



African Swine Fever Virus and Host Response: Transcriptome Profiling of the Georgia 2007/1 Strain and Porcine Macrophages

Gwenny Cackett,^a Raquel Portugal,^b Dorota Matelska,^a Linda Dixon,^b Finn Werner^a

^aInstitute for Structural and Molecular Biology, University College London, London, United Kingdom

^bPirbright Institute, Pirbright, Surrey, United Kingdom

Gwenny Cackett and Raquel Portugal contributed equally to this article. Author order was determined alphabetically.

ABSTRACT African swine fever virus (ASFV) has a major global economic impact. With a case fatality in domestic pigs approaching 100%, it currently presents the largest threat to animal farming. Although genomic differences between attenuated and highly virulent ASFV strains have been identified, the molecular determinants for virulence at the level of gene expression have remained opaque. Here, we characterize the transcriptome of ASFV genotype II Georgia 2007/1 (GRG) during infection of the physiologically relevant host cells, porcine macrophages. In this study, we applied cap analysis gene expression sequencing (CAGE-seq) to map the 5' ends of viral mRNAs at 5 and 16 h postinfection. A bioinformatics analysis of the sequence context surrounding the transcription start sites (TSSs) enabled us to characterize the global early and late promoter landscape of GRG. We compared transcriptome maps of the GRG isolate and the lab-attenuated BA71V strain that highlighted GRG virulence-specific transcripts belonging to multigene families, including two predicted MGF 100 genes, I7L and I8L. In parallel, we monitored transcriptome changes in the infected host macrophage cells. Of the 9,384 macrophage genes studied, transcripts for 652 host genes were differentially regulated between 5 and 16 h postinfection compared with only 25 between uninfected cells and 5 h postinfection. NF- κ B activated genes and lysosome components such as S100 were upregulated, and chemokines such as CCL24, CXCL2, CXCL5, and CXCL8 were downregulated.

IMPORTANCE African swine fever virus (ASFV) causes hemorrhagic fever in domestic pigs, with case fatality rates approaching 100% and no approved vaccines or antivirals. The highly virulent ASFV Georgia 2007/1 strain (GRG) was the first isolated when ASFV spread from Africa to the Caucasus region in 2007, then spreading through Eastern Europe and, more recently, across Asia. We used an RNA-based next-generation sequencing technique called CAGE-seq to map the starts of viral genes across the GRG DNA genome. This has allowed us to investigate which viral genes are expressed during early or late stages of infection and how this is controlled, comparing their expression to the nonvirulent ASFV-BA71V strain to identify key genes that play a role in virulence. In parallel, we investigated how host cells respond to infection, which revealed how the ASFV suppresses components of the host immune response to ultimately win the arms race against its porcine host.

KEYWORDS African swine fever virus, CAGE-seq, RNA-seq, gene expression, innate immunity, promoters, transcription, transcriptome, virology

African swine fever virus (ASFV) originated in sub-Saharan Africa, where it remains endemic. However, this was followed by the introduction in 2007 of a genotype II isolate to Georgia (1) and subsequent spread in Russia and Europe. The virus was then introduced to China in 2018 (2), from here it spread rapidly across Asia, strongly emphasizing

Editor Rozanne M. Sandri-Goldin, University of California, Irvine

Copyright © 2022 Cackett et al. This is an open-access article distributed under the terms of the [Creative Commons Attribution 4.0 International license](https://creativecommons.org/licenses/by/4.0/).

Address correspondence to Linda Dixon, linda.dixon@pirbright.ac.uk, or Finn Werner, f.werner@ucl.ac.uk.

The authors declare no conflict of interest.

Received 23 November 2021

Accepted 9 December 2021

Accepted manuscript posted online

12 January 2022

Published 9 March 2022

this disease as a severe threat to global food security. ASFV is the only characterized member of the *Asfarviridae* family (3) in the recently classified *Nucleocytoviricota* (ICTV Master Species List 2019.v1) phylum (4, 5). ASFV has a linear double-stranded DNA (dsDNA) genome of ~170 to 193 kbp encoding ~150 to ~200 open reading frames (ORFs). Until recently, little was known about either the transcripts expressed from the ASFV genome or the mechanisms of ASFV transcription. Much of what is known about transcription is extrapolated from vaccinia virus (VACV), a distantly related *Nucleocytoviricota* member, from the *Poxviridae* family (6). ASFV encodes a eukaryotic-like 8-subunit RNA polymerase (RNAP), an mRNA capping enzyme, and poly-A polymerase, all of which are carried within mature virus particles (7). These virions are transcription competent upon solubilization *in vitro* (8) and support mRNA modification by including a 5'-methylated cap and a 3' polyadenylated (polyA) tail of ~33 nucleotides long (8, 9).

Viral genes are typically classified according to their temporal expression patterns; ASFV genes have historically been categorized as “immediate early” when expressed immediately following infection, as “early genes” following the onset of viral protein synthesis, as “intermediate genes” after the onset of viral DNA replication, or as “late genes” thereafter. The temporal regulation of transcription is likely enabled by different sets of general transcription initiation factors that recognize distinct early or late promoter motifs (EPM and LPM, respectively), as we previously investigated in the ASFV-BA71V strain (10) and address further in this study. EPM recognition is likely enabled by the ASFV homologue of heterodimeric VACV early transcription factor (VETF), consisting of D1133L (D6) and G1340L (A7) gene products, which bind the Poxvirus early gene promoter motif (11–13), which the ASFV EPM strongly resembles. Both ASFV-D6 and ASFV-A7 are late genes, i.e., synthesized late during infection (10) and packaged into virus particles (7). The ASFV LPM is less well defined than the EPM, but a possible initiation factor involved in its recognition is the ASFV-encoded viral homolog of the eukaryotic TATA-binding protein (TBP), expressed during early infection (10). By analogy with the VACV system, additional factors, including homologs of A1, A2, and G8, may also contribute to late transcription initiation (6).

We have recently carried out a detailed and comprehensive ASFV whole-genome expression analysis using complementary next-generation sequencing (NGS) results and computational approaches to characterize the ASFV transcriptome following BA71V infection of Vero cells at 5 h postinfection (hpi) and 16 hpi (10). Most of our knowledge about the molecular biology of ASFV, including gene expression, has been derived from cell culture-adapted, attenuated virus strains, such as BA71V infecting Vero tissue culture cells (9, 10). These model systems provide convenient models to study the replication cycle but have deletions of many genes that are not essential for replication but have important roles in virulence within its natural porcine hosts (14–16). To date, 24 ASFV genotypes have been identified in Africa (16–23), while all strains spreading across Asia and Europe belong to the type II genotype. Most of these are highly virulent in domestic pigs and wild boar, including the ASFV Georgia 2007/1 (GRG) (24) and the Chinese ASFV Heilongjiang and 2018 (Pig/HLJ/18) (25) isolates. Though a number of less virulent isolates have been identified in wild boar in the Baltic states and domestic pigs in China (26–29). It is crucial to understand the similarities and commonalities between ASFV strains and to characterize the host response to these in order to understand the molecular determinants for ASFV pathogenicity. Information about the gene content and genome organization can be gained from comparing virus genome sequences. However, only functional genomics such as transcriptome or proteome analyses can provide information about the differences in gene expression programs and the host responses to infection.

On the genome level, most differences between virulent (e.g., GRG) and attenuated (e.g., lab-attenuated BA71V) ASFV strains reside toward the genome termini. Figure 1a shows a whole-genome comparison of GRG (left) and BA71V (right) strains with the sequence conservation color-coded in different shades of blue. The regions toward the ends of the genome are more dynamic compared to the central region, which is highly

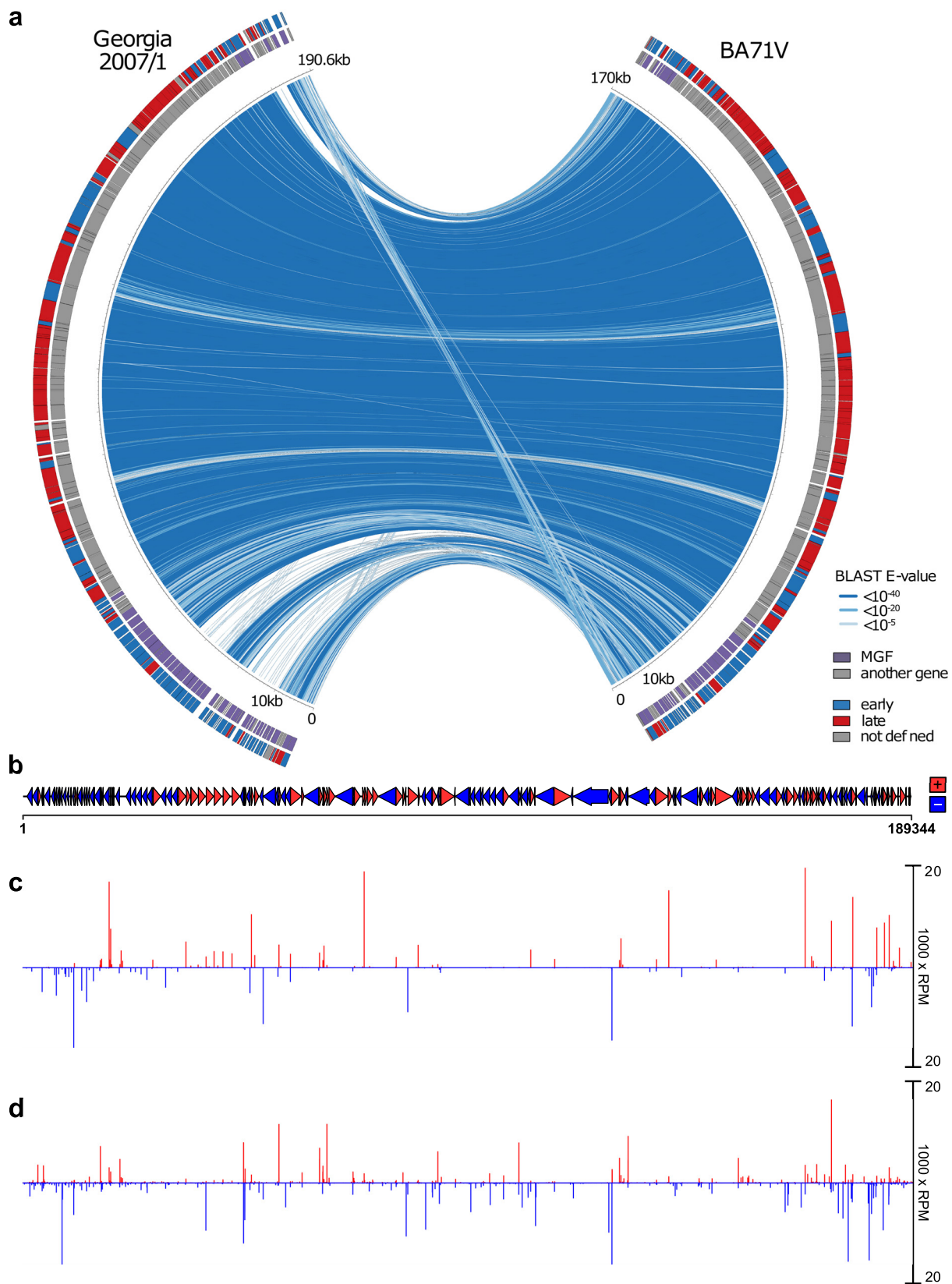


FIG 1 Functional genome annotation of ASFV GRG. (a) Comparison between the genomes of BA71V and GRG, generated with Circos (<http://circos.ca/>). Blue lines represent sequence conservation (BLAST E values per 100 nt). The inner ring represents genes defined as MGF (Continued on next page)

conserved, as genes at the termini are prone to deletion, duplication, insertion, and fusion (17, 30). Most of the GRG-specific genes are expressed early during infection (early genes are color-coded blue in the outer arch of Fig. 1a), and many belong to multigene families (MGFs, purple in the inner arch). The functions of many MGF members remain poorly understood, though variation among MGFs is linked to virulence (31), and deleting members of MGF 360 and 505 families has been shown to reduce virulence (32, 33). Deletion of MGF 505-7R or MGF 110-9L also partially attenuated the virus in pigs (34, 35). In contrast, deletion of MGF 110-1L and MGF 100-1R did not reduce virus virulence (17). Members of MGF 110 are highly expressed on both the mRNA and protein levels in infections with the BA71V isolate or OURT88/3 (10, 36), suggesting MGF 110 holds importance during infection. Overall, the functions of MGF 360 and 505 members are better characterized than other MGFs, playing a role in evading the host type I interferon (IFN) response (15, 32, 37–40). In summary, comparing the expression of ASFV genes, especially MGFs, between the virulent GRG and the lab-adapted BA71V strains, is fundamental in the identification of virulence factors and better MGF characterization.

Macrophages are the primary target cells for ASFV; they are important immune effector cells that display remarkable plasticity allowing efficient response to environmental signals (41). There are some studies which have investigated how host macrophages respond to infection, including a microarray analysis of primary swine macrophage cells infected with virulent GRG (42). There are two RNA-seq studies of whole blood or tissues isolated from pigs postmortem, which were either infected with a low pathogenic ASFV-OURT 88/3 or ASFV-GRG (43) or infected with a pathogenic Chinese isolate, ASFV-SY18 (44). Recently, two reports have been published about the transcriptomic response of porcine macrophages to infection with a virulent Chinese genotype II isolate using a low multiplicity of infection (MOI, 1) and classical RNA-seq (45, 46), but due to different experimental conditions the various results are somewhat challenging to compare with other studies. It must also be remembered that neither these classical RNA-seq nor microarray analyses have sufficient resolution to accurately capture viral gene expression in the compact ASFV genome alongside that of the host.

Here, we applied CAGE-seq to characterize the transcriptome of the highly virulent GRG isolate (24) in primary porcine macrophages, the biologically relevant target cells for ASFV infection. In this study, we used a high multiplicity of infection (MOI, 5) so that transcripts expressed during a single-cycle time course could be measured without the complication of variable proportions of uninfected cells being present. We investigated the differential gene expression patterns of viral mRNAs at early and late time points of 5 and 16 hpi and mapped the viral promoter motifs. Importantly, we have compared the expression levels and temporal regulation of genes conserved in both the virulent GRG isolate and the attenuated tissue-culture-adapted BA71V strain. With a few exceptions, both mRNA expression levels and temporal regulation of the conserved genes are surprisingly similar. This confirms that it is not deregulation of their conserved genes, but the virulent isolate-specific genes, which are the key determinants for ASFV virulence. Most of these genes are MGF members, likely involved in suppression of the host immune response. Indeed, transcriptome analysis of the porcine macrophages upon GRG infection reflects a modulation of host immune response genes, although the bulk of the ~9,000 genes studied did not significantly change expression levels during infection.

