and finally readthrough transcripts which are generated from RNAP reaching the end of the template strand. **(C)** Following the process in **(B)**, denatured samples were run on an 11% TBE-polyacrylamide 7 M urea denaturing sequencing gel for a range of scaffolds. The sequences of template and non-template strands, as well as the lengths of transcribed products, are shown in Supplementary Table 5. The polyA and polyT transcripts were synthetic based on a previous work (69), while *CP312R* (polyT), *CP530R* (no motif), and *D117L* (polyA) were native ASFV terminators. **(D)** Transcripts from native *CP312R* polyT motif (9 nt), followed by *CP312R* 7T, 5T, and 3T as the same scaffolds with subsequent replacement of 2T with 2A in the sequence (see Supplementary Table 5). **(E)** Transcripts from native *E184L* polyT motif (9 nt), followed by *E184L* 6T, 5T, 4T, and 3T, as the same scaffolds with subsequent replacement of a T with an A in the sequence. **(F)** Schematic summary of how the *CP312R* scaffold sequence generates transcripts with specific lengths in the presence and absence of GTP in transcription reactions. **(G)** *In-vitro* reactions from *CP312R* 7T in **(C)** ran on a TGX 4%−15% gel under native conditions in TG buffer. Lanes where GTP was omitted from the reactions are indicated, inducing a pausing prior to the terminator motif, wherein the sequence contains only 2 G's.

due its dependence on RNA hairpin secondary structures that are not found in ASFV (Supplementary Figure 5) (4, 72). Alternatively, VACV-like (24) promoter-proximal RNAP slippage in ASFV could generate transcripts with varying lengths (4). The ASFV *CP312R* and *E184L* terminators, each having a stretch of nine T residues, turned out to be very efficient terminators *in vitro* without significant transcription readthrough. In stark contrast, the ASFV *D117L* and *B646L* terminators associated with polyA motifs did not lead to termination *in vitro* but only produced runoff transcripts

(Figure 8C). To test the correlation between the number of T residues and termination efficiency, we utilized mutant variants of the CP312R terminator varying the number of T's from 9 to 7, 5, and 3 (Figure 8D). Our results show a dose–response-dependent decrease in terminated transcripts and a concomitant increase in terminator readthrough RNA when decreasing the number of T residues. Almost no readthrough was observed with nine T's, and no polyT-dependent termination could be observed with three T's. To ascertain that this phenomenon is not restricted to the *CP312R*

terminator, we introduced similar variations in T content in the terminator of the *E184L* gene (Figure 8E). The results were directly comparable to *CP312R*, cementing the view that the number of T residues determines the termination efficiency of ASFV RNAP.

This indicates that the core RNAP alone is capable of terminating transcription, without any added termination factors, on polyT motif of ≥5 nt in length. Denaturing gels as in Figures 8C–E cannot discriminate between paused TECs and the outcome of *bona fide* termination, where the RNA has dissociated from the TEC. To distinguish pausing from termination, we separated released RNA from TEC by native gel electrophoresis. Both polyT-associated (red triangle) and runoff transcripts (green triangle) are present in native and denaturing conditions showing that the RNA has been released from the TEC during termination. We cannot rule out that a small fraction of RNA was dissociated subsequent to the transcription reaction during electrophoresis. To rule out this background signal, we included a control using a paused TEC generated by nucleotide limitation. The strong *CP312R* scaffold encodes a 30-nt transcript (Figure 8F) in the absence of GTP (Figure 8G, detailed layout in Supplementary Figure 8A). This stable paused TEC forms a band with low mobility on native gels, which demonstrates that the majority of the RNA remains associated with RNAP in paused complexes, with only a minor proportion of RNA released ("−GTP" in Figure 8G). Without nucleotide restriction ("+GTP"), the signal of the retained complex decreases significantly as they terminate transcription and release the RNA. The same pattern occurred throughout all ASFV scaffolds used (Supplementary Figures 7B, C).

In summary, the ASFV core RNAP can terminate transcription at polyT motifs independently of transcription termination factors, and the exact mRNA 3′ end shows some flexibility. The polyA signals associated with some late gene terminators *in vivo* cannot enable termination without additional factors *in vitro*.

Discussion

In the current study, we have applied LRS to analyze genome-wide termination in ASFV and compared it to our previous 3′ end mapping using SRS 3′ RNA-seq (4). Both studies indicated a clear-cut relationship between early genes and the use of the polyT terminator motif. Late transcription termination is also associated with polyT terminators, but also A-rich motifs, while 3′ end formation appears more variable. The mechanism by which ASFV uses the polyT terminator motif generating a polyU tract at the 3′ ends of transcripts appears to share similarities to bacteria (11), archaea (10, 69), and RNAPIII (12, 63, 64). In contrast to bacterial intrinsic terminators that are reliant on an RNA hairpin upstream of the polyU (11), we have found no evidence of stem-loop formation associated with ASFV terminators. While a polyT signal is necessary and sufficient to terminate transcription in archaea, the termination factor archaeal cleavage and polyadenylation specificity factor 1 (aCPSF1) that is homologous to the RNAPII termination factors CPSF73 assists termination in a fashion that is enhanced by the recognition of RNA polyU stretches by the KH domains of aCPSF1, which are upstream of the termination site (8, 73, 74).

Our *in-vitro* transcription assays demonstrate that the ASFV core RNAP is able to terminate transcription without the strict requirement of predicted termination factors (Table 1). A run of five consecutive T's (or U's in the RNA sequence) is sufficient for ASFV RNAP to stop transcription elongation, while the longer the motif length, the stronger the stop signal, with nine T's being sufficient to abolish transcription readthrough. Furthermore, we see evidence that the RNAP also releases transcripts following polyT terminators, suggesting ASFV RNAP is capable of intrinsic termination activity in response to this signal (Figure 8)—akin to intrinsic transcription termination in archaea (10, 69). Of course, as in archaea where polyT readthrough is also common (75), this intrinsic activity does not exclude the possibility that ASFV carries out factor-dependent transcription termination. The polyA terminator sequences were not able to terminate ASFV core RNAP *in vitro*, but it may also be the case that these sites are generated by other means, such as processing by RNases, subsequent to termination. ASFV may utilize a similar termination mechanism to archaea, whereby RNAP can intrinsically terminate at a polyT, but termination factors enhance the process in a polyT signal-dependent fashion (8, 73, 74). ASFV encodes multiple predicted transcription termination factors, though is not clear whether these enhance transcript release in an ATP-dependent manner, akin to the VACV system (see below), following pausing at a polyT terminator.

The utilization of termination factors in VACV differs between early, intermediate, and late gene transcription. During intermediate and late VACV infection, termination is thought to be facilitated by the H5 factor aka VLTF-4 (76, 77), reminiscent of activities by CPSF and the cleavage stimulatory factor (CstF), in their recognition of a polyadenylation signal (PAS) and promoting cleavage (78–80). VACV G2 and the DNA helicase A18 interact with H5 *in vivo* (81), where A18 facilitates transcript release in an ATP-dependent fashion (36, 82). Interestingly, H5 and G2 have also been reported to enhance elongation (30, 31), suggesting that the interplay between H5, G2, and A18 can shift the balance between elongation and termination. However, the mechanism and structural basis of termination by these factors remains opaque. Termination of early VACV genes is not reliant on A18, but by D11 (aka NPH-I) (83, 84) or I8 (NPH-II) (85–87), which terminate transcription immediately following pyrimidine-rich sequences, both in the absence and presence of the upstream UUUUUNU motif (23).

Like VACV, ASFV encodes NPH-I (*Q706L*) and NPH-II (*B962L*) homologs, both of which are found in viral particles suggesting their role during early transcription (3), similar to VACV (85). Recent structural studies suggest that VACV NPH-I facilitates promoter escape by an unusual upstream "DNA scrunching" mechanism (88). Importantly, the complete VACV–RNAP complex (including NPH-I) is capable of site-specific transcription termination (20, 89). In ASFV, no homolog for H5 nor G2 has been identified, though there are two ASFV homologs for the intermediate and late termination factors VACV-A18:*QP509L* and *A859L* (32–35, 90). ASFV-QP509L is the best candidate for an A18 homolog as both are ~500 residues in length and closely related at the sequence level (EMBOSS Needle (91) sequence identity: 19.1%). ASFV-A859L encodes a larger product that only partially aligns to residues 45–793 with VACV-A18, according to

similar pairwise alignment (sequence identity: 14.0%). Transcriptome analyses show that *QP509L*, *A859L*, and *Q706L* genes are all upregulated during late ASFV infection. Only VACV-A18 is present in VACV particles indicating a function for early gene transcription (92), while its ASFV homologs QP509L and A859L are not found in ASFV particles (3). Neither *QP509L* nor *A859L* is essential for the virus as either could be deleted individually in ASFV (93, 94); as the double knockout was not prepared, it is possible that QP509L and A859L are functionally redundant. *Q706L* and *QP509L* are both expressed in mid to late infection (95), but only Q706L is packaged in viral particles (3). Intriguingly, their knockdown hindered ASFV replication and disrupted late transcription, but did not affect early gene expression (96). Future work would perhaps benefit from LRS following infection with the A18 homolog deletion strains or putative termination factor knockdowns described above, to assess how the absence of these factors affects termination readthrough. It would be expected that the absence of *bona fide* transcription termination factors would lead to an increase in readthrough, while knocking out elongation factors should have the opposing effect. Similar experiments in VACV involved mutations made to factors like A18 (82), NPH-I (D11L) (97), NPH-II, or I8 (98) and provided vital evidence for their predicted roles in transcription termination.

LRS was previously used by Olasz and colleagues (61, 62) to investigate RNA extracted from ASFV-infected porcine macrophages. The study included a comparison of LRS to SRS data obtained by traditional RNA-seq and the analysis of ASFV transcript isoforms with their 5′ and 3′ ends (62). Their results confirmed the commonly reported diversity in viral transcript lengths [as summarized previously (90)] and importantly showed that long transcripts included multiple consecutive ORFs. The relatively low sequencing depth of this study limited its reach, and the pooling of samples precluded an analysis of differential gene expression or varying read lengths throughout the ASFV infection time course. As temporal gene expression is key to understanding ASFV biology, we sought to fill this knowledge gap with the current study, which demonstrates significant changes in the 3′ end landscape between early and late infection.

One of the remaining key questions is how the non-polyT TTSs we detected factor into the ASFV transcriptomic landscape. The 3′ RNA-seq initially failed to identify many late gene TTSs, compared with those from early genes. LRS indicated that late gene transcript 3′ ends are indeed enriched at polyT motifs, but also at a polyA (Figure 7), which would have been filtered out from our 3′ RNA-seq data to remove potential mispriming (4). RNA 3′ end enrichment at polyA motifs is clearly more common during late infection and for late-classified genes (Figure 7G), though the reason for this remains enigmatic. Due to the AT-rich nature of the ASFV genome, polyT sequences are widespread, ensuring most of the non-polyT TTSs we detected were close to polyT sequences downstream (Figure 7F). Future work should investigate how 3′ end formation occurs among those non-canonical non-polyT TTSs, which become more frequent during late infection. It will be also important to investigate whether they are indeed *bona fide* termination sites or arising through some other mechanism such as 3′ end processing. Transcription terminator readthrough of convergent genes can produce mRNA

3′ regions that are complementary to each other and thus form dsRNAs, which in turn can trigger antiviral responses (99, 100) including the interferon response, RNA interference by Dicer (dsRNA targeting) (101, 102), and activation of RNase L (ssRNA cleavage) (103). The latter acts in concert with oligoadenylate synthetase gene 1 to inhibit ASFV replication (104).

While the role and extent of RNA 3′ processing of ASFV mRNAs remains unclear, it is possible and even likely that some of the mapped RNA 3′ ends are not "nascent" termination products, but generated by processing, and this applies particularly to late infection. Figure 9A summarizes the broad patterns of transcription termination in ASFV. While this accounts for intrinsic factor-independent termination as demonstrated *in vitro* (Figure 8), ASFV encodes several termination factor candidates (Table 1) that contribute to the formation of the viral transcription termination landscape. Those found in ASFV particles (3) are likely utilized during the early stages of infection (Figure 9B). However, their mechanisms of action remain poorly understood. Likewise, the role of other putative termination factors or as-yet undiscovered candidates remains opaque (Figure 9C).

A vital tool in the production of attenuated ASFV strains is a detailed gene expression landscape, playing a key role in vaccine development. Compromising ASFV pathogenicity, attenuating the virus can be achieved by many different means, and the molecular machineries responsible for viral gene expression are key targets genes in this context—as well as for the screening and development of specific inhibitors with the potential to act as antiviral drugs.

Summary

- The transcription termination landscape of ASFV is highly complex.
- Termination sites can be associated with either polyT or polyA DNA motifs or not be associated with any sequence signatures.
- PolyT terminators of different lengths are the dominant termination signals, with a larger number of T residues resulting in more efficient termination and less transcription readthrough.
- PolyT motifs become less common during late infection.
- ASFV polyT terminators are not associated with RNA hairpin secondary structures.
- PolyT terminators are recognized directly by the eight-subunit ASFV core RNAP independently of termination factors.
- ASFV polyT terminator architecture is evolutionary conserved with archaeal RNAP and eukaryotic RNAPIII systems.
- mRNA 3′ complexity increases greatly during late infection due to variations in TTS usage and possibly due to increased RNA processing and degradation.
- Transcription readthrough is a prominent feature of ASFV terminators and results in polycistronic mRNAs, although
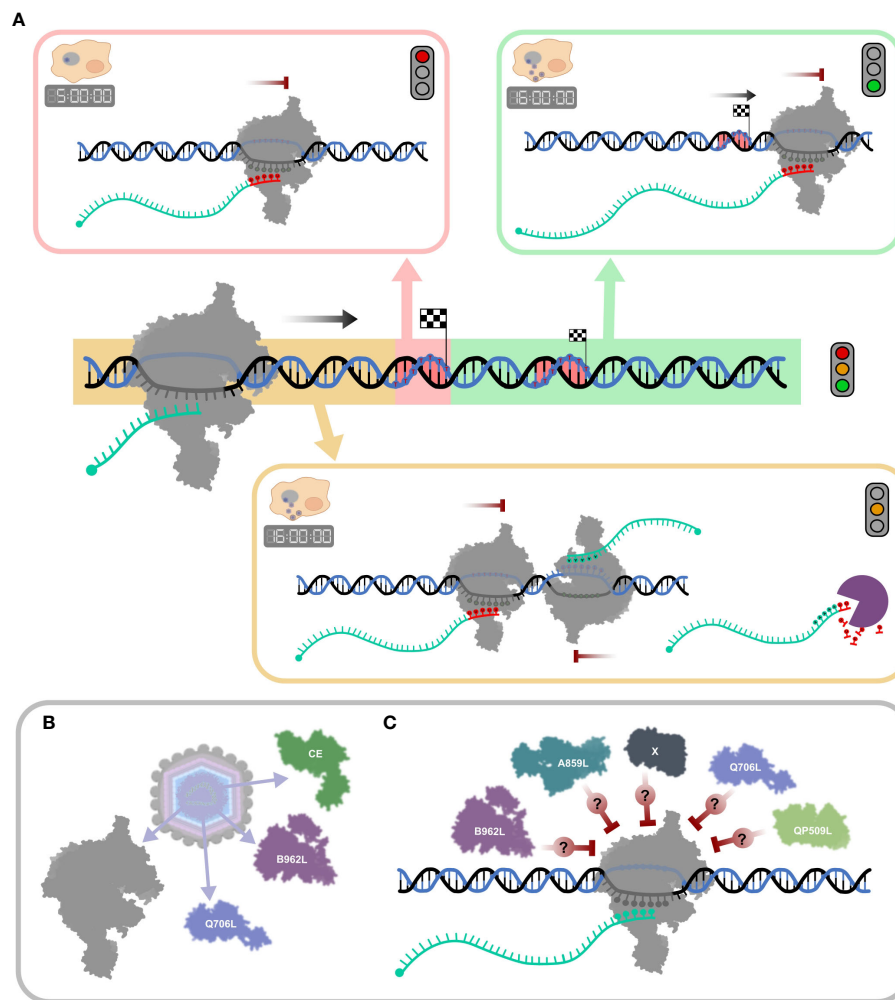
FIGURE 9