FIG 1 Legend (Continued)

members (purple) and all others (gray). The outer ring shows annotated genes which we have defined as early or late according to downregulation or upregulation between 5 hpi and 16 hpi from DESeq2 analysis. (b) GRG annotated ORFs ($n = 189$) are represented as arrows and colored according to strand. (c and d) CAGE-seq peaks across the GRG genome at 5 hpi (c) and 16 hpi (d), normalized coverage reads per million mapped reads (RPM) of 5' ends of CAGE-seq reads. The coverage was capped at 20,000 RPM for visualization, though multiple peaks exceeded this. deepTools (116) was used to convert bam files to bigwig format and import them into RStudio for visual representation via the packages ggplot, ggbio, and rtracklayer, and gggenes was used to generate the ORF map in panel b.

RESULTS

Genome-wide transcription start site-mapping. We infected primary porcine alveolar macrophages with ASFV GRG at a high multiplicity of infection (MOI, 5.0), isolated total RNA at 5 hpi and 16 hpi, and sequenced using CAGE-seq (see Table S1a in the supplemental material). The resulting mRNA 5' ends were mapped to the GRG genome (Fig. 1b), resulting in the annotation of 229 and 786 transcription start sites (TSSs) at 5 and 16 hpi, respectively (Fig. 1c and d, from Table S1b and c, respectively). The majority of TSSs were identified within 500 bp upstream of the start codon of a given ORF, a probable location for a bona fide gene TSS. The strongest and closest TSSs upstream of ORFs were annotated as "primary" TSSs (pTSSs, listed in Table S1d), and in this manner we could account for the TSSs of 177 out of 189 GRG ORFs annotated in the GRG genome (GenBank accession number [FR682468.1](https://www.ncbi.nlm.nih.gov/nuccore/FR682468.1)). TSS signals below the threshold for detection included MGF_110-11L, C62L, and E66L, the remainder being short ORFs designated "ASFV_G_ACD," predicted solely from the GRG genome sequence (24). The E66L ORF was originally predicted from only the BA71V genome sequence, but likewise was undetectable with CAGE-seq (10), making its expression unlikely. Our TSS mapping identified novel ORFs (nORFs) downstream of the TSS, which were included in the curated GRG genome map (Table S1d includes pTSSs of annotated ORFs and nORFs in gene feature file, or GFF, format, details of nORF-generating TSSs are found in Table S2a). In addition to ORF-associated TSSs, some were located within ORFs (intra-ORF or ioTSS) or in between them (inter-ORF TSS), and all detected TSSs are listed in Table S1b and c.

Expression of GRG genes during early and late infection. Having annotated TSSs across the GRG genome, we quantified the viral mRNAs originating from pTSSs from CAGE-seq data, normalizing against the total number of reads mapping to the ASFV genome (i.e. RPM, or reads per million mapped reads per sample). We compared gene expression between early and late infection and simplistically defined genes as early or late if they were significantly down- or upregulated (respectively), using DESeq2 (47). In summary, 165 of the 177 detectable genes were differentially expressed (adjusted *P* value, or *padj* < 0.05, Table S1e). Those showing no significant change were D345L, DP79L, I8L, MGF_100-1R, A859L, QP383R, B475L, E301R, DP63R, C147L, and I177L. A total of 87 of those 165 differentially expressed genes were significantly downregulated, thus representing the early genes, while 78 of the 165 genes were upregulated or late genes. The majority of MGFs were early genes, apart from MGF 505-2R, MGF 360-2L, and MGF 100-1L (Fig. 2a). Figure 2b shows the expression patterns of GRG-exclusively expressed genes, which we defined as only having a detectable CAGE-seq TSS in GRG, and not in BA71V (regardless of the presence in the BA71V genome). These, unsurprisingly, consist of many MGFs (19), all of which were early genes (Fig. 2b), barring MGF 100-1L. In addition, genes I9R, I10L, and I11L and several of the newly annotated short ORFs were specific to GRG.

We extracted the top 20 most highly expressed genes of GRG (as RPM) during 5 hpi (Fig. 2c) and 16 hpi (Fig. 2d) postinfection. The following 10 genes are shared between both top 20 lists: MGF 110-3L, A151R, MGF 110-7L, MGF 110-5L-6L, I73R, 285L, CP312R, ASFV_G_ACD_00600, MGF 110-4L, and CP204L. It is important to note that the relative expression values (RPM) for genes at 5 hpi are significantly higher than those at 16 hpi. This is consistent with our observations in the BA71V strain (10) and due to the increase in global viral transcript levels during late infection discussed below. Table S1f includes all the GRG annotated ORFs, their TSS locations during early and late infection, their relative distances if these TSS locations differ, and their respective 5' untranslated region (UTR) lengths.

GRG and BA71V share strong similarity between conserved gene expression. Next, we carried out a direct comparison of mRNA levels from 132 conserved genes between the virulent GRG and attenuated BA71V (10) strain making use of our previously published CAGE-seq data. The relative transcript levels (RPM) of the genes conserved between the two strains showed a significant correlation at 5 hpi (Fig. 3a) and 16 hpi (Fig. 3b), supported by the heatmap in Fig. S1 and the RPM for each gene,

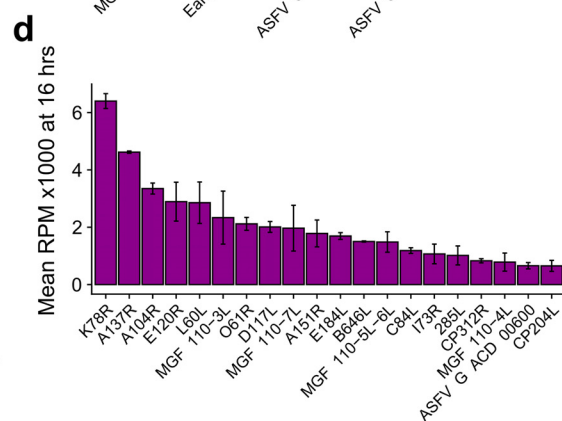
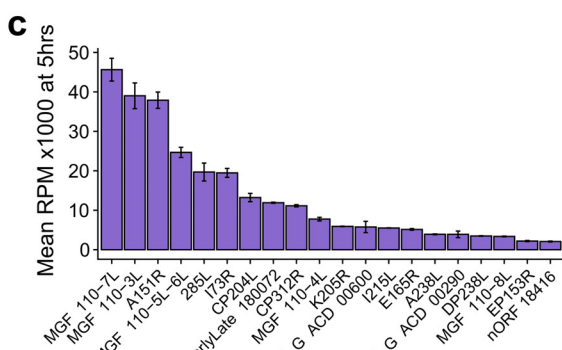
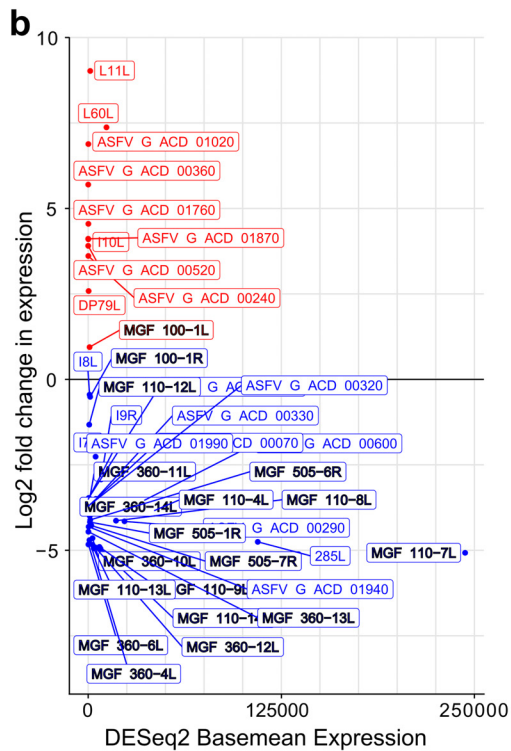
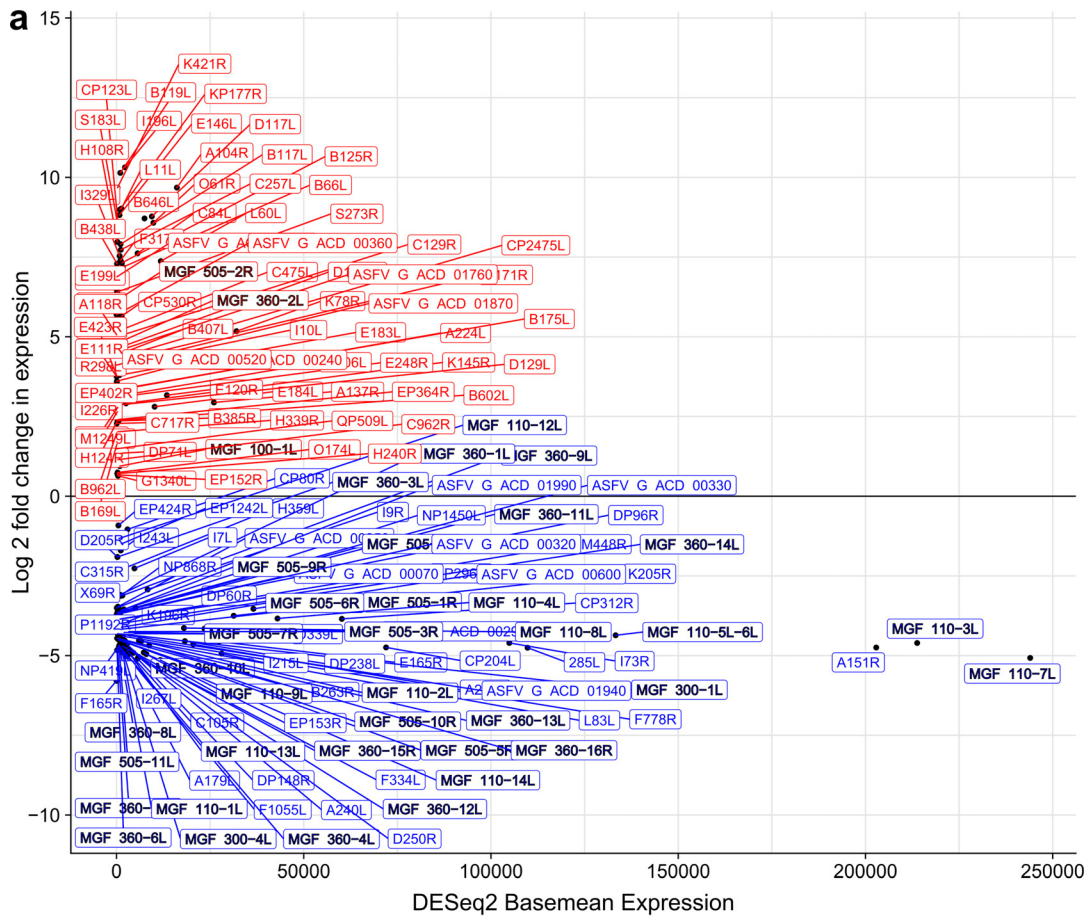


FIG 2 Summary of GRG gene expression. (a) Expression profiles for 164 genes for which we annotated pTSSs from CAGE-seq and which showed significant differential expression. Log₂ fold change and base mean expression values were from DESeq2 (Continued on next page)

across both time points and replicates, showing a strong congruence between the two strains. Of the 132 conserved genes, 125 showed significant differential expression in both strains; 119 of these 125 showed the same down- or upregulated patterns of significant differential expression from 5 hpi to 16 hpi (Fig. 3c; early genes in blue, late genes in red). The exceptions are D205R, CP80R, C315R, NP419L, F165R, and DP148R (MGF 360-18R), encoding RNA polymerase subunits RPB5 and RPB10 (15), transcription factor IIB (TFIIB) (15), DNA ligase (48), a putative signal peptide-containing protein, and a virulence factor (49), respectively. The ASFV-TFIIB homolog (C315R) is classified as an early gene in GRG but not in BA71V, in line with the predominantly early-expressed TBP (B263R), its predicted interaction partner. It is worth noting however, that D205R, CP80R, and C315R are close to the threshold of significance, with transcripts being detected at both 5 hpi and 16 hpi (Table S1e).

Increased and pervasive transcription during late infection. During late infection of BA71V (10), we noted an increase in genome-wide mRNA abundance, as well as an increasing number of TSSs and transcription termination sites, reminiscent of pervasive transcription observed during late infection of vaccinia virus (50). To quantify and compare the global mRNA increase in both BA71V and GRG, we calculated the ratio of read coverage per nucleotide, at 16 hpi versus 5 hpi (\log_2 transformed ratio of RPM), across the viral genome (Fig. 4a; increase shown above and decrease below the x axis). This dramatic increase is due to the overall increase of virus mRNAs present, which is visible in both strains (Fig. 4b), with an ~ 2 -fold increase in GRG from 5 hpi to 16 hpi, versus ~ 8 -fold in BA71V (Fig. 4c).

This observation can at least in part be attributed to the larger number of viral genomes during late infection, with increased levels of viral RNAP and associated factors available for transcription, following viral protein synthesis. Viral DNA-binding proteins, such as histone-like A104R (51), may remain associated with the genome originating from the virus particle in early infection. This could suppress spurious transcription initiation compared to that of freshly replicated nascent genomes that are highly abundant in late infection. In order to test whether the increased mRNA levels correlated with the increased number of viral genomes in the cell, we determined the viral genome copy number by using quantitative PCR (qPCR against the p72 capsid gene sequence) using purified total DNA from infected cells isolated at 0 hpi, 5 hpi, and 16 hpi and normalized values to the total amount of input DNA. Using this approach, we observed genome copy levels that were consistent from 0 hpi to 5 hpi, consistent with this being pre-DNA replication, followed by a substantial increase at 16 hpi, which was more pronounced in BA71V infection (Fig. 4d). This corresponded to a 15-fold increase in GRG genome copy numbers from late compared to early times postinfection of porcine macrophages and a 30-fold increase in BA71V during infection of Vero cells (Fig. 4e). In summary, the ASFV transcriptome changes both qualitatively and quantitatively as infection progresses, and the increase of virus mRNAs during late infection is accompanied by the dramatic increase in viral genome copies. Interestingly, the increase in viral transcripts and genome copies was less dramatic in the virulent GRG strain.