Schematic summary of ASFV transcription termination and putative mechanisms of RNA 3′ end formation. **(A)** The mechanisms for "correct" termination (red highlight), premature termination (yellow highlight), and terminator readthrough (green highlight) are illustrated in boxes. During the early stages of infection (5 hpi), termination is dominated by concise mRNA 3′ formation associated with strong polyU stretches at RNA 3′ ends (red nt). A more complex mRNA 3′ end landscape can be observed in late infection (16 hpi), alongside concise termination, abundant terminator readthrough (green), and premature 3′ end formation (yellow), many of which are not associated with polyT motifs. Transcripts appearing as prematurely terminated include mRNA 3′ ends generated by *bona fide* termination, head-on collisions of RNAPs which transcribe convergent gene pairs, or alternatively by mRNA degradation or processing. **(B)** ASFV particles include at least two termination factor candidates, Q706L and B962L, and the CE, important for termination in VACV (described in Table 1). **(C)** The ASFV genome encodes additional putative termination factors including A859L and QP509L, but their molecular mechanisms and exact roles during termination are still not well understood (96).

it remains uncertain to which extent downstream ORFs are translated.

- The molecular mechanisms of non-polyT-facilitated termination are unknown but may involve termination factors, head-on collisions with RNAP transcription elongation complexes (TECs) on convergent genes, collisions with replication forks, and other DNA-binding factors such as chromatin proteins (e.g., A104R).

Future directions

The recombinant ASFV RNAP and *in-vitro* transcription assays will now empower us to characterize the molecular mechanisms of transcription termination in ASFV under rigorous conditions. The combination of functional genomics studies, like the one reported here, and *in-vitro* experiments will enable a comprehensive yet detailed understanding of transcription termination in ASFV. The ensuing burning research questions include the following:

- What is the structure and function of the predicted ASFV termination factors (Table 1)?
- Which genes are terminated by which termination factors during early and late infection?
- How does the depletion of termination factors alter the mRNA 3′ patterns *in vivo*? Does 3′ end formation differ between ASFV strains?

- How are ASFV termination factors and their mechanisms related to other NCLDV viruses and to cellular RNAP transcription systems?
- Does chromatinization by DNA-binding factors modulate transcription initiation and termination of ASFV genes?
- Are downstream ORFs in multicistronic mRNAs translated in ASFV-infected cells?

Materials and methods

Long-read sequencing of ASFV-BA71V: Oxford Nanopore MinION library preparation and sequencing

A total of four RNA samples were extracted under the same conditions as for previous BA71V transcriptomic work (4): two replicates from 5 h and 16 h post-infection. RNA sequencing libraries were prepared with the Direct RNA Sequencing Kit (Oxford Nanopore or "ONT," SQK-RNA002) according to the manufacturer's instructions. In brief, any polyadenylated RNA present was annealed to an oligo-dT primer along with an adapter to facilitate reverse transcription (RT) from native 3′ ends of transcripts with SuperScript IV (Thermo Fisher, Waltham, USA). Magnetic beads (Agencourt AMPure XP, Beckman Coulter, Brea, USA) were mixed with 1 µl of RNasin Ribonuclease Inhibitor (Promega, Madison, USA) per 100 µl beads. Following RT, the sample was mixed with this bead–inhibitor mix, and using a magnetic rack, the beads were cleaned with 70% EtOH. After elution from the beads in nuclease-free water, a sequencing adapter was ligated to the RNA–DNA hybrid, bound to fresh magnetic beads, and washed with the kit's wash buffer and magnetic rack, before eluting in its elution buffer. Samples were prepared for loading onto a MinION flow cell (FLO-MINSP6) according to the manufacturer's instructions. Samples were sequenced until there were no remaining pores in the flow cell (1 flow cell per sample, taking up to 72 h), and the results were output in FAST5 format.

Basecalling and mapping of nanopore sequencing reads

Guppy (v4.4.2, Oxford Nanopore) was used for basecalling, i.e., converting the FAST5 voltage signal files into FASTQ sequencing data files (parameters: –flowcell FLO-MIN106 –kit SQK-RNA002 –trim_strategy none –fast5_out –reverse_sequence on –calib_detect on -r). Output FASTQ files were then concatenated to generate FASTQ files for each sample, containing both Vero host and ASFV-BA71V reads. Minimap2 (105) was used to map reads from FASTQ files to either the ASFV-BA71V (U18466.2) or Vero (GCF_000409795.2 *Chlorocebus sabaeus* 1.1) genomes, after generating index.mmi files for each genome to improve mapping speed. Mapping command: minimap2 -ax splice -uf -d -a [genome file].mmi [sample file].fastq > [sample file].sam. For visualization, SAM files were converted to BAM files using Samtools (106), which was also used to extract lengths of reads mapping to each genome.

Genome-wide comparison to LRS 5′ and 3′ ends to 5′ CAGE-seq and 3′ RNA-seq

BAM files were sorted and indexed using IGV Tools (107) before BEDTools (108) was used to return LRS reads which overlapped our newly annotated BA71V genome in GFF3 format [from Cackett et al. (4)], using the command: bedtools intersect -wo -s -a [.gff3 file] -b [.bam file]. Each resulting table was imported into RStudio (109) (Version 1.1.456, R version: 3.6.3) in which most subsequent analysis and data visualization took place, predominantly using the packages dplyr (110) and ggplot2 (111). At this stage, results from replicates were pooled into two groups: ASFV reads overlapping genes at 5 h and reads overlapping genes at 16 h. For each read that overlapped an annotated gene on the same strand, these were filtered according to firstly whether the 5′ end of the read was within 100 nt of the CAGE-seq-annotated TSS. Then, reads were filtered according to where the 3′ ends of reads were located, relative to the 3′ RNA-seq-annotated pTTS (4): at the pTTS (classed as "correct" termination), upstream of the pTTS ("premature"), and downstream of the pTTS ("readthrough"). For all the reads which mapped close to the 5′ ends, all reads would be assigned to either of these three categories, with their frequencies per gene being compared between early and late genes (assigned from CAGE-seq), between time points overall, or between timepoints when the pTTS contained a polyT sequence motif, and according to the polyT length (number of consecutive T's). All this was saved as a data.frame in RStudio for downstream analysis.