Correcting the bias of temporal expression patterns. The standard methods of defining differential gene expression are well established in transcriptomics using programs like DESeq2 (47). This is a very convenient and powerful tool which captures the nuances of differential expression in complex organisms. However, virus transcription is often characterized by more extreme changes, typically ranging from zero to millions of reads. Furthermore, in both BA71V and GRG strains, the genome-wide mRNA levels

FIG 2 Legend (Continued)

analysis of raw counts (see Materials and Methods). Genes are colored according to their \log_2 fold change in expression as red (positive, upregulated from 5 hpi to 16 hpi) or blue (negative, downregulated). MGFs are emphasized with a black outline to highlight their overrepresentation in the group of downregulated genes. (b) Expression profiles for 41 genes (excluding nORFs) only detected as being expressed in GRG and not BA71V; format is as in panel a. (c) Expression (RPM) of the 20 highest-expressed genes at 5 hpi; error bars represent the standard deviation between replicates. (d) Expression (RPM) of the 20 highest-expressed genes at 16 hpi; error bars are the standard deviation between replicates.

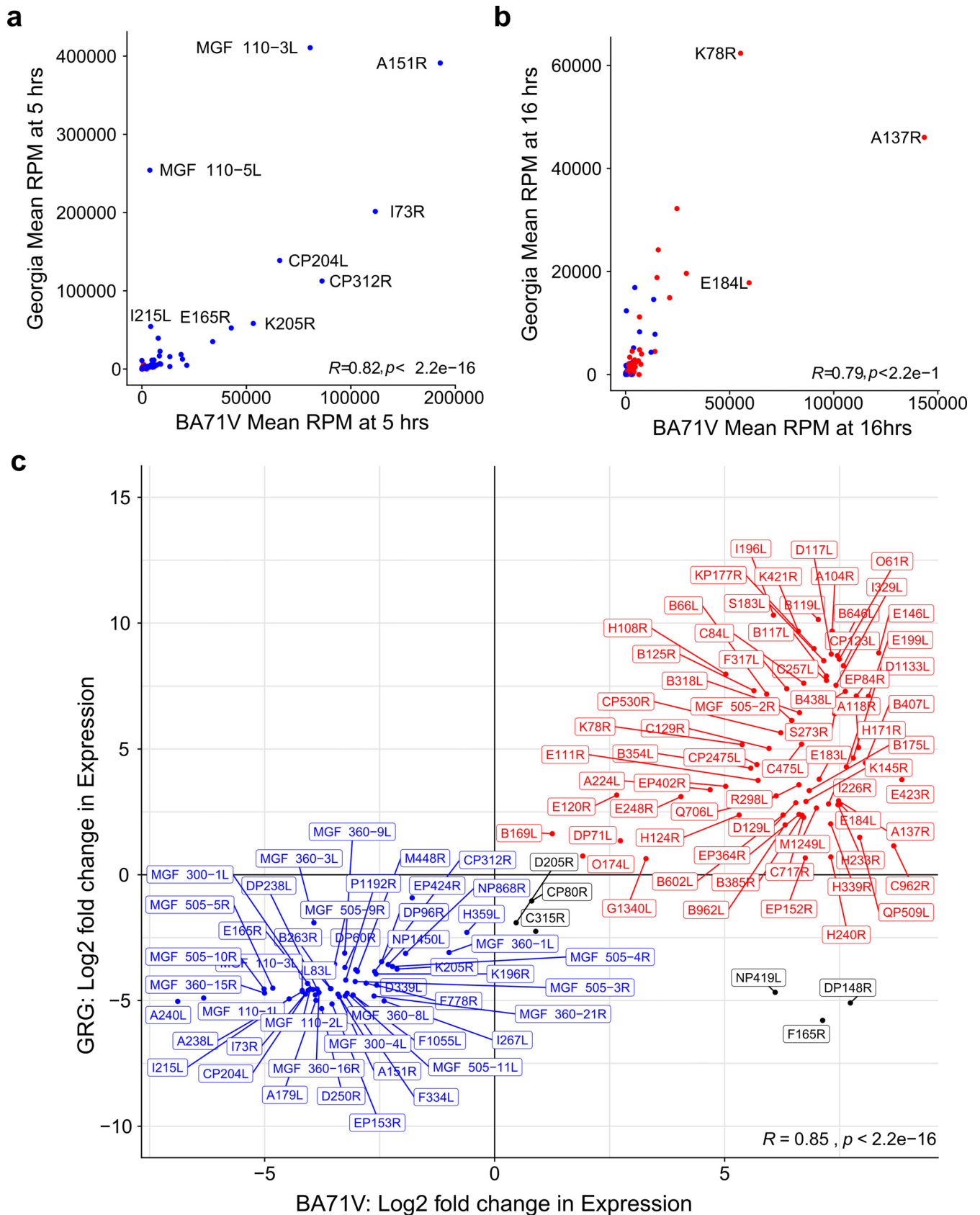


FIG 3 Comparison of gene expression profiles for genes shared between GRG and BA71V. (a and b) Scatterplots of mean RPM across replicates for shared genes at 5 hpi (a) and 16 hpi (b), colored according to whether genes show significant downregulation (blue) or upregulation (red) according to DESeq2 (Continued on next page)

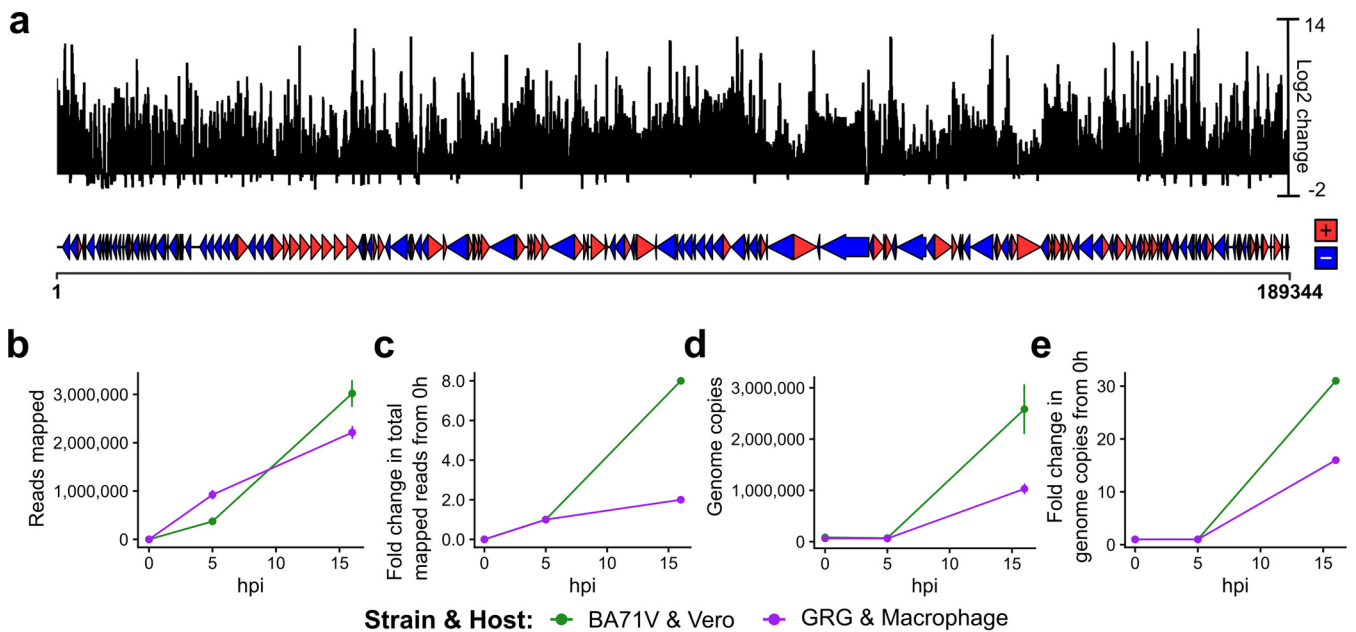


FIG 4 Increase in virus genome copy number mRNA levels during late infection. (a) The “log₂ change” represents log₂ of the ratio of CAGE-seq reads (normalized per million mapped reads) at 16 hpi versus 5 hpi per nucleotide across the genome. Alignment comparisons and calculations were done with deepTools (116). (b) Replicate means of CAGE-seq reads mapped to either the BA71V (green) or GRG (purple) genomes throughout infection. (c) Fold change in CAGE-seq reads during infection, calculated via mean value across 2 replicates, but with the assumption that the number of reads at 0 hpi is 0, therefore dividing by the values from 5 hpi. (d) Change in genome copies from DNA qPCR of the B646L gene, dividing by the value at 0 hpi to represent “1 genome copy per infected cell.” (e) Fold change in genome copies present at 0 hpi, 5 hpi, and 16 hpi from qPCR in panel d. Panel d was calculated as for panel c, but with actual values for 0 hpi.

and total ASFV reads increase over the infection time course (Fig. 4 and Table S1a). As a consequence, such normalization against the total mapped transcripts per sample (RPM) generates overestimated relative expression values at 5 hpi and understates those at 16 hpi (10). In order to validate the early-late expression patterns derived from CAGE-seq, we carried out reverse transcriptase PCR (RT-PCR) for selected viral genes, as this signal is proportionate to the number of specific mRNAs regardless of the level of other transcripts—with the minor caveat that it can pick up readthrough transcripts from upstream genes. We tested differentially expressed conserved genes, including GRG early (MGF 505-7R, MGF 505-9R, NP419L) and D345L, which showed stable relative expression values (RPM values in Supplementary Table 1e). All selected genes showed a consistently stronger RT-PCR signal during late infection in both BA71V and GRG (Fig. 5a to d). The exception is NP419L, whose levels were largely unchanged, and this is an example of how a gene whose transcript levels remain constant would be considered downregulated, when almost all other mRNA levels increase (Fig. 5b).

The standard normalization of NGS reads against total mapped reads (RPM) is regularly used, as it enables a statistical comparison between samples and conditions, subject to experimental variations (52). Keeping this in mind, we used an additional method of analyzing the “raw” read counts to represent global ASFV transcript levels that are not skewed by the normalization against total mapped reads. Figure 5 shows a side-by-side comparison of RT-PCR results and the CAGE-seq data normalized (RPM) or expressed as raw counts beneath each RT-PCR gel. Unlike CAGE-seq, RT-PCR will

FIG 3 Legend (Continued)

analysis in GRG. In both panels b and c, genes with RPM values above 40,000 RPM in either strain are labeled. (c) Comparison of log₂ fold change in expression values of genes in GRG and BA71V; in blue are downregulated (early) genes in both strains, and red are upregulated (late) genes in both strains, while the genes which disagree in their differential expression patterns between strains are in black. R represents the Pearson correlation coefficient for each individual plot in panels a, b, and c. Due to inconsistencies in their genome annotations, two genes were omitted from the BA71V-GRG transcriptome comparisons in Fig. 2b and 3a to d—EP296R in GRG is known as E296R in BA71V, and C122R (GRG) is the old nomenclature for C105R (BA71V), which are now correctly named in Table S1e and Fig. 2a. Both genes showed the same early expression patterns in BA71V (10) and GRG (Table S1e) and so would strengthen the patterns observed.

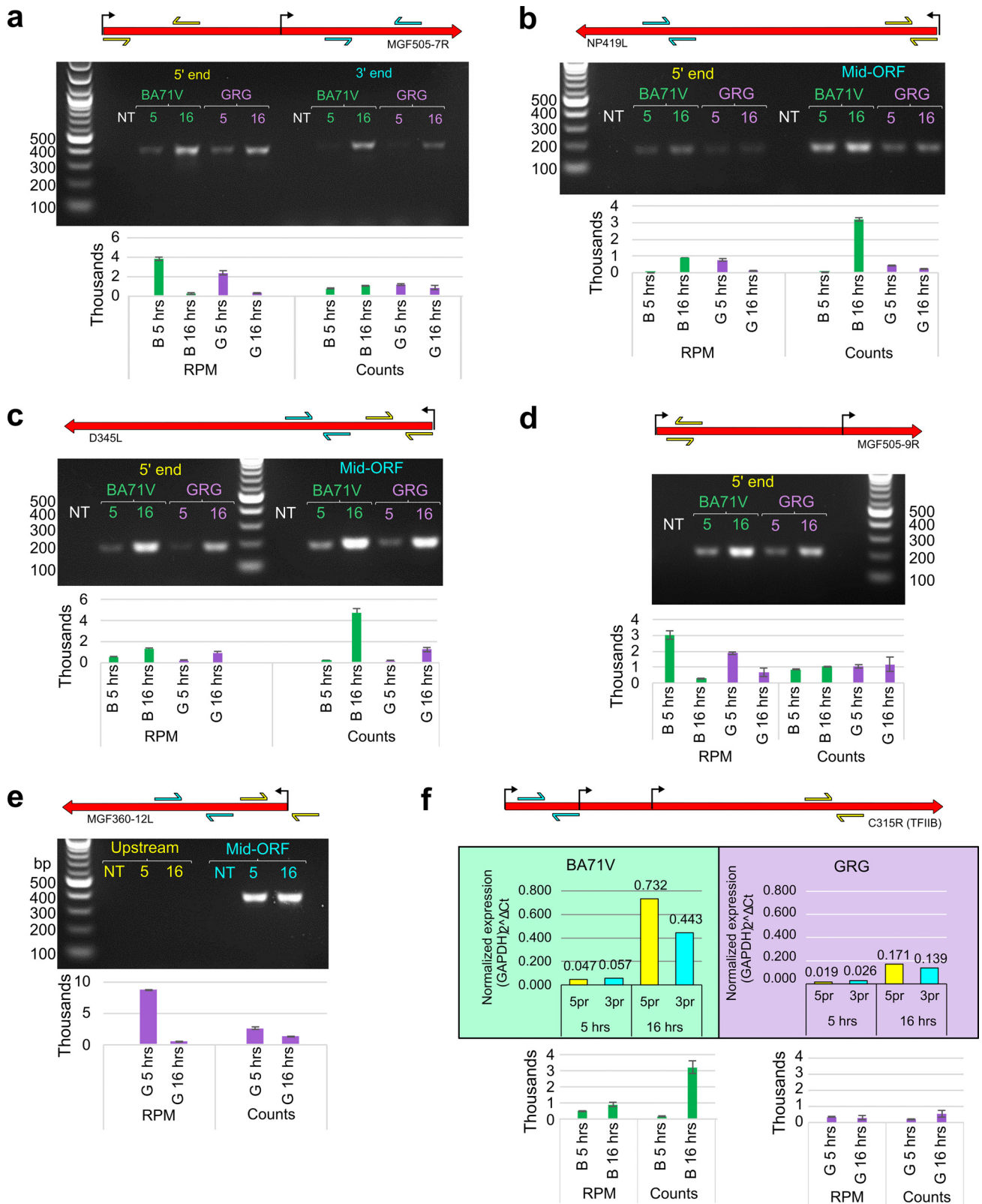


FIG 5 (a to f) RT-PCR results of genes for comparison to CAGE-seq data from (a) MGF 505-7R, (b) NP419L, (c) D345L, (d) MGF 360-12L, (e) MGF 505-9R, and (f) qRT-PCR results of C315R (ASFV-TFIIB) (NT, no template control). For each panel at the top is a diagrammatic representation of each gene's TSSs (bent arrow, including both pTSS and iOTSSs), and annotated ORF (red arrow); the arrow pairs in cyan or yellow represent the primers used for PCR (see Materials and Methods for primer sequences). Beneath each PCR result are bar charts representing the CAGE-seq results as either normalized (mean RPM) or raw (mean read counts) data; error bars show the range of values from each replicate.