Relationship between polyT presence and termination types at TTSs

Genome-wide polyT occurrences were identified via searching for any "TTTT" motif across each strand of the ASFV-BA71V genome using IGV Tools "Find Motif" function, and polyT locations were exported in BED format. BED files were then sorted (command: sort -k1,1 -k2,2n [.bed] > [sorted.bed]). Sorted BED file coordinates were merged with BEDTools, to combine consecutive stretches of >4 T's into a single annotation in the BED file for each strand (command: bedtools merge -i [sorted.bed] > [merge.bed]). BAM alignments were converted to BW format via deepTools (112), with one file for each strand (command: bamCoverage –bam [.bam file] –outFileName [.bw file] –outFileFormat bigwig –binSize 1 –filterRNAstrand [forward/reverse] –normalizeUsing CPM).

Visualization of aligned LRS reads

Alignments shown in Figures 1, 3, 5, 6, as well as in Supplementary Figures 1, 6, 7, were all generated with R in RStudio after deepTools (112) was used to separately pool the BA71V-aligned reads for the 5-h and 16-h time points and convert them from BAM to BED format before conversion to GFF. Each of these GFF files was imported into RStudio using import.gff3() from

the package rtracklayer (113). Annotations such as TTS or polyT locations were imported in six-column BED format using import.bed(). Alignment figures were generated via the packages ggplot, ggbio, and rtracklayer, and the gggenes package (65) was used to generate the gene maps beneath each alignment.

Genome-wide transcription termination patterns

The layout of ASFV genes initially summarized in Figure 4 was manually annotated for the 153 ASFV-BA71V genes, according to each gene's relationship to its closest neighboring genes. Genes whose neighbors shared the same strand and therefore direction were classed as "contiguous," "clashing" if its 3′ end was directed toward the 5′ end of its neighbor, and "diverging" if two genes' 5′ ends began together and were directed away from one another on opposing strands. Due to the compact nature of the BA71V genome and some genes overlapping, these categorizations were not mutually exclusive. Genes were defined as only either clashing or contiguous (diverging genes could be either), only according to the direction of the gene downstream. For reads that matched their 5′ ends within 100 nt of TSSs of annotated genes, a BED file only containing the last nt at the 3′ end was extracted. BEDTools slop was used to expand the 3′ end nt location of each read plus and minus 20 nt on either side. BEDTools getfasta was then used to extract the genomic sequences within these regions. Fasta files were filtered for duplicate sequences using sRNAtoolbox (114) to reduce bias from highly expressed genes—with many identical terminator motifs. The filtered sequences were then searched for enriched motifs using the MEME Suite (115) (searching for three motifs 5–20 nt in length, in zoops mode).

Defining TTSs from LRS and enriched predicting motifs

To annotate TTSs *de novo* using the LRS data, each sample coverage BAM file was first converted into a BW file via deepTools (112) with only coverage for the last 3′ end nt of each read. TTS prediction was carried out using these BW files as input, using the CAGEfightR (116) package in R, as carried out for our 3′ RNA-seq TTS prediction (4). The 376 TTSs found via CAGEfightR for peak calling were annotated according to their position relative to the closest ORF using BEDTools closest -s and manually, if applicable. There were four different "TTS types": firstly, the vast majority of TTSs were defined as primary or non-primary TTSs (pTTS or npTTS), pTTS meaning that the highest number of reads originating from a particular gene's TSS would have their 3′ ends located at this specific site. Secondly, if there were further enriched sites used by genes, which were utilized less than its pTTS, these were defined as npTTSs. The designation of pTTS or npTTS was done via calculating the percentage of each gene's reads (5′ ends matching) terminating at a particular TTS (3′ ends within 100 nt) and manually checked from alignments in IGV. Thirdly, if after matching to a gene's TSS, the 3′ end was found within that gene's ORF, this was described as "intra-ORF." Lastly, for reads whose 5′ ends did not match to any annotated gene, these were called "intergenic."

These different annotations also confirmed that often TTSs could be used by multiple genes—defined as reads whose 3′ ends were within 100 nt of the same TTS but whose 5′ ends were within 100 nt of TSSs for different genes.

After annotating the LRS TTSs with their originating TSSs and defining each as pTTS (111 TTSs), 179 npTTS, or 87 intra-ORF-TTS, these TTS sequences were searched for enriched motifs. This was carried out as described above, except with the region 10 bp up- and downstream of each TTS, rather than the 3′ ends of the reads. The MEME suite was also similarly used to find three motifs for each TTS type (in zoops mode, 5–21 nt in length). All motifs were made using the sequence output from MEME, listing all the sequences contributing to each motif, converted to .fasta format, and input into WebLogo 3 (117) (www.weblogo.threeplusone.com/create.cgi). According to the appearance of each TTS motif, they were defined as a "polyT" or "non_polyT." These TTS types were compared in RStudio using mainly dplyr (110) to the expression of the main gene user of each TTS, i.e., which gene matched its TSS to the highest proportion of 5′ read ends, whose 3′ ends then matched to that TTS. The location of each LRS TTS along with details of its matched gene and expression and their surrounding sequences are listed in Supplementary Table 2. Supplementary Table 3 contains a summary of all LRS TTS locations across the BA71V genome, including in BED file format, named according to the most common gene users of each site.

Prediction of minimal folding energy

RNAfold (118) was used to predict RNA minimal folding energy of 50 nt upstream of all the annotated TTSs including the TTS position. BEDTools random was used to extract 10,000 genomic sequences of the same length to serve as a background. The values in kcal/mol are listed in Supplementary Table 4.

Scaffolds for *in-vitro* transcription elongation assays

The RNA (RNA14 from Hirtreiter et al. (69), sequence: AUUUAGACCAGGCG) was ordered from GenScript, Piscataway, USA, and 10 µM of RNA14 was ^{32}P-labeled with [γ-^{32}P] ATP (Hartmann Analytic, Braunschweig, Germany), 1 µl of PNK, and 5 µl of PNK buffer (M0201S, NEB, Ipswich, USA), with 0.5 µl of RNasin Plus Ribonuclease Inhibitor (Promega, Madison, USA) before making up the volume to 50 µl with RNase/DNase-free H_2O. After incubation for 1.5 h at 37°C, free [γ-32P] ATP was removed from the reactions via MicroSpin G-25 desalting columns (Cytiva, Marlborough, USA) as per the manufacturer's instructions. The sequences for synthetic polyA and polyT template (T) sequences were from Hirtreiter et al. (69) Their corresponding non-template (NT) sequences were designed to anneal directly downstream of the annealed RNA, and this design was used for all RNA:dsDNA scaffolds. DNA oligos used for generating scaffolds used in transcription assays were ordered from Integrated DNA Technologies, Coralville, USA and Merck Life Science, Darmstadt,

Germany (Supplementary Table 5). The DNA and RNA were annealed in a ratio of 10 µM:2.5 µM:2.5 µM of RNA:T:NT strands, respectively, in reaction volumes of 50 µl, with 1 µl of RNasin Plus Ribonuclease Inhibitor, 2 µl of 25× annealing buffer (250 mM of Tris–HCl pH 7.5, 1.25 M of NaCl), and the remaining volume made up with water. Annealing was carried out at 96°C for 1 min before 2 min at room temperature and then placed on ice before use in assays.

In-vitro transcription elongation using recombinant ASFV-RNAP

Purified recombinant ASFV RNAP (60 nM) [produced as previously described (19)] was preincubated for 10 min at 37°C, with 0.5 µl of RNasin Plus Ribonuclease Inhibitor (Promega, Madison, USA) and 5.6 µl of the RNA:dsDNA scaffold prepared as above. Heparin solution was added to a final concentration of 1 µg/ml per reaction, followed by 5 min incubation at 37°C. Reactions were started by adding 2.6 µl of a master mix to generate final concentrations per reaction of 300 µM for each ATP, GTP, CTP, and UTP, as well as 25 mM of Tris–HCl pH 8, 3 mM of $MgCl_2$, 50 mM of KCl, 7 mM of DTT, and 2.5 mg/ml of BSA. Reactions were then incubated for 30 min at 37°C and stopped either via loading directly onto a native gel (4%–20% TGX Bio-Rad, Hercules, USA) run in TG buffer or by the addition of denaturing Gel Loading Buffer II (Thermo Fisher Scientific) and boiling for 5 min at 95°C before loading onto a denaturing 11% TBE-polyacrylamide 7 M urea gel. The ladder used for polyacrylamide gel electrophoresis was the Decade Markers System (AM7778, Thermo Fisher Scientific).