detect transcripts originating from readthrough of transcripts initiated from upstream TSSs, including intra-ORF TSSs (ioTSSs). To detect such “contamination,” we used multiple primer combinations in upstream and downstream segments of the gene (Fig. 5c, cyan and yellow arrows) to capture and account for possible variations. Overall, our comparative analyses show that the normalized data (RPM) of early genes such as MGF 505-7R and 9R indeed skew and overemphasize their early expression, while the raw counts are in better agreement with the mRNA levels detected by RT-PCR. In contrast, late genes such as NP419L and D345L would be categorized as late using all three quantification methods, in agreement with GRG CAGE-seq but not BA71V from Fig. 3c. We validated the expression pattern of the early GRG-specific gene MGF 360-12L (Fig. 5e). While the RPM values indicated a very strong decrease in mRNA levels from early to late time points, the decrease in raw counts was less pronounced and more congruent with the RT-PCR analysis, showing a specific signal with nearly equal intensity during early and late infection. Lastly, we used qRT-PCR to quantify C315R transcript levels, as this was close to the early versus late threshold, (a log 2-fold change of 0 in Fig. 3c), which showed again that qRT-PCR better agreed with the raw counts.

An improved temporal classification of ASFV genes. Based on the considerations described above, we prepared a revised classification of temporal gene expression of the genes conserved between the two strains based on raw counts. The heatmap in Fig. 6a shows the mRNA levels at early and late infection stages of BA71V and GRG strains (all in duplicates) with the genes clustered into five subcategories (Fig. 6a, 1 to 5) according to their early and late expression pattern, which are shown in Fig. 6b. Genes that are expressed at high or intermediate levels during early infection, but that also show high or intermediate mRNA levels during late infection, are classified as early genes belonging to cluster-1 (8 genes, levels high to high [H-H]), cluster-4 (33 genes, mid to mid [M-M]) and cluster-5 (16 genes, low-mid to low-mid [LM-LM]). Genes with low or undetectable mRNA levels during early infection, which increase to intermediate or high levels during late infection are classified as late genes and belong to cluster-2 (15 genes, low to high [L-H]) and cluster-3 (60 genes, low to mid [L-M]), respectively. Overall, the clustered heatmap based on raw counts shows a similar but more emphasized pattern compared to the normalized (RPM) data (compare Fig. 6 and Fig. S1). Calculating the percentage of reads per gene, which can be detected at 16 hpi compared to 5 hpi, reveals that only a small number of genes have most ($\geq 70\%$) of their reads originating during early infection, 30 genes in the GRG strain and 5 genes in the BA71V strain. For over half of the BA71V-GRG conserved genes, 90 to 100% of reads can be detected during late infection (Figure 6c). For all GRG genes, this generates a significant difference between the raw counts per gene between time points (Fig. 6d).

Below, we discuss specific examples of genes subcategorized in specific clusters. I73R is among the top 20 most-expressed genes during both early and late infection according to the normalized RPM values (Fig. 2c and d) and resides in cluster-1 (H-H) (Fig. 6a). While I73R is expressed during early infection, the mRNA levels remain high with $>1/3$ of all reads detected during late infection in both strains when calculated as raw counts (34% in GRG and 45% in BA71V). This new analysis firmly locates I73R in cluster-1 (H-H), and it is classified confidently as an early gene. Notably, our new approach results in biologically meaningful subcategories of genes that are likely to be coregulated; e.g. the eight key genes that encode the ASFV transcription system, including RNAP subunits RPB1 (NP1450L), RPB2 (EP1242L), RPB3 (H359L), RPB5 (D205R), RPB7 (D339L), and RPB10 (CP80R), the transcription initiation factor TBP (B263R), and the capping enzyme (NP868R) belong to cluster-4 (M-M), and transcription factors TFII5 (I243L) and TFII8 (C315R) belong to cluster-5 (LM-LM). The overall mRNA levels of cluster-4 and -5 genes are different but remain largely unchanged during early and late infection, consistent with the transcription machinery being required throughout infection. In contrast, the mRNAs encoding the transcription initiation factors D6 (D1133L) and A7 (G1340L) are only present at low levels during early but

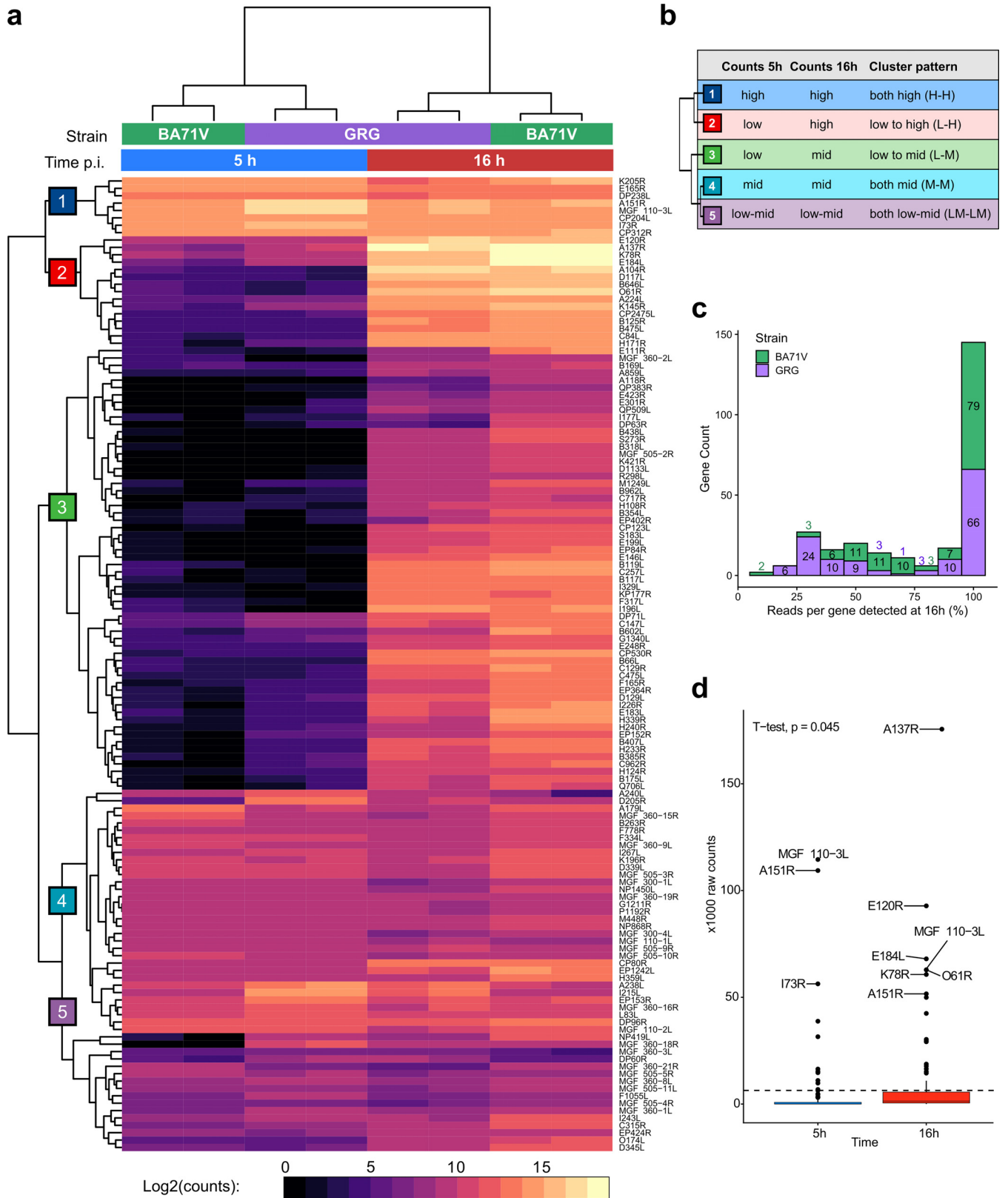


FIG 6 Comparison of the raw read counts for genes shared between BA71V and GRG. (a) Clustered heatmap representation of raw counts for genes shared between BA71V and GRG, generated with pheatmap. (b) Broad patterns represented by genes in the 5 clusters indicated in panel a. (c) Histogram showing the percentage of the total raw reads per gene which are detected at 16 hpi versus 5 hpi and comparing the distribution of percentages between GRG and BA71V. (d) Mean read counts from GRG at 5-hpi versus 16-hpi replicates, showing a significant increase (*t* test, *P* value = 0.045) from 5 hpi to 16 hpi.

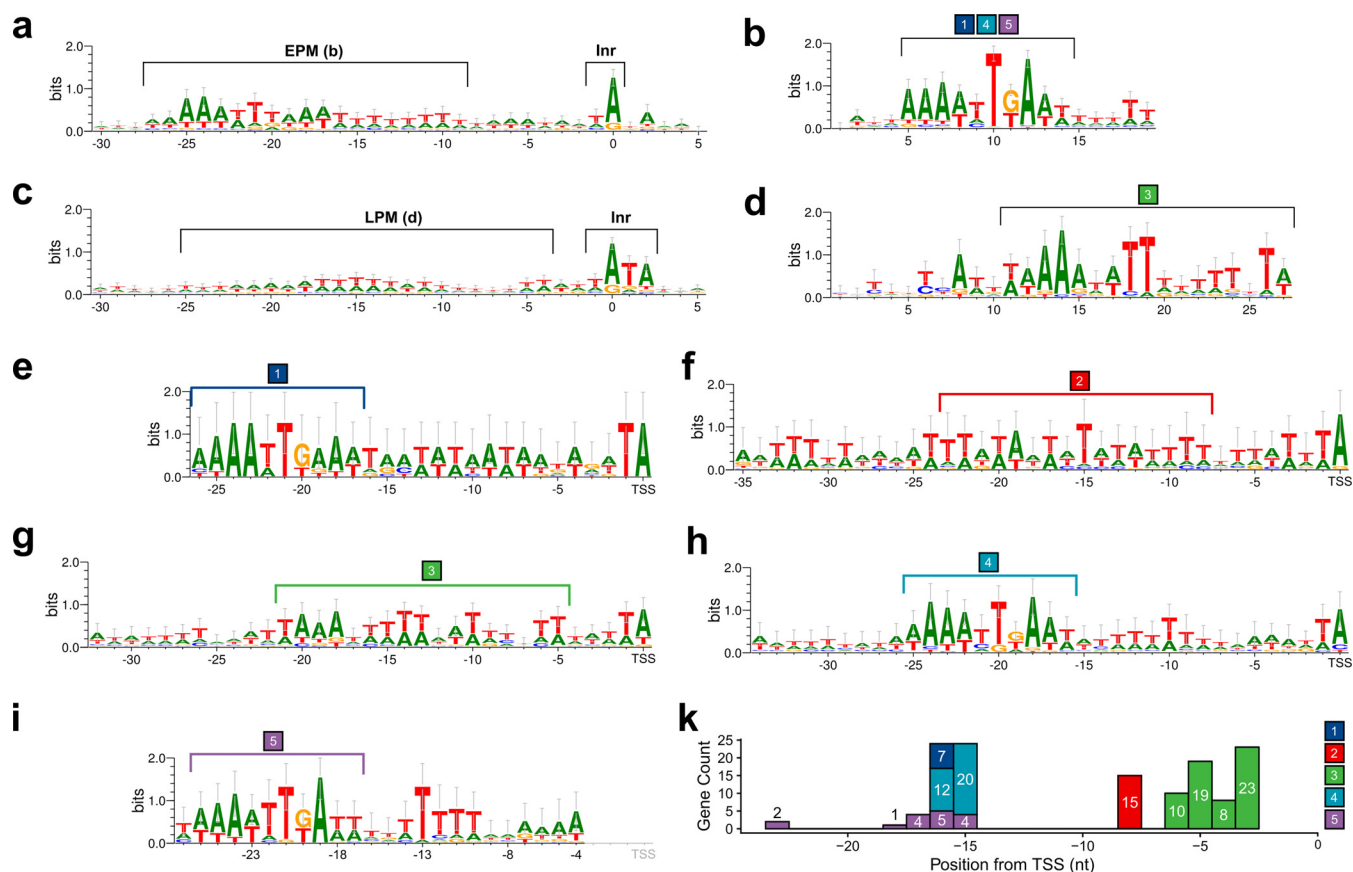


FIG 7 Promoter motifs and initiators detected in early and late ASFV GRG TSSs, including alternative TSSs and those for nORFs. (a) Consensus of 30 bp upstream and 5 bp downstream of all 134 early TSSs including nORFs, with the conserved EPM (10) and Inr annotated. (b) 30 bp upstream and 5 bp downstream of all 234 late gene and nORF TSSs, with the LPM and Inr annotated. (c) The conserved EPM detected via a MEME motif search of 35 bp upstream found upstream of 133 out of 134 early TSSs (E value = 3.1×10^{-69}). The conserved LPM detected via a MEME motif search of 35 bp upstream for 46 out of 234 late gene TSSs (E value = 2.6×10^{-003}). The locations of the EPM shown in panel b and the LPM shown in panel d are annotated with brackets in panels a and b, respectively. (e to i) Motifs detected via a MEME search of 35 bp upstream of genes in the following clusters from Fig. 6: cluster-1 (7 genes, E value = 9.1×10^{-12}), -2 (15 genes, E value = 2.6×10^{-48}), -3 (60 genes, E value = 1.0×10^{-167}), -4 (32 genes, E value = 4.7×10^{-105}), -5 (16 genes, E value = 5.7×10^{-036}), respectively. For ease of comparison, panels e, g, i, and f and h are aligned at the TSS position. All motifs were generated using WebLogo 3 (117). (k) The distribution of MEME motif-end distances, from the last nucleotide (in colored bracket), to their respective downstream TSSs.

increase during late infection and thus belong to cluster-3 (L-M), classifying them as late genes. This is meaningful since the heterodimeric D6-A7 factor is packaged into viral particles (7), presumably during the late stage of the infection cycle. The mRNAs of the major capsid protein p72 (B646L) and the histone-like-protein A104R (51, 53) follow a similar late pattern but are present at even higher levels during late infection and therefore belong to cluster-2 (L-H).

Architecture of ASFV promoter motifs. In order to characterize early promoter motifs (EPM) in the GRG strain, we extracted sequences 35 bp upstream of all early gene TSSs and carried out multiple sequence alignments. As expected, this region shows a conserved sequence signature in good agreement with our bioinformatics analyses of EPMs in the BA71V strain, including the correct distance between the EPM and the TSS (9 to 10 nt from the EPM 3' end) and the "TA" motif characteristic of the early gene initiator (Inr) element (Fig. 7a) (10). A motif search using MEME (54) identified a core EPM (cEPM) motif with the sequence 5'-AAAATTGAAT-3' (Fig. 7b) within the longer EPM. The cEPM is highly conserved and is present in almost all promoters controlling genes belonging to cluster-1, -4, and -5 (Table S3). A MEME analysis of sequences 35 bp upstream of late genes (Fig. 7c) provided a 17-bp AT-rich core late promoter motif (cLPM, Fig. 7d); however, this could only be detected in 46 of the late promoters.

In an attempt to improve the promoter motif analyses and deconvolute putative sequence elements further, we probed the promoter sequence context of the five

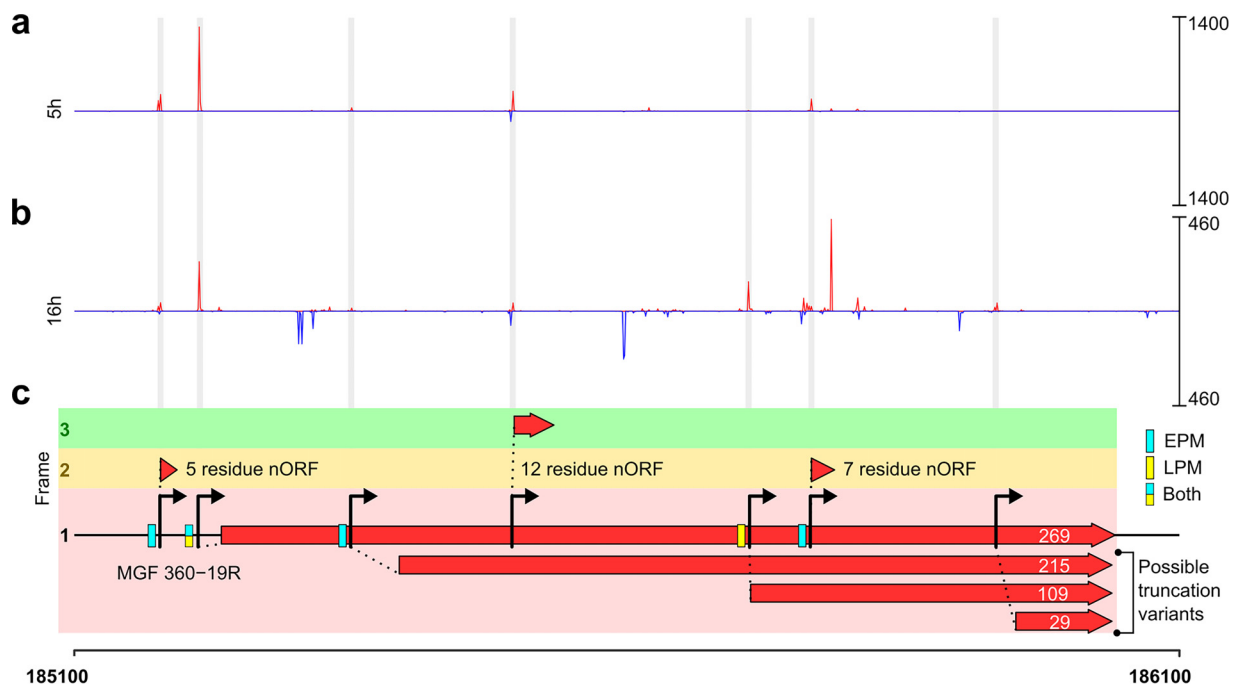


FIG 8 The TSSs of MGF 360-19R. Panels a (5 hpi) and b (16 hpi) show CAGE-seq 5'-end data from these time points. In red are reads from the plus strand, and blue shows reads from the minus strand; the RPM scales are on the right. (c) TSSs are annotated with arrows if they can generate a minimum of 5 residue-ORFs downstream, and gray bars indicate where they are located on the CAGE-seq coverage in panels a and b. ORFs identified downstream of TSSs are shown as red arrows (visualized with the R package gggenes), including three short nORFs out of frame with MGF 360-19R. Also shown are three in-frame truncation variants, from TSSs detected inside the full-length MGF 360-19R 269-residue ORF, downstream of its pTSS at 185213. Blue or yellow boxes upstream of TSSs indicate whether the EPM or LPM (respectively) could be detected within 35 nt upstream of the TSS using FIMO searching (118).

clusters (cluster-1 to cluster-5 in Fig. 7e to i, respectively) of temporally expressed genes with MEME (Table S3). The early gene promoters of cluster-1 (H-H), -4 (M-M), and -5 (LM-LM) are each associated with different expression levels, and all of them contain the cEPM located 15 to 16 nt upstream of the TSS, with two exceptions that are characterized by relatively low mRNA levels (Fig. 7k). Interestingly, cluster-2 (L-H) promoters are characterized by a conserved motif with significant similarity to a eukaryotic TATA box promoter element that binds the TBP-containing TFIID transcription initiation factor (Fig. 7f, highlighted with red bracket, detected via Tomtom [55] analysis of the MEME motif output). Cluster-3 (L-M) promoters contain a long motif akin to the cLPM, derived from searching all late gene promoter sequences, which is similar to the LPM identified in BA71V (Fig. 7d and g, green bracket). All motifs described in the cluster analysis above could be detected with statistical significance (P value < 0.05) via MEME in every gene in each respective cluster with only two exceptions—MGF 110-3L from cluster-1 and MGF 360-19R from cluster-4 (for the latter see details below).

Updating genome annotations using transcriptomics data. TSS annotation provides a useful tool for reannotating predicted ORFs in genomes like ASFV (10) where many of the gene products have not been fully characterized and usually rely on prediction from genome sequence alone. We have provided the updated ORF map of the GRG genome in GFF format (Table S1f). This analysis identified an MGF 360-19R ortholog (Fig. 8), demonstrating how transcriptomics enhances automated annotation of ASFV genomes by predicting ORFs from TSSs. The newly identified MGF 360-19R was included in subsequent DESeq2 analysis showing it was not highly or significantly differentially expressed (Table S1e). Another important feature is the identification of intra-ORF TSSs (ioTSSs) within MGF 360-19R that potentially direct the synthesis of N-terminally truncated protein variants expressed during either early or late infection. The presence of EPM and LPM promoter motifs lends further credence to the ioTSSs (Fig. 8). Similar truncation variants were previously reported for I243L and I226R (56)

and in BA71V (10). In addition, we detected multiple TSSs within MGF 360-19R encoding very short putative novel ORFs (nORF) 5, 7, or 12 amino acid (aa) residues long; since these ioTSSs were present in both early and late infection, they are not all likely to be due to pervasive transcription during late infection.

We investigated the occurrence of ioTSSs genome-wide and uncovered many TSSs with ORFs downstream that were not annotated in the GRG genome (Table S2a). These ORFs could be divided into the following subcategories: in-frame truncation variants (Table S2b, akin to MGF 360-19R in Fig. 8), nORFs (Table S2c), and simply mis-annotated ORFs. All updated annotations are found in Table S1f. Putative truncation variants generated from ioTSSs were predominantly identified during late infection, suggesting these could be a by-product of pervasive transcription. Therefore, those detected early or throughout infection are perhaps more interesting; they span a variety of protein functional groups, and many gene products are entirely uncharacterized (Fig. 9a). The truncation variants additionally showed a size variation of 5' UTRs between the ioTSSs and the downstream start codon (Fig. 9b). An example of a mis-annotation would be the CP204L (phosphoprotein p30, Fig. 9c) gene, which is predicted to be 201 residues long. The TSS determined by CAGE-seq and validated by rapid amplification of cDNA ends (5'-RACE) is located downstream of the annotated start codon; based on our results, we reannotated the start codon of CP204L which results in a shorter ORF of 193 amino acids (Fig. 9c).

Our GRG TSS map led to the discovery of many short nORFs, which are often overlooked in automated ORF annotations due to a minimum size, e.g., 60 residues in the original BA71V annotation (15). Some short ORFs have been predicted for the GRG genome, including those labeled "ASFV_G_ACD" in the Georgia 2007/1 genome annotation (19). However, their expression was not initially supported by experimental evidence, though we have now demonstrated their expression via CAGE-seq (Fig. 2b, Table S1e). We have now identified TSSs for most of these short ORFs, indicating at minimum that they are transcribed. As described above, we noted that TSSs were found throughout the genome in intergenic regions in addition to those identified upstream of the 190 annotated GRG ORFs (including MGF 360-19R, Table S2c). Our systematic, genome-wide approach identified 175 novel putative short ORFs. BLASTP (57) alignments showed that 13 were homologous to ORFs predicted in other strains, including DP146L and pNG4 from BA71V. We validated the TSSs for these candidates using 5'-RACE, which demonstrates the presence of these mRNAs and their associated TSSs at both time points (Fig. 9d and e, respectively), compared to our CAGE-seq data (Fig. 9f and g, respectively).

Putative single-SH2 domain protein-encoding genes in MGF 100. Our understanding of the ASFV genome is hampered by the large number of genes with unknown functions. We approached this problem by searching for conserved domains of uncharacterized MGF members *in silico*. MGF 100 genes form the smallest multigene family and include three short (100 to 150 aa) paralogs located at both genome ends (right, R and left, L)—1R, 1L (MGF_100-2L or DP141L in BA71V) and 3L (DP146L in BA71V). We predicted the two highly similar GRG ORFs I7L and I8L (51% sequence identity) to belong to the MGF 100 family (Fig. 10a), as designated in the Malawi LIL20/1 strain (58). Both I7L and I8L show similar overall transcript levels to the annotated MGF 100 members, 1L and 1R, though newly annotated MGF 100-3L (nORF_180573) was expressed at much higher levels. I7L and I8L are both early genes like MGF 100-3L, while MGF 100-1L and 1R are expressed late and not significantly changing, respectively (Table S1e). Several lines of evidence suggest that I7L and I8L play a role during infection. I7L and I8L are expressed early and at high levels; their deletion along with L9R, L10L, and L11L ORFs reduces virulence in swine (59), and their loss is associated with the adaptation of the GRG2007/1 strain to tissue culture infection (60). To gain insight into the function of MGF family members, including I7L and I8L, we generated computational homology models of MGF 100-1L -1R, I7L, and I8L using Phyre2 (61) (Fig. 10b). The structures selected by the algorithm for the modeling of MGF 100 proteins included suppressor of cytokine signaling proteins 1 and 2, and the PI3-kinase subunit alpha, all of which are characterized by Src homology 2 (SH2) domains (Fig. 10b and Table S2d). Canonical SH2

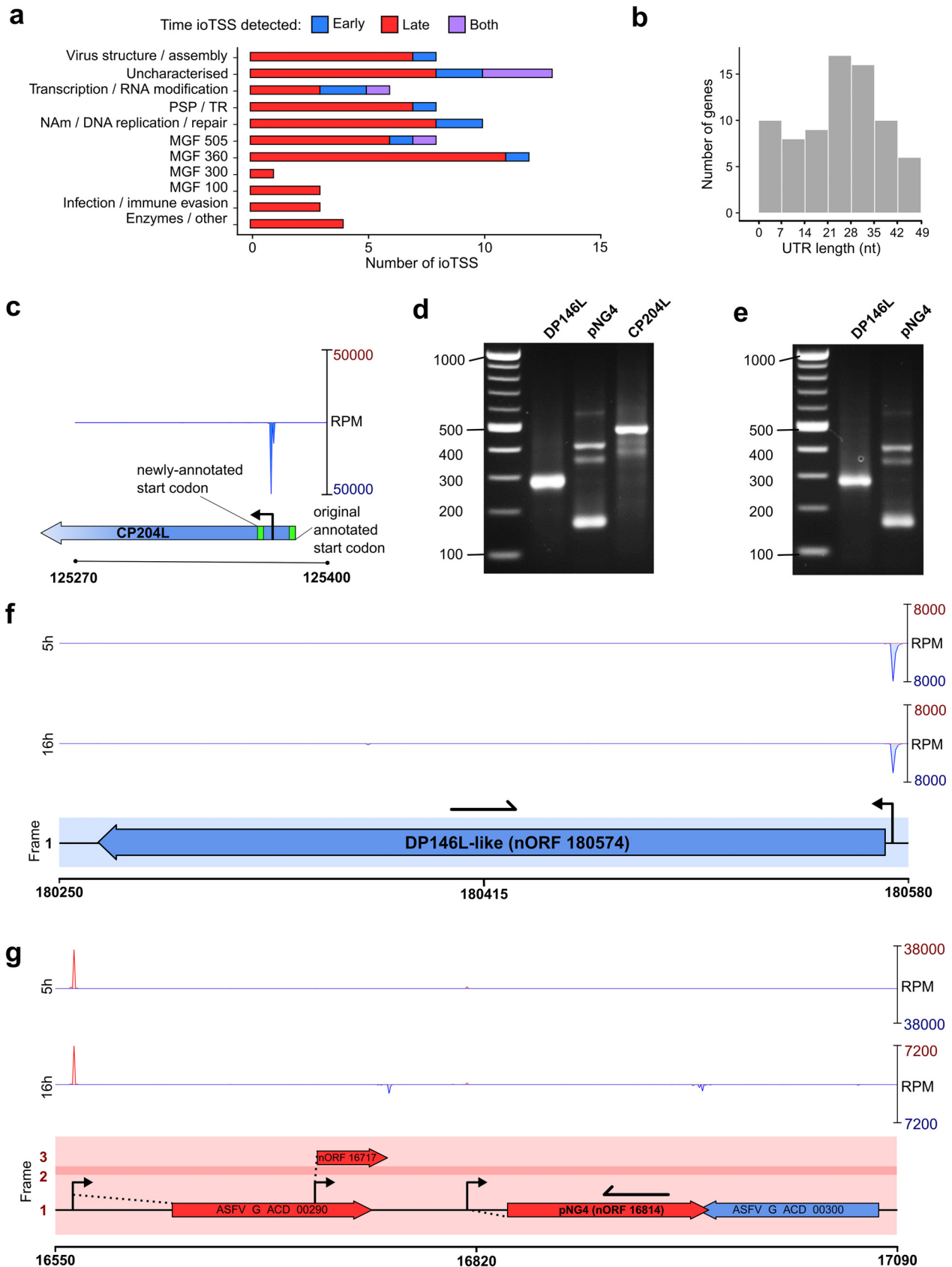


FIG 9 Summary of intra-ORF TSSs (ioTSSs) and nORFs detected in the GRG genome (further information is in Table S2). (a) Summary of the gene types in which ioTSSs were detected, showing an overrepresentation of MGFs, especially from families 360 and 505; furthermore, the majority of (Continued on next page)

domains bind to phosphorylated tyrosine residues and are an integral part of signaling cascades involved in the immune response (62). HHpred searches (63) predicted that, indeed, all MGF 100 members in BA71V and GRG include SH2 domains (Fig. 10c).