Data availability statement

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found below: https://www.ncbi.nlm.nih.gov/sra/, PRJNA1045388.

Author contributions

GC: Conceptualization, Data curation, Formal analysis, Funding acquisition, Investigation, Methodology, Visualization, Writing – original draft, Writing – review & editing. MS: Methodology, Validation, Writing – review & editing. RP: Writing – review & editing, Methodology, Validation. CD: Writing – review & editing, Methodology. LD: Funding acquisition, Project administration, Supervision, Writing – review & editing. FW: Conceptualization, Funding acquisition, Project administration, Supervision, Writing – review & editing.

Funding

Acknowledgments

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fimmu.2024.1350267/full#supplementary-material

References

1. Werner F, Grohmann D. Evolution of multisubunit RNA polymerases in the three domains of life. *Nat Rev Microbiol* (2011) 9:85–98. doi: 10.1038/nrmicro2507

2. Salas ML, Kuznar J, Vinuela E. Polyadenylation, methylation, and capping of the RNA synthesized *in vitro* by African swine fever virus. *Virology* (1981) 113:484–91. doi: 10.1016/0042-6822(81)90176-8

3. Alejo A, Matamoros T, Guerra M, Andres G. A proteomic atlas of the African swine fever virus particle. *J Virol* (2018) 92:JVI.01293–18. doi: 10.1128/JVI.01293-18

4. Cackett G, Matelska D, Sýkora M, Portugal R, Malecki M, Bahler J, et al. The African swine fever virus transcriptome. *J Virol* (2020) 94:e00119-20. doi: 10.1128/JVI.00119-20.

5. Nemeroff ME, Barabino SM, Li Y, Keller W, Krug RM. Influenza virus NS1 protein interacts with the cellular 30 kDa subunit of CPSF and inhibits 3′end formation of cellular pre-mRNAs. *Mol Cell* (1998) 1:991–1000. doi: 10.1016/S1097-2765(00)80099-4.

6. Grosso AR, Leite AP, Carvalho S, Matos MR, Martins FB, Vitor AC, et al. Pervasive transcription read-through promotes aberrant expression of oncogenes and RNA chimeras in renal carcinoma. *Elife* (2015) 4:e09214. doi: 10.7554/eLife.09214.

7. Hadar S, Meller A, Saida N, Shalgi R. Stress-induced transcriptional readthrough into neighboring genes is linked to intron retention. *iScience* (2022) 25:105543. doi: 10.1016/j.isci.2022.105543.

8. Yue L, Li J, Zhang B, Qi L, Li Z, Zhao F, et al. The conserved ribonuclease aCPSF1 triggers genome-wide transcription termination of Archaea via a 3′-end cleavage mode. *Nucleic Acids Res* (2020) 48:9589–605. doi: 10.1093/nar/gkaa702.

9. Peters JM, Mooney RA, Grass JA, Jessen ED, Tran F, Landick R. Rho and NusG suppress pervasive antisense transcription in Escherichia coli. *Genes Dev* (2012) 26:2621–33. doi: 10.1101/gad.196741.112.

10. Santangelo TJ, Cubonova L, Skinner KM, Reeve JN. Archaeal intrinsic transcription termination *in vivo*. *J Bacteriol* (2009) 191:7102–8. doi: 10.1128/JB.00982-09.

11. Yarnell WS, Roberts JW. Mechanism of intrinsic transcription termination and antitermination. *Science* (1999) 284:611–5. doi: 10.1126/science.284.5414.611.

12. Nielsen S, Yuzenkova Y, Zenkin N. Mechanism of eukaryotic RNA polymerase III transcription termination. *Science* (2013) 340:1577–80. doi: 10.1126/science.1237934.

13. Schmidt MC, Chamberlin MJ. nusA Protein of Escherichia coli is an efficient transcription termination factor for certain terminator sites. *J Mol Biol* (1987) 195:809–18. doi: 10.1016/0022-2836(87)90486-4.

14. Guglielmini J, Woo AC, Krupovic M, Forterre P, Gaia M. Diversification of giant and large eukaryotic dsDNA viruses predated the origin of modern eukaryotes. *Proc Natl Acad Sci USA* (2019) 116:19585–92. doi: 10.1073/pnas.1912006116.

15. Mason SW, Wallisch M, Grummt I. RNA polymerase I transcription termination: similar mechanisms are employed by yeast and mammals. *J Mol Biol* (1997) 268:229–34. doi: 10.1006/jmbi.1997.0976.

16. Lang WH, Morrow BE, Ju Q, Warner JR, Reeder RH. A model for transcription termination by RNA polymerase I. *Cell* (1994) 79:527–34. doi: 10.1016/0092-8674(94)90261-5.

17. Lang WH, Reeder RH. Transcription termination of RNA polymerase I due to a T-rich element interacting with Reb1p. *Proc Natl Acad Sci USA* (1995) 92:9781–5. doi: 10.1073/pnas.92.21.9781.

18. Lang WH, Reeder RH. The REB1 site is an essential component of a terminator for RNA polymerase I in Saccharomyces cerevisiae. *Mol Cell Biol* (1993) 13:649–58. doi: 10.1128/MCB.13.1.649.

19. Pilotto S, Sýkora M, Cackett G, Dulson C, Werner F. Structure of the recombinant RNA polymerase from African Swine Fever Virus. *Res Square [Preprint]* (2023). doi: 10.21203/rs.3.rs-3254640/v1

20. Grimm C, Hillen HS, Bedenk K, Bartuli J, Neyer S, Zhang Q, et al. Structural basis of poxvirus transcription: vaccinia RNA polymerase complexes. *Cell* (2019) 179:1537–1550.e19. doi: 10.1016/j.cell.2019.11.024

21. Vilborg A, Steitz JA. Readthrough transcription: How are DoGs made and what do they do? *RNA Biol* (2017) 14:632. doi: 10.1080/15476286.2016.1149680.

22. Beaudoing E, Freier S, Wyatt JR, Claverie JM, Gautheret D. Patterns of variant polyadenylation signal usage in human genes. *Genome Res* (2000) 10:1001–10. doi: 10.1101/gr.10.7.1001.

23. Yang Z, Bruno DP, Martens CA, Porcella SF, Moss B. Genome-wide analysis of the 5′ and 3′ ends of vaccinia virus early mRNAs delineates regulatory sequences of annotated and anomalous transcripts. *J Virol* (2011) 85:5897–909. doi: 10.1128/JVI.00428-11.

24. Yang Z, Martens CA, Bruno DP, Porcella SF, Moss B. Pervasive initiation and 3′-end formation of poxvirus postreplicative RNAs. *J Biol Chem* (2012) 287:31050–60. doi: 10.1074/jbc.M112.390054.

25. Chung CS, Chen CH, Ho MY, Huang CY, Liao CL, Chang W. Vaccinia virus proteome: identification of proteins in vaccinia virus intracellular mature virion particles. *J Virol* (2006) 80:2127–40. doi: 10.1128/JVI.80.5.2127-2140.2006.

26. Pena L, Yàñez RJ, Revilla Y, Viñuela E, Salas ML. African swine fever virus guanylyltransferase. *Virology* (1993) 193:319–28. doi: 10.1006/viro.1993.1128.