The response of the porcine macrophage transcriptome to ASFV infection. In order to evaluate the impact of ASFV on the gene expression of the host cell, we analyzed transcriptomic changes of infected porcine macrophages using the CAGE-seq data from the control (uninfected cells), 5 hpi, and 16 hpi. We annotated 9,384 macrophage-expressed protein-coding genes with CAGE-defined TSSs (Table S4). Although primary macrophages are known to vary largely in their transcription profile, the CAGE-seq reads were highly similar between RNA samples obtained from macrophages from two different animals in this study (Spearman's correlation coefficients ≥ 0.77).

As TSSs are not well annotated for the swine genome, we annotated them *de novo* using our CAGE-seq data with the RECLU pipeline. A total of 37,159 peaks could be identified, out of which around half (18,575) matched unique CAGE-derived peaks annotated in Robert et al. (64); i.e., they were located closer than 100 nt from the previously described peaks. Mapping CAGE-seq peaks to annotated swine protein-coding genes led to the identification of TSSs for 9,384 macrophage-expressed protein-coding genes (Table S4). The remaining 11,904 swine protein-coding genes did not have assigned TSSs, and therefore their expression levels were not assessed. The majority of genes were assigned with multiple TSSs, and these TSS-assigned genes corresponded to many critical functional macrophage markers, including genes encoding 56 cytokines and chemokines (including CXCL2, PPBP, CXCL8, and CXCL5 as the most highly expressed), 10 S100 calcium binding proteins (S100A12, S100A8, and S100A9 in the top expressed genes), as well as interferon and tumor necrosis factor (TNF) receptors (IFNGR1, IFNGR2, IFNAR1, IFNAR2, IFNLR1, TNFRSF10B, TNFRSF1B, TNFRSF1A, etc.) and typical M1/M2 marker genes such as TNF, ARG1, CCL24, and NOS2 (Table S5).

The 9,384 genes with annotated promoters were subjected to differential expression analysis using DESeq2 to compare the 5- and 16-h infected cell time points with control noninfected cells (c, 5, and 16) in a pairwise manner, i.e., between each condition. Expression of only 25 host genes was significantly deregulated between the control and 5 hpi, compared to 652 genes between 5 hpi and 16 hpi and 1,325 genes between mock-infected and 16 hpi (at a false-discover rate [FDR] of 0.05) (65). Based on the pairwise comparisons, we could distinguish major response profiles of the host genes. Late response genes, whose expression was significantly deregulated between both the uninfected control and 16 hpi, and 5 and 16 hpi, and early response genes, whose expression was significantly deregulated between the control and 5 hpi, but not 5 to 16 hpi (Fig. 11a). The latter category included only 20 genes, whereas more than 500 genes showed the late differentially regulated response; 344 genes were upregulated, and 180 genes were downregulated. The majority of the >9,000 genes analyzed therefore were not differentially regulated. Comparison of differences between expression levels in the different samples indicates that macrophage differentially expressed transcription programs change mostly between 5 and 16 hpi (Fig. 11b and c). The upregulated late response genes with the highest expression levels included several S100 calcium binding proteins. In contrast, expression of important cytokines (including CCL24, CXCL2, CXCL5, and CXCL8) significantly decreased from 5 hpi to 16 hpi (Fig. 11d).

To investigate the transcriptional response pathways and shed light on possible transcription factors involved in the macrophage response to ASFV infection, we searched for DNA motifs enriched in promoters of the four categories of deregulated genes in Fig. 11a.

FIG 9 Legend (Continued)

iotSSs are detected at 16 hpi. (b) For iotSSs in-frame with the original, the subsequent UTR lengths are summarized, i.e., the distance from the TSS to the next in-frame ATG start codon, which could generate a truncation variant. (c) Example of a mis-annotation for CP204L, whereby the pTSS is downstream from the predicted start codon. (d and e) Results of 5' RACE for three genes (DP146L, pNG4, and CP204L; see Materials and Methods for primers) at 5 hpi and 16 hpi, respectively. (f and g) Examples of genome regions around DP146L (f) and pNG4 (g), wherein iotSSs were detected with capacity for altering ORF length in subsequent transcripts and therefore protein output. Primers used for 5' RACE for DP146L and pNG4 are represented as black arrows in panel f and g, respectively.

a

| Genotype | ASFV strain | Pathogenic | Source | 1R | 1L/2L (DP141L) | 3L (DP146L) | I7L | I8L |
|----------|----------------------|------------|--------------------|----|-------------------|----------------|-----|-----|
| I | BA71V | N | NC_001659.2 | | ← ← | | | |
| I | BA71 | Y | NC_044942 | // | ← ← | ← ← | ← ← | ← ← |
| I | Portugal, OURT88/3 | N | AM712240.1 | → | // | ← ← | ← ← | ← ← |
| I | Benin 97/1 | Y | AM712239.1 | | ← ← | ← ← | ← ← | ← ← |
| II | Georgia 2007/1 | Y | FR682468.2 | → | // | ← ← | ← ← | ← ← |
| II | Georgia 2007/1-VP110 | N | Krug <i>et al.</i> | → | // | ← ← | | |
| II | China/2018/AnhuiXCGQ | Y | MK128995.1 | → | // | ← ← | ← ← | ← ← |
| IX | Ken05/Tk1 | Y | KM111294.1 | → | // | ← ← | ← ← | ← ← |
| X | Kenya 1950 | Y | AY261360.1 | → | // | ← ← | ← ← | ← ← |

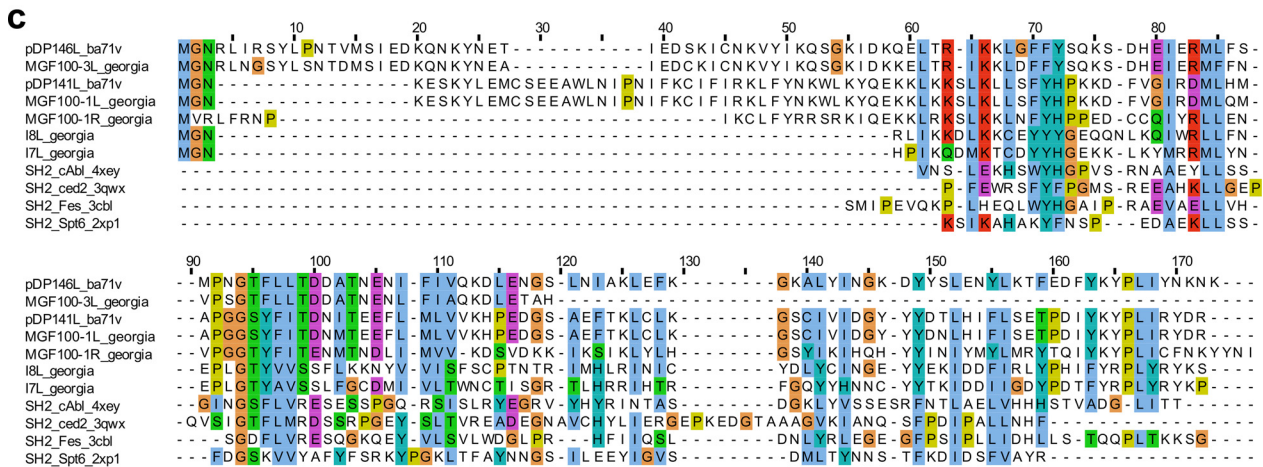
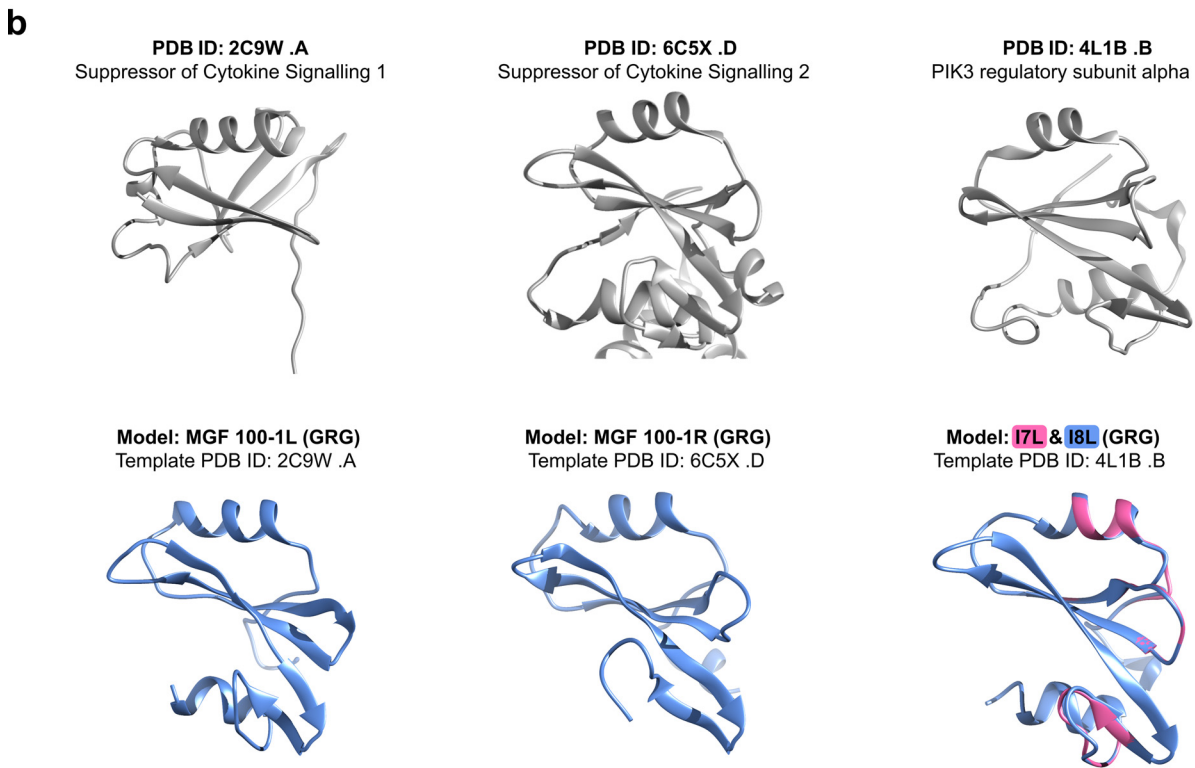


FIG 10 MGF 100 genes likely encode SH2-domain factors. (a) Occurrence of MGF 100 genes in selected ASFV strains, with genotype and pathogenicity indicated (as yes [Y] or no [N]). “1L/2L” refers to the genes MGF 100-2L (DP141L in BA71V) and MGF 100-1L in the GRG genome (Continued on next page)