27. Shuman S, Broyles SS, Moss B. Purification and characterization of a transcription termination factor from vaccinia virions. *J Biol Chem* (1987) 262:12372–80. doi: 10.1016/S0021-9258(18)45363-X.

28. Yanez RJ, Rodriguez JM, Boursnell M, Rodriguez JF, Vinuela E. Two putative African swine fever virus helicases similar to yeast 'DEAH' pre-mRNA processing proteins and vaccinia virus ATPases D11L and D6R. *Gene* (1993) 134:161–74. doi: 10.1016/0378-1119(93)90090-p

29. Deng L, Shuman S, Vaccinia NPH-I. a DExH-box ATPase, is the energy coupling factor for mRNA transcription termination. *Genes Dev* (1998) 12:538–46. doi: 10.1101/gad.12.4.538.

30. Kovacs GR, Moss B. The vaccinia virus H5R gene encodes late gene transcription factor 4: purification, cloning, and overexpression. *J Virol* (1996) 70:6796–802. doi: 10.1128/jvi.70.10.6796-6802.1996.

31. Black EP, Condit RC. Phenotypic characterization of mutants in vaccinia virus gene G2R, a putative transcription elongation factor. *J Virol* (1996) 70:47–54. doi: 10.1128/jvi.70.1.47-54.1996.

32. Yáñez RJ, Rodriguez JM, Nogal ML, Yuste L, Enriquez C, Rodriguez JF, et al. Analysis of the complete nucleotide sequence of African swine fever virus. *Virology* (1995) 208:249–78. doi: 10.1006/viro.1995.1149.

33. Baylis SA, Twigg SRF, Vydelingum S, Dixon LK, Smith GL. Three African swine fever virus genes encoding proteins with homology to putative helicases of vaccinia virus. *J Gen Virol* (1993) 74:1969–74. doi: 10.1099/0022-1317-74-9-1969.

34. Yutin N, Wolf YI, Raoult D, Koonin EV. Eukaryotic large nucleo-cytoplasmic DNA viruses: clusters of orthologous genes and reconstruction of viral genome evolution. *Virol J* (2009) 6:223. doi: 10.1186/1743-422X-6-223.

35. Iyer LM, Aravind L, Koonin EV. Common origin of four diverse families of large eukaryotic DNA viruses. *J Virol* (2001) 75:11720–34. doi: 10.1128/JVI.75.23.11720-11734.2001.

36. Lackner CA, Condit RC. Vaccinia virus gene A18R DNA helicase is a transcript release factor. *J Biol Chem* (2000) 275:1485–94. doi: 10.1074/jbc.275.2.1485.

37. Koonin EV, Senkevich TG. Vaccinia virus encodes four putative DNA and/or RNA helicases distantly related to each other. *J Gen Virol* (1992) 73:989–93. doi: 10.1099/0022-1317-73-4-989.

38. Shuman S. Vaccinia virus RNA helicase: An essential enzyme related to the DE-H family of RNA-dependent NTPases. *Proc Natl Acad Sci USA* (1992) 89:10935–9. doi: 10.1073/pnas.89.22.10935.

39. Almazán F, Rodríguez JM, Angulo A, Viñuela E, Rodriguez JF. Transcriptional mapping of a late gene coding for the p12 attachment protein of African swine fever virus. *J Virol* (1993) 67:553–6. doi: 10.1128/jvi.67.1.553-556.1993.

40. Almazán F, Rodriguez JM, Andres G, Perez R, Viñuela E, Rodriguez JF. Transcriptional analysis of multigene family 110 of African swine fever virus. *J Virol* (1992) 66:6655–67. doi: 10.1128/jvi.66.11.6655-6667.1992.

41. Rodriguez JM, Yáñez RJ, Pan R, Rodriguez JF, Salas ML, Viñuela E. Multigene families in African swine fever virus: family 505. *J Virol* (1994) 68:2746–51. doi: 10.1128/jvi.68.4.2746-2751.1994.

42. González A, Calvo V, Almazán F, Almendral JM, Ramirez JC, de la Vega I, et al. Multigene families in African swine fever virus: family 360. *J Virol* (1990) 64:2073–81. doi: 10.1128/JVI.64.5.2073-2081.1990

43. Almendral JM, Almazán F, Blasco R, Viñuela E. Multigene families in African swine fever virus: family 110. *J Virol* (1990) 64:2064–72. doi: 10.1128/jvi.64.5.2064-2072.1990.

44. Simón-Mateo C, Freije JMP, Andrés G, López-Otin C, Viñuela E. Mapping and sequence of the gene encoding protein pl7, a major African swine fever virus structural protein. *Virology* (1995) 206:1140–4. doi: 10.1006/viro.1995.1039.

45. Rodríguez JM, Salas ML, Viñuela E. Intermediate class of mRNAs in African swine fever virus. *J Virol* (1996) 70:8584–9. doi: 10.1128/jvi.70.12.8584-8589.1996.

46. Rodriguez JM, Salas ML, Viñuela E. Genes homologous to ubiquitin-conjugating proteins and eukaryotic transcription factor SII in African swine fever virus. *Virology* (1992) 186:40–52. doi: 10.1016/0042-6822(92)90059-X.

47. Kuznar J, Salas ML, Viñuela E. DNA-dependent RNA polymerase in African swine fever virus. *Virology* (1980) 101:169–75. doi: 10.1016/0042-6822(80)90493-6.

48. Rodríguez JM, Yáñez RJ, Rodríguez JF, Viñuela E, Salas ML. The DNA polymerase-encoding gene of African swine fever virus: sequence and transcriptional analysis. *Gene* (1993) 136:103–10. doi: 10.1016/0378-1119(93)90453-A.

49. Galindo I, Viñuela E, Carrascosa AL. Characterization of the african swine fever virus protein p49: a new late structural polypeptide. *J Gen Virol* (2000) 81:59–65. doi: 10.1099/0022-1317-81-1-59.

50. Portugal R, Leitão A, Martins C. Characterization of African swine fever virus IAP homologue expression in porcine macrophages infected with different virulence isolates. *Vet Microbiol* (2009) 139:140–6. doi: 10.1016/j.vetmic.2009.04.030.

51. Rodriguez F, Alcaraz C, Eiras A, Yáñez RJ, Rodriguez JM, Alonso C, et al. Characterization and molecular basis of heterogeneity of the African swine fever virus envelope protein p54. *J Virol* (1994) 68:7244–52. doi: 10.1128/jvi.68.11.7244-7252.1994.

52. Alejo A, Yáñez RJ, Rodríguez JM, Viñuela E, Salas ML. African swine fever virus trans-prenyltransferase. *J Biol Chem* (1997) 272:9417–23. doi: 10.1074/jbc.272.14.9417.

53. Yáñez RJ, Rodriguez JM, Rodriguez JF, Salas ML, Viñuela E. African swine fever virus thymidylate kinase gene: Sequence and transcriptional mapping. *J Gen Virol* (1993) 74:1633–8. doi: 10.1099/0022-1317-74-8-1633.

54. Goatley LC, Twigg SR, Miskin JE, Monaghan P, St-Arnaud R, Smith GL, et al. The African swine fever virus protein j4R binds to the alpha chain of nascent polypeptide-associated complex. *J Virol* (2002) 76:9991–9. doi: 10.1128/JVI.76.19.9991-9999.2002.

55. Epifano C, Krijnse-Locker J, Salas ML, Rodriguez JM, Salas J. The African swine fever virus nonstructural protein pB602L is required for formation of the icosahedral capsid of the virus particle. *J Virol* (2006) 80:12260–70. doi: 10.1128/JVI.01323-06.

56. Chacón MR, Almazán F, Nogal ML, Viñuela E, Rodríguez JF. The African swine fever virus IAP homolog is a late structural polypeptide. *Virology* (1995) 214:670–4. doi: 10.1006/viro.1995.0083.

57. Galindo I, Almazán F, Bustos MJ, Viñuela E, Carrascosa AL. African swine fever virus EP153R open reading frame encodes a glycoprotein involved in the

hemadsorption of infected cells. *Virology* (2000) 266:340–51. doi: 10.1006/viro.1999.0080.

58. Rodríguez JM, Yáñez RJ, Almazán F, Viñuela E, Rodriguez JF. African swine fever virus encodes a CD2 homolog responsible for the adhesion of erythrocytes to infected cells. *J Virol* (1993) 67:5312–20. doi: 10.1128/jvi.67.9.5312-5320.1993.

59. Oliveros M, Garcia-Escudero R, Alejo A, Viñuela E, Salas ML, Salas J. African swine fever virus dUTPase is a highly specific enzyme required for efficient replication in swine macrophages. *J Virol* (1999) 73:8934–43. doi: 10.1128/JVI.73.11.8934-8943.1999.

60. Jain M, Olsen HE, Paten B, Akeson M. The Oxford Nanopore MinION: delivery of nanopore sequencing to the genomics community. *Genome Biol* (2016) 17:1–11. doi: 10.1186/s13059-016-1103-0.

61. Olasz F, Tombacz D, Torma G, Csabai Z, Moldovan N, Dormo A, et al. Short and long-read sequencing survey of the dynamic transcriptomes of African swine fever virus and the host cells. *Front Genet* (2020) 11:2020.02.27. doi: 10.3389/fgene.2020.00758

62. Torma G, Tombacz D, Csabai Z, Moldovan N, Meszaros I, Zadori Z, et al. Combined short and long-read sequencing reveals a complex transcriptomic architecture of African swine fever virus. *Viruses* (2021) 13:579. doi: 10.3390/v13040579.

63. Hou H, Li Y, Wang M, Liu A, Yu Z, Chen K, et al. Structural insights into RNA polymerase III-mediated transcription termination through trapping poly-deoxythymidine. *Nat Commun* (2021) 12:6135. doi: 10.1038/s41467-021-26402-9.

64. Girbig M, Xie J, Grotsch H, Libri D, Porrua O, Muller CW. Architecture of the yeast Pol III pre-termination complex and pausing mechanism on poly(dT) termination signals. *Cell Rep* (2022) 40:111316. doi: 10.1016/j.celrep.2022.111316.

65. Wilkins D. *gggenes: Draw Gene Arrow Maps in 'ggplot2'*. R package version 0.4.1. (2020). Available at: https://CRAN.R-project.org/package=gggenes.

66. Soneson C, Yao Y, Bratus-Neuenschwander A, Patrignani A, Robinson MD, Hussain S. A comprehensive examination of Nanopore native RNA sequencing for characterization of complex transcriptomes. *Nat Commun* (2019) 10:3359. doi: 10.1038/s41467-019-11272-z

67. Ju X, Li D, Liu S. Full-length RNA profiling reveals pervasive bidirectional transcription terminators in bacteria. *Nat Microbiol* (2019) 4:1907–18. doi: 10.1038/s41564-019-0500-z.

68. Manley J. A complex protein assembly catalyzes polyadenylation of mRNA precursors. *Curr Opin Genet Dev* (1995) 5:222–8. doi: 10.1016/0959-437X(95)80012-3.

69. Hirtreiter A, Grohmann D, Werner F. Molecular mechanisms of RNA polymerase–the F/E (RPB4/7) complex is required for high processivity *in vitro*. *Nucleic Acids Res* (2010) 38:585–96. doi: 10.1093/nar/gkp928.

70. Arnvig KB. Riboswitches: choosing the best platform. *Biochem Soc Trans* (2019) 47:1091–9. doi: 10.1042/BST20180507.

71. Spitalny P, Thomm M. A polymerase III-like reinitiation mechanism is operating in regulation of histone expression in archaea. *Mol Microbiol* (2008) 67:958–70. doi: 10.1111/j.1365-2958.2007.06084.x.

72. Gusarov I, Nudler E. The mechanism of intrinsic transcription termination. *Mol Cell* (1999) 3:495–504. doi: 10.1016/S1097-2765(00)80477-3.

73. Porrua O, Boudvillain M, Libri D. Transcription termination: variations on common themes. *Trends Genet* (2016) 32:508–22. doi: 10.1016/j.tig.2016.05.007.

74. Li J, Yue L, Li Z, Zhang W, Zhang B, Zhao F, et al. aCPSF1 cooperates with terminator U-tract to dictate archaeal transcription termination efficacy. *Elife* (2021) 10:e70464. doi: 10.7554/eLife.70464

75. Stockl R, Nissl L, Reichelt R, Rachel R, Grohmann D, Grunberger F. The transcriptional regulator EarA and intergenic terminator sequences modulate archaellation in Pyrococcus furiosus. *Front Microbiol* (2023) 14:1241399. doi: 10.3389/fmicb.2023.1241399.

76. Howard ST, Ray CA, Patel DD, Antczak JB, Pickup DJ. A 43-nucleotide RNA cis-acting element governs the site-specific formation of the 3′ end of a poxvirus late mRNA. *Virology* (1999) 255:190–204. doi: 10.1006/viro.1998.9547

77. D'Costa SM, Bainbridge TW, Condit RC. Purification and properties of the vaccinia virus mRNA processing factor. *J Biol Chem* (2008) 283:5267–75. doi: 10.1074/jbc.M709258200.

78. Colgan DF, Manley JL. Mechanism and regulation of mRNA polyadenylation. *Genes Dev* (1997) 11:2755–66. doi: 10.1101/gad.11.21.2755.

79. Zhao J, Hyman L, Moore C. Formation of mRNA 3′ Ends in Eukaryotes: Mechanism, Regulation, and Interrelationships with Other Steps in mRNA Synthesis. *Microbiol Mol Biol Rev* (1999) 63:405–45. doi: 10.1128/MMBR.63.2.405-445.1999.

80. Millevoi S, Vagner S. Molecular mechanisms of eukaryotic pre-mRNA 3' end processing regulation. *Nucleic Acids Res* (2009) 38:2757–74. doi: 10.1093/nar/gkp1176

81. Black EP, Moussatche N, Condit RC. Characterization of the interactions among vaccinia virus transcription factors G2R, A18R, and H5R. *Virology* (1998) 245:313–22. doi: 10.1006/viro.1998.9166.

82. Xiang Y, Simpson DA, Spiegel J, Zhou A, Silverman RH, Condit RC. The vaccinia virus A18R DNA helicase is a postreplicative negative transcription elongation factor. *J Virol* (1998) 72:7012–23. doi: 10.1128/JVI.72.9.7012-7023.1998.

83. Li J, Pennington MJ, Broyles SS. Temperature-sensitive mutations in the gene encoding the small subunit of the vaccinia virus early transcription factor impair

84. Hindman R, Gollnick P. Nucleoside triphosphate phosphohydrolase I (NPH I) functions as a 5' to 3' Translocase in transcription termination of vaccinia early genes. *J Biol Chem* (2016) 291:14826–38. doi: 10.1074/jbc.M116.730135.

85. Gross CH, Shuman S. Vaccinia virions lacking the RNA helicase nucleoside triphosphate phosphohydrolase II are defective in early transcription. *J Virol* (1996) 70:8549–57. doi: 10.1128/jvi.70.12.8549-8557.1996.

86. Yuen L, Moss B. Oligonucleotide sequence signaling transcriptional termination of vaccinia virus early genes. *Proc Natl Acad Sci USA* (1987) 84:6417–21. doi: 10.1073/pnas.84.18.6417.