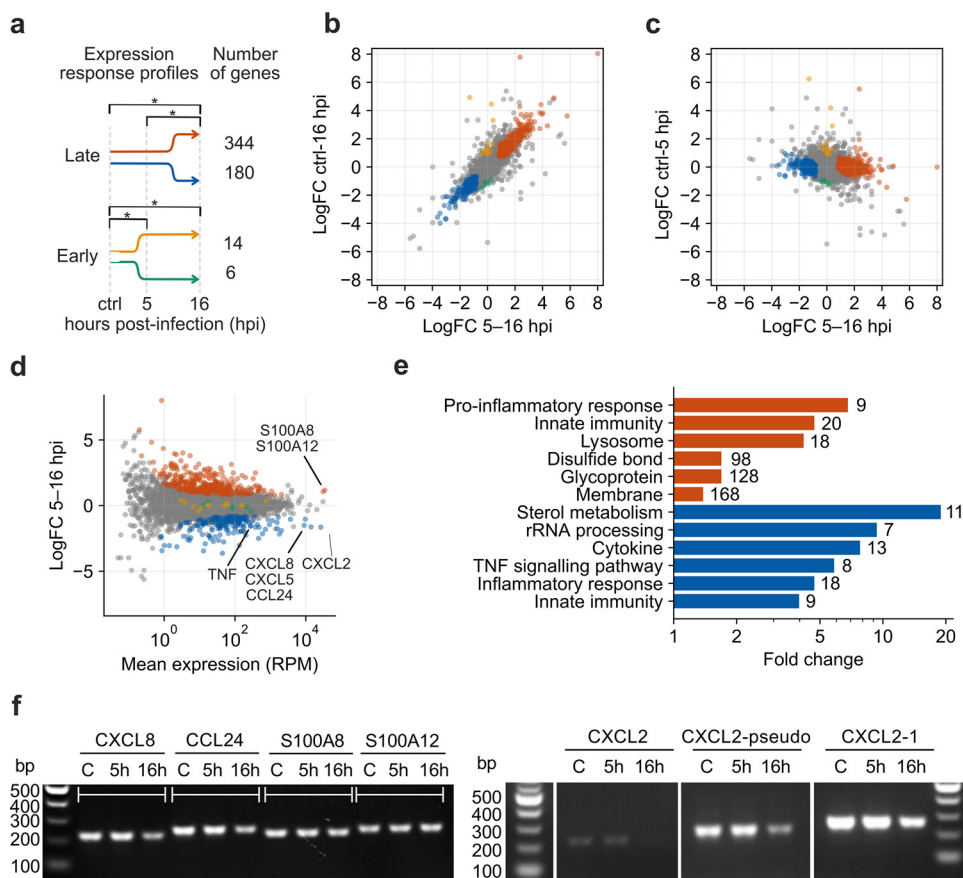


FIG 11 Changes in the swine macrophage transcriptome upon ASFV GRG infection. (a) Major expression response profiles of the pig macrophage transcriptome. Late response genes are significantly deregulated (false-discovery rate, < 0.05) in one direction both between mock-infected (ctrl) and 16 hpi as well as between 5 and 16 hpi, but not between mock-infected and 5 hpi. Early response genes are significantly deregulated in one direction both between ctrl and 5 hpi as well as ctrl and 16 hpi, but not between 5 and 16 hpi. (b) Relationship of log fold changes (logFC) of TSS-derived gene expression levels of the total 9,384 swine genes expressed in macrophages between 5 and 16 hpi and ctrl and 16 hpi. Colors correspond to the response groups from panel a. (c) Relationship of log fold changes of TSS-derived gene expression levels of the total 9,384 swine genes expressed in macrophages between 5 and 16 hpi and ctrl and 5 hpi. (d) MA plot of the TSS-derived gene expression levels between 5 and 16 hpi based on differential expression analysis with edgeR (111, 119). (e) Representative overrepresented functional annotations of the upregulated (red) and downregulated (blue) macrophage genes following late transcription response (Benjamini-corrected *P* value lower than 0.05). Numbers on the right of the bars indicate the total number of genes from a given group annotated with a given annotation. (f) RT-PCR of four genes of interest indicated in panel d. "C" is the uninfected macrophage control; "NTC" is the nontemplate control for each PCR, excluding template DNA. See Materials and Methods for the primers used.

Both late response promoter sets were significantly enriched with motifs, some of which contained submotifs known to be recognized by human transcription factors (Fig. S2). The highest-scored motif found in promoters of upregulated genes contained a submotif recognized by a family of human interferon regulatory factors (IRF9, IRF8, and IRF8; (Fig. S2a) that play essential roles in the antiviral response. Interestingly, both upregulated and downregulated promoters (Fig. S2b and c, respectively) were enriched with extended RELA/p65 motifs. p65 is a Rel-like domain-containing subunit of the NF- κ B complex, regulated by I-kappa-B, whose analog is encoded by ASFV. This pathway is a known target for ASFV in controlling host transcription (66–69).

FIG 10 Legend (Continued)

(GenBank accession number [FR682468.1](#)) annotation. (b) The top panel illustrates representative SH2 domain structures (suppressor of cytokine signaling 1 and 2 and the PI3K alpha), and the bottom panel shows structural homology models of MGF 100 members 1L, 1R, and I7L and I8L superimposed. The Phyre2 algorithm (56) was used to predict models for MGF 100 members (Table S2d), and the structures at the top were detected as the top hits for each of the MGF 100 models shown in the lower panel. (c) Structure-guided multiple sequence alignment of selected MGF 100 member models, alongside known SH2 domain structures (annotated as SH2_name_PDB number).

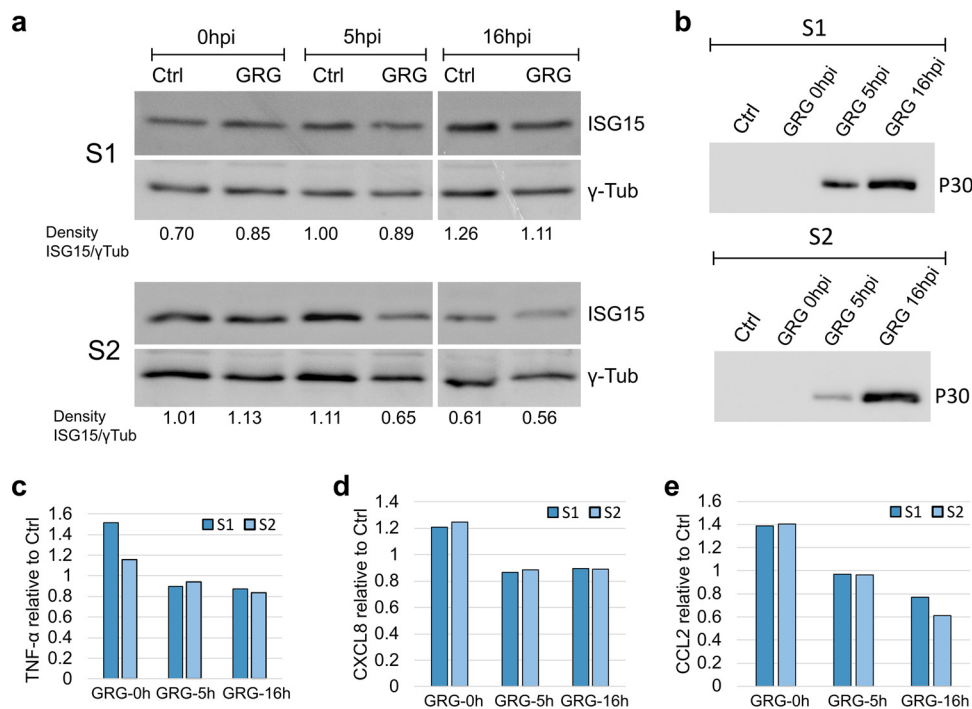


FIG 12 Protein expression at different times during infection of swine macrophages with ASFV-GRG. (a and b) Two different batches of macrophages (S1 and S2) were infected with an MOI of 5 or left uninfected as a control (Ctrl), and at 0, 5, and 16 hpi cellular extracts were collected and analyzed via SDS-PAGE Western blotting for the presence of ISG15 and γ -tubulin as a protein loading control (a) and for the presence of viral protein P30 as control of ASFV infection (b). (c to e) Results from ELISAs for detection of porcine TNF- α , IL-8/CXCL8, and CCL2/MCP-1, respectively, in culture supernatants. Results are presented as “relative to control” values (y axis of panels c to e) calculated by performing ELISAs in parallel for control and GRG infection at each time point.

To understand functional changes in the macrophage transcriptome, we also performed gene set enrichment analysis using annotations of human homologs. The top enriched functional annotations in the upregulated late response genes include glycoproteins and disulfide bonds, transmembrane proteins, and innate immunity, as well as positive regulation of inflammatory response (Fig. 11e). In contrast, sterol metabolism, rRNA processing, cytokines, TNF signaling pathway, and inflammatory response, as well as innate immunity were the top enriched functional clusters among the downregulated late response genes. Interestingly, the genes associated with innate immunity appear overrepresented in both up- and downregulated gene subsets, yet cytokines are 8-fold enriched only in the downregulated genes. The mRNA levels of genes of interest were additionally verified using RT-PCR (Fig. 11f).

Protein expression of selected genes. In order to determine whether the regulation exerted by GRG on host transcription of immunomodulatory genes could also translate to protein levels, we selected representative proteins whose genes showed significant changes. ISG15 expression, part of the antiviral response genes of the type I IFN stimulation pathway, was measured with Western blotting (Fig. 12a), with ASFV infection being monitored via P30 levels (Fig. 12b). Cytokines released from infected protospacer-adjacent motif (PAMS) were quantified using enzyme-linked immunosorbent assay (ELISAs) for pig cytokines, TNF- α , CXCL8, and CCL2 (Figure 12c, d, and e, respectively). As shown in Fig. 12, the release/expression for all the tested proteins during GRG infection were similar or decreased in comparison to the control uninfected cells at both 5 hpi and 16 hpi, while the production of viral protein P30 increased, confirming an effective viral infection.

DISCUSSION

In order to shed light on the gene expression determinants for ASF virulence, we focused our analyses on the similarities and differences in gene expression between a

highly virulent Georgia 2007/1 isolate and a nonvirulent, lab-adapted strain, BA71V. Previous annotation identified 125 ASFV ORFs that are conserved between all ASFV strain genomes irrespective of their virulence (16). They represent a core set of genes required for the virus to produce infectious progeny and include gene products such as those involved in virus genome replication, virion assembly, RNA transcription, and modification. These genes are located in the central region of the genome (Fig. 1a). Besides such essential genes, about one-third are nonessential for replication but have roles in evading host defense pathways. Some genes are conserved between isolates but are not necessarily essential core genes, for example, the apoptosis inhibitors Bcl-2 family member A179L and IAP family member A224L (70). Other nonessential genes, especially MGF members, vary in number between isolates. Our transcriptomic analysis captured TSS signals from 119 genes both shared between the BA71V and GRG genomes, which also matched expression patterns during early and late infection, according to CAGE-seq (Fig. 3 and 4a to c). Outliers include DP148R, which is unsurprising, given its promoter region is deleted in BA71V, and its coding region is interrupted by a frameshift mutation; therefore, functional protein expression is unlikely. DP148R is a nonessential, early-expressed virulence factor in the Benin 97/1 strain (49), consistent with our GRG data. Many additional GRG genes, lost from BA71V, are MGFs, which are mostly upregulated during early infection and located at the ends of the linear genome (Fig. 1a). MGFs have evolved on the virus genome by gene duplication and do not share significant similarity to other proteins, though some conserved domains, including ankyrin repeats, are present in some MGF 360 and 505 family members (17, 19).

Using advanced sequence searches and computational homology modeling, we predict the members of the MGF 100 family to encode SH2 domains, including I7L and I8L. Although SH2 domains are primarily specific to eukaryotes, rare cases of horizontally transferred SH2 domains found in viruses are implicated in hijacking host cell pTyr signaling (71). A large family of “superbinding” SH2 domains were discovered in *Legionella*. Its members, including single SH2 domain proteins, are likely effector proteins during infection (72). We also identified another MGF 100 member in the GRG genome as one of our nORFs, a partial 100-residue copy of DP146L (MGF 100-3L) (Table S2c). Unlike its annotated MGF 100-1L and MGF 100-1R cousins, it was downregulated from 5 hpi to 16 hpi (Table S1e). Together with I7L and I8L, GRG encodes a total of 5 MGF 100 genes (Fig. 10a). Interestingly, loss of MGF 100 members was observed during the process of adapting a virulent Georgia strain to grow in cultured cell lines (60). Deletion of MGF 100-1R, from a virulent genotype II Chinese strain (73), or of I8L from Georgia 2010 was shown not to reduce virulence of the virus in pigs or reduce virus replication in porcine macrophages (74). However, simultaneous deletion of genes I7L, I8L, I9L, I10L, and I11L from a Chinese virulent isolate reduced virulence, and surviving pigs were protected against challenge (59). In summary, although deletion of some individual MGF 100 genes does not lead to attenuation, deletion of I7L and I8L, in combination with I9L, I10L, and I11L did have an impact.

The Georgia 2007/1 genome was recently resequenced, which identified a small number of genome changes affecting mapped ORFs and identified new ORFs (18). Adjacent to the covalently cross-linked genome termini, the BA71V genome contains terminal inverted repeats of >2 kbp, in which two short ORFs were identified (DP93R, DP86L). These were not included in previous GRG sequence annotations; however, our nORFs included a 55-residue homolog of DP96R, which was a late, but not highly expressed, gene. These are yet further examples of how transcriptomics aid in improving ASFV genome annotation. Functional data are available for only a few proteins coded by ORFs not conserved between BA71V and GRG. This includes the p22 protein (KP177R), which is expressed on the cell membrane during early infection and also incorporated into the virus particle inner envelope. The function of the KP177R-like GRG gene I10L has not been studied but may provide an antigenically divergent variant of P22, enabling evasion of the host immune response (19). We found that KP177R was highly expressed at 16 hpi, while I10L was also expressed late, but at much lower

levels. Their function is unknown, though the presence of an SH2 domain indicates possible roles in signaling pathways (7, 19, 75).

MGF 110 members are among the most highly expressed genes during early infection both in GRG (this study) and in BA71V (10), suggesting great importance during infection, at least in porcine macrophages and Vero cells, respectively. However, MGF 110 remains poorly characterized, and 13 orthologues were identified thus far, with the numbers present varying between isolates (30). MGF 110 proteins possess cysteine-rich motifs, optimal for an oxidizing environment as found in the endoplasmic reticulum (ER) lumen or outside the cell, and MGF 110-4L (XP124L) contains a KDEL signal for retaining the protein in the ER (76). Since highly virulent isolates have few copies of these genes (for example, only 5 in the Benin 97/1 genome), it was assumed they are not important for virulence in pigs (17), but their high expression warrants further investigation, which has recently begun in the form of deletion mutants. For example, deletion of MGF 110-9L from a Chinese genotype II virulent strain, reduced virulence (35), whereas deletion of MGF 110-1L from Georgia 2010 (77) did not substantially affect virulence.

There is, however, good evidence that MGF 360 and 505 carry out important roles in evading the host type I interferon (IFN) response—the main host antiviral defense pathway (37). Evidence for the role of MGF 360 and 505 genes in virulence was obtained from deletions in tissue-culture adapted and field-attenuated isolates, as well as targeted gene deletions. This correlated with induction of the type I interferon response, which itself is inhibited in macrophages infected with virulent ASFV isolates (32, 38, 39). Deletions of these MGF 360 and 505 genes also correlated with an increased sensitivity of ASFV replication to pretreatment of the macrophage cells with type I IFN (40). Thus, the MGF 360 and 505 genes have roles in inhibiting type I IFN induction and increasing sensitivity to type I IFN. However, it remains unknown if these MGF 360 and MGF 505 genes act synergistically or if some have a more important role than others in type I IFN suppression. Our DESeq2 analysis did show that members of both these families showed very similar patterns of early expression (Fig. 2 and 3), conserved cEPM-containing promoters, and almost exclusive presence in clusters-1 (H-H), -4 (M-M), and -5 (LM-LM) (Fig. 6 and 7), consistent with ASFV prioritizing inhibition of the host immune response during early infection.