87. Taylor SD, Solem A, Kawaoka J, Pyle AM. The NPH-II helicase displays efficient DNA x RNA helicase activity and a pronounced purine sequence bias. *J Biol Chem* (2010) 285:11692–703. doi: 10.1074/jbc.M109.088559.

88. Grimm C, Bartuli J, Boettcher B, Szalay AA, Fischer U. Structural basis of the complete poxvirus transcription initiation process. *Nat Struct Mol Biol* (2021) 28:779–88. doi: 10.1038/s41594-021-00655-w.

89. Hillen HS, Bartuli J, Grimm C, Dienemann C, Bedenk K, Szalay AA, et al. Structural basis of poxvirus transcription: transcribing and capping vaccinia complexes. *Cell* (2019) 179:1525–1536.e12. doi: 10.1016/j.cell.2019.11.023

90. Cackett G, Sykora M, Werner F. Transcriptome view of a killer: African swine fever virus. *Biochem Soc Trans* (2020) 48:1569–81. doi: 10.1042/BST20191108.

91. Rice P, Longden I, Bleasby A. EMBOSS: the European molecular biology open software suite. *Trends Genet TIG* (2000) 16:276–7. doi: 10.1016/S0168-9525(00)02024-2.

92. Bayliss CD, Condit RC. The vaccinia virus A18R gene product is a DNA-dependent ATPase. *J Biol Chem* (1995) 270:1550–6. doi: 10.1074/jbc.270.4.1550.

93. Li D, Wu P, Liu H, Feng T, Yang W, Ru Y, et al. A QP509L/QP383R-Deleted African Swine Fever Virus Is Highly Attenuated in Swine but Does Not Confer Protection against Parental Virus Challenge. *J Virol* (2022) 96:e015002. doi: 10.1128/JVI.01500-21.

94. Ramirez-Medina E, Vuono EA, Pruitt S, Rai A, Espinoza N, Velazquez-Salinas L, et al. Evaluation of an ASFV RNA helicase gene A859L for virus replication and swine virulence. *Viruses* (2021) 14:10. doi: 10.3390/v14010010.

95. Cackett G, Portugal R, Matelska D, Dixon L, Werner F. African swine fever virus and host response: transcriptome profiling of the Georgia 2007/1 strain and porcine macrophages. *J Virol* (2022) 96:e0193921. doi: 10.1128/jvi.01939-21.

96. Freitas FB, Frouco G, Martins C, Ferreira F. The QP509L and Q706L superfamily II RNA helicases of African swine fever virus are required for viral replication, having non-redundant activities. *Emerg Microbes Infect* (2019) 8:291–302. doi: 10.1080/22221751.2019.1578624.

97. Christen LM, Sanders M, Wiler C, Niles EG. Vaccinia virus nucleoside triphosphate phosphohydrolase I is an essential viral early gene transcription termination factor. *Virology* (1998) 245:360–71. doi: 10.1006/viro.1998.9177.

98. Bayliss CD, Smith GL. Vaccinia virion protein I8R has both DNA and RNA helicase activities: implications for vaccinia virus transcription. *J Virol* (1996) 70:794–800. doi: 10.1128/jvi.70.2.794-800.1996.

99. Guo Z, Li Y, Ding SW. Small RNA-based antimicrobial immunity. *Nat Rev Immunol* (2019) 19:31–44. doi: 10.1038/s41577-018-0071-x.

100. Hovanessian AG, Kerr IM. Synthesis of an oligonucleotide inhibitor of protein synthesis in rabbit reticulocyte lysates analogous to that formed in extracts from interferon-treated cells. *Eur J Biochem* (1978) 84:149–59. doi: 10.1111/j.1432-1033.1978.tb12151.x.

101. Feng Y, Zhang X, Graves P, Zeng Y. A comprehensive analysis of precursor microRNA cleavage by human Dicer. *RNA* (2012) 18:2083–92. doi: 10.1261/rna.033688.112.

102. Gurung C, Fendereski M, Sapkota K, Guo J, Huang F, Guo YL. Dicer represses the interferon response and the double-stranded RNA-activated protein kinase pathway in mouse embryonic stem cells. *J Biol Chem* (2021) 296:100264. doi: 10.1016/j.jbc.2021.100264.

103. Donovan J, Rath S, Kolet-Mandrikov D, Korennykh A. Rapid RNase L-driven arrest of protein synthesis in the dsRNA response without degradation of translation machinery. *RNA* (2017) 23:1660–71. doi: 10.1261/rna.062000.117.

104. Sun H, Wu M, Zhang Z, Wang Y, Yang J, Liu Z, et al. OAS1 suppresses African swine fever virus replication by recruiting TRIM21 to degrade viral major capsid protein. *J Virol* (2023) 97:e0121723. doi: 10.1128/jvi.01217-23.

105. Li H. Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics* (2018) 34:3094–100. doi: 10.1093/bioinformatics/bty191.

106. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. The sequence alignment/Map format and SAMtools. *Bioinformatics* (2009) 25:2078–9. doi: 10.1093/bioinformatics/btp352.

107. Thorvaldsdottir H, Robinson JT, Mesirov JP. Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. *Briefings Bioinf* (2013) 14:178–92. doi: 10.1093/bib/bbs017.

108. Quinlan AR, Hall IM. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* (2010) 26:841–2. doi: 10.1093/bioinformatics/btq033.

109. RStudioTeam. *Integrated development for R. @ in* RStudio Vol. 42 Boston, MA: RStudioe, Inc (2020). p. 14.

110. Wickham H, François R, Henry L, Müller K, Vaughan D. *dplyr: A Grammar of Data Manipulation*. R package version 1.1.4. (2023). Available at: https://dplyr.tidyverse.org.

111. Wickham H. *ggplot2: Elegant Graphics for Data Analysis* New York: Springer New York, NY (2009). doi: 10.1007/978-0-387-98141-3.

112. Ramírez F, Dündar F, Diehl S, Grüning BA, Manke T. deepTools: a flexible platform for exploring deep-sequencing data. *Nucleic Acids Res* (2014) 42:W187–91. doi: 10.1093/nar/gku365.

113. Lawrence M, Gentleman R, Carey V. rtracklayer: an R package for interfacing with genome browsers. *Bioinformatics* (2009) 25:1841–2. doi: 10.1093/bioinformatics/btp328.

114. Aparicio-Puerta E, Lebron R, Rueda A, Gomez-Martin C, Giannoukakos S, Jaspez D, et al. sRNAbench and sRNAtoolbox 2019: intuitive fast small RNA profiling and differential expression. *Nucleic Acids Res* (2019) 47:W530–5. doi: 10.1093/nar/gkz415.

115. Bailey TL, Johnson J, Grant CE, Noble WS. The MEME suite. *Nucleic Acids Res* (2015) 43:W39–49. doi: 10.1093/nar/gkv416.

116. Thodberg M, Thieffry A, Vitting-Seerup K, Andersson R, Sandelin A. CAGEfightR: Analysis of 5′-end data using R/Bioconductor. *BMC Bioinf* (2019) 20:487. doi: 10.1186/s12859-019-3029-5.

117. Crooks GE, Hon G, Chandonia JM, Brenner SE. WebLogo: A sequence logo generator. *Genome Res* (2004) 14:1188–90. doi: 10.1101/gr.849004.

118. Lorenz R, Bernhart SH, Honer Zu Siederdissen C, Tafer H, Flamm C, Stadler PF, et al. ViennaRNA package 2.0. *Algorithms Mol Biol* (2011) 6:26. doi: 10.1186/1748-7188-6-26.

119. Cackett G. University College London. (2022).