An interesting pattern which emerged during our CAGE-seq analysis was the clear prevalence of iOTSSs within ORFs, especially in MGFs (Fig. 8 and 9). However, it is not clear whether subsequent in-frame truncation variants generate stable proteins, nor what their function could be. Perhaps even more interesting was the discovery of 176 nORFs (including MGF 360-19R) with clear TSSs according to CAGE-seq, highlighting the power of transcriptomics to better annotate sequenced genomes. We were able to detect previously unannotated genes from other strains, and partial duplications of genes already encoded in GRG (Table S2).

The increase in transcription across the ASFV genome during late infection (10) appears ubiquitous. At least 50 genes have previously been investigated in single gene expression studies using Northern blotting or primer extension (for review see references 10 and 78). Transcripts from over two-thirds of these genes were detected during late infection, and a quarter had transcripts detected during both early and late infection. Therefore, clear evidence using several techniques now supports this increase in ASFV transcripts at late times postinfection. It is not entirely clear whether it is due to pervasive transcription, high mRNA stability, or a combination of factors. However, there is a correlated increase in viral genome copies, potentially available as templates for pervasive transcription. The increase in genome copies is more pronounced in BA71V than in GRG, which likewise is reflected in the increase in transcripts during late infection (Fig. 4).

Our transcriptomic analysis of the porcine macrophage host revealed 522 genes whose expression patterns significantly changed between 5 and 16 h postinfection (Fig. 11a), and only 20 genes were found to change between the control cells and those infected for

5 hpi. In aggregate, this reflects a relatively slow host response to ASFV infection following expression of early ASFV genes. We observed mild downregulation of some genes, e.g., ACTB coding for β -actin, eIF4A, and eIF4E (Table S5), resembling patterns previously shown by RT-qPCR (79). The macrophage transcriptome mainly shuts down immunomodulation between 5 hpi to 16 hpi postinfection; cytokines appeared highly expressed at 5 hpi but downregulated from 5 hpi to 16 hpi. Of the 54 cytokine genes we detected, expression of 13 was decreased—four interleukin genes (interleukin-1A [IL1A], IL-1B, IL-19, IL-27), four proinflammatory chemokines (CCL24, CXCL2, CXCL5, CXCL8), and tumor necrosis factor (TNF) genes. Since inflammatory responses serve as the first line of host defense against viral infections, viruses have developed ways to neutralize host proinflammatory pathways. ASFV encodes a structural analog of I κ B, A238L, which was proposed to act as a molecular off-switch for NF- κ B-targeted pro-inflammatory cytokines (68). In our study, A238L is one of the most expressed ASFV genes at 5 hpi, but is significantly downregulated afterward (Fig. 2c). Accordingly, swine homologs of human NF- κ B target genes were significantly overrepresented (3.8-fold) among downregulated macrophage genes (Fisher's exact P value < 1e-5, based on human NF- κ B target genes from <https://www.bu.edu/nf-kb/gene-resources/target-genes/>). Downregulated genes include interleukins 1A, 1B, 8, and 27 (IL1A, IL1B, CXCL8, IL-27), TNF, as well as a target for common nonsteroidal anti-inflammatory drugs, prostaglandin-endoperoxide synthase 2 (PTGS2 or COX-2). Interestingly, promoters of both up- and downregulated genes (Fig. S2) contained a motif with the sequence preferentially recognized by the human p65-NF- κ B complex (80). Expression of TNF, a well-known marker gene for acute immune reaction and M1 polarization, was recorded at a high level in control samples and at 5 hpi but significantly dropped at 16 hpi. It has been already shown that ASFV inhibits transcription of TNF and other proinflammatory cytokines (68). On the other hand, the downregulation of TNF stands in contrast to previous results from ASFV-E75 strain-infected macrophages *in vitro*, where TNF expression increased significantly after 6 hpi (81). Therefore, the different time courses of TNF expression induced by the moderately virulent E75 and more virulent Georgia strain may reflect different macrophage activation programs (82).

We investigated if the modulation of transcription we observed by CAGE-seq during GRG infection of PAMS was also observed at the protein level. We analyzed the secretion or expression of different immunomodulators (cytokines CCL2, CXCL8, TNF- α , and interferon-stimulated gene ISG15) at different times following infection of PAMS. We confirmed that the infection did not lead to an increase of these mediators at either 5 h or 16 h of infection. Secretion or expression of these proteins was similar or slightly decreased in infected cells in comparison to control noninfected cells. The results indicated that the control by virulent Georgia 2007/1 of host cell responses to infection we observed at the transcription level can lead to control also at the level of the protein production. Interestingly, CCL2 transcription was somewhat upregulated at late infection (Table S5), whereas its protein release to the supernatant was decreased (Fig. 12e). ASFV has been shown to prioritize expression of its encoded proteins by sequestering components of the host translation machinery to viral factories (83). The levels or functions of host proteins may also be modulated by targeting for posttranslational modification or degradation (83–85). Therefore, in addition to control at the transcriptional level ASFV, may modulate the production of immunomodulatory host proteins at a later step, as seems to occur for CCL2, a known chemoattractant for myeloid and lymphoid cells (86), which could be an important target for regulation by ASFV.

Four S100 family members are among the host genes that are upregulated after 5 hpi (Fig. 11b), including S100A8, S100A11, S100A12, and S100A13. S100A8 and S100A12 are among the most highly expressed genes on average throughout infection. S100 proteins are calcium-binding cytosolic proteins that are released and serve as a danger signal and stimulate inflammation (87). Once released from the cell, S100A12 and S100A8 function as endogenous agonists to bind TLR4 and induce apoptosis and autophagy in various cell types (87). S100A8 and S100A9 were also found in the RNA-seq whole-blood study as the

top upregulated upon infection of the pigs with Georgia 2007/1, but not of a low-pathogenic ASFV isolate OURT 88/3 (43).

Previous studies described global swine transcriptome changes upon ASFV infection using short read sequencing (Illumina), including the RNA-seq described above (43) and a microarray study of primary swine macrophage cell cultures infected with the GRG strain, at six time points postinfection (42). Although these varied in designs and selected methods, results of both these works give some indication of the main host immune responses and how ASFV could evade them. The latter microarray study indicated similar suppression of inflammatory response after 16 hpi as we observed in this study, with expression of many cytokines downregulated relative to noninfected macrophages (42). More recently, there have been several transcriptomic studies using classical RNA-seq of ASFV infections from Chinese isolates (44–46). Fan et al. (44) investigated the transcriptomic and proteomic response within tissues of pigs following ASFV infection and death, though this was not directly comparable to our own analysis in PAMs, due to their observations being of a far later infection stage (postmortem) than our 16-h time point. The two studies most comparable to ours were carried out on a Chinese genotype II pathogenic strain during infection of PAMs. Ju et al. (45) investigated 6, 12, and 24 hpi, while Yang et al. (46) investigated 12, 24, and 36 hpi. However, comparison of the overlapping time points of 12 hpi and 24 hpi did not yield similar host gene expression changes, possibly due to variation among primary macrophages or due to the low MOI of 1 used in both studies. In summary, these differences highlight that our understanding of the host-virus relationship during ASFV infection is still not well understood, and further work is needed to understand why such substantial variation in host gene expression can arise.

A further important note is that all of the studies described above used classical RNA-seq-based methods, the nucleotide resolution of which is not sufficient to investigate differential expression of both the virus and host simultaneously. Investigating the viral transcriptome is especially difficult in a compact genome like that of ASFV, where transcription readthrough can undermine results from classical RNA-sequencing techniques (10, 88). A recent investigation into ASFV RNA transcripts using a long-read-based Oxford Nanopore Technologies (ONT) method provides fascinating insight into their length and readthrough heterogeneity. This new method highlighted how misleading short-read sequencing with classical RNA-seq can be when quantifying ASFV gene expression, due to the abundance of readthrough occurring in ASFV, generating transcripts covering multiple viral ORFs. This study did, however, unfortunately lack the read coverage for in-depth analysis of host transcripts alongside that of viral transcripts (89–90).

Here, we have demonstrated that CAGE-seq is an exceptionally powerful tool for quantifying the relative expression of viral genes across the ASFV genome, as well as making direct comparison between strains for expression of shared genes, and further highlighting the importance of highly expressed but still functionally uncharacterized viral genes. CAGE-seq conveniently circumvents the issue in compact viral genomes like those of ASFV and VACV, of transcripts reading through into downstream genes which cannot be distinguished from classical short-read RNA-seq (10, 43, 91). Furthermore, it enables us to effectively annotate genome-wide the 5' ends of capped viral transcripts, and thus TSSs of viral genes, and subsequently their temporal promoters. This 5'-end resolution in ASFV is still not achievable via ONT long-read sequencing (89–90). We have now expanded on promoter motifs we previously described (Fig. 7) to identify 5 clusters of genes (Fig. 6) with distinct patterns of expression. Three of these clusters (cluster-1, high to high levels; -4, mid to mid; and -5, low-mid to low-mid) have slightly differing promoters, with a highly conserved core EPM. This is akin to the early gene promoter of VACV (88) for VETF recognition and early gene transcription initiation (13, 92, 93). We have found that late genes can be categorized into two types that increase either from low to extremely high expression levels (e.g., p72-encoding B646L) in cluster-2 or from low to medium expression levels in cluster-3 (e.g., VETF-encoding genes). The promoters of these genes show resemblance to

the eukaryotic TATA box (94) or the BA71V LPM (10), respectively. Our analysis additionally shows the potential for a variety of non-pTSSs—alternative ones used for different times in infection, iOTSSs which could generate in-frame truncation variants of ORFs, sense or antisense transcripts relative to annotated ORFs, and finally, TSSs generating nORFs, which predominantly have no known homologs.

In summary, it is becoming increasingly clear that the transcriptomic landscape of ASFV and its host during infection is far more complex than originally anticipated. Much of this raises further questions about the basal mechanisms underlying ASFV transcription and how it is regulated over the infection time course. Which subsets of initiation factors enable the RNAPs to recognize early and late promoters? Does ASFV include intermediate genes, and what factors enables their expression? What is the molecular basis of the pervasive transcription during late infection? The field of ASFV transcription has been understudied and underappreciated, and considering the severe threat that ASF poses for the global food system and food security, we now need to step up and focus our attention and resources to study the fundamental biology of ASFV to develop effective antiviral drugs and vaccines.

MATERIALS AND METHODS

GRG-infection of macrophages and RNA-extraction. Primary porcine alveolar macrophage cells were collected from two animals following approval by the local Animal Welfare and Ethical Review Board at The Pirbright Institute. Cells were seeded in 6-well plates (2×10^6 cells/well) with RPMI medium (with GlutaMAX), supplemented with 10% pig serum, 100 IU/mL penicillin, and 100 μ g/mL streptomycin. They were infected as 2 replicate wells for 5 hpi or 16 hpi with a multiplicity of infection (MOI) of 5 of the ASFV Georgia 2007/1 strain, while uninfected cells were seeded in parallel as a control (mock-infection). Total RNA was extracted according to the manufacturer's instructions for extraction with TRIzol lysis reagent (Thermo Fisher Scientific), and the subsequent RNAs were resuspended in 50 μ L RNase-free water and DNase-treated (Turbo DNase-free kit; Invitrogen). RNA quality was assessed via Bioanalyzer (Agilent 2100). Then, 5 μ g of each sample was ethanol-precipitated before being sent to CAGE-seq (Kabushiki Kaisha DNAFORM, Japan). Samples were named as follows: uninfected cells or "mock" (C1-ctrl and C2-ctrl), at 5 hpi postinfection (samples G1-5h and G2-5h), and at 16 hpi postinfection (G3-16h and G4-16h).

CAGE-sequencing and mapping to GRG and *Sus scrofa* genomes. Library preparation and CAGE-sequencing of RNA samples were carried out by CAGE-seq (Kabushiki Kaisha DNAFORM, Japan). Library preparation produced single-end indexed cDNA libraries for sequencing; in brief, this included reverse transcription with random primers and oxidation and biotinylation of the 5' mRNA cap, followed by RNase ONE treatment removing RNA not protected in a cDNA-RNA hybrid. Two rounds of cap-trapping using streptavidin beads washed away uncapped RNA-cDNA hybrids. Next, RNase ONE and RNase H treatment degraded any remaining RNA, and cDNA strands were subsequently released from the streptavidin beads and quality assessed via Bioanalyzer. Single-strand index linker and 3' linker were ligated to released cDNA strands, and primer containing the Illumina sequencer priming site was used for second-strand synthesis. Samples were sequenced using the Illumina NextSeq 500 platform, producing 76-bp reads. FastQC (95) analysis was carried out on all FASTQ files at Kabushiki Kaisha DNAFORM, and CAGE-seq reads showed consistent read quality across their read length and therefore were mapped in their entirety to the GRG genome (GenBank accession number [FR682468.1](#)) in our work using Bowtie2 (96) and to the *Sus scrofa* ([GCF_000003025.6](#)) genome with HISAT2 (96, 97) by Kabushiki Kaisha DNAFORM.

Transcription start site-mapping across the viral GRG genome. CAGE-seq mapped sample BAM files were converted to BigWig (BW) format with BEDtools (98) genomecov, to produce per-strand BW files of 5' read ends. Stranded BW files were input for TSS-prediction in RStudio (99) with the Bioconductor (100) package CAGEfightR (101). Genomic feature locations were imported as a TxDb object from the GRG genome (GenBank accession number [FR682468.1](#)) gene feature file (GFF3). CAGEfightR was used to quantify the CAGE reads mapping at base pair resolution to the GRG genome—at CAGE TSSs, separately for the 5-hpi and 16-hpi replicates. TSS values were normalized by tags-per-million for each sample and pooled, and only TSSs supported by presence in both replicates were kept. TSSs were assigned to clusters, if within 25 bp of one another, filtering out pooled, RPM-normalized TSS counts below 25 bp for 5-hpi samples, or 50 bp for 16 hpi, and assigned a "thick" value as the highest TSS peak within that cluster. A higher cutoff for 16 hpi was used to minimize the extra noise of pervasive transcription observed during late infection (10). TSS clusters were assigned to annotated the GRG genome (GenBank accession number [FR682468.1](#)) ORFs using BEDtools intersect, if its highest point (thick region) was located within 500 bp upstream of an ORF, "CDS" if within the ORF, or "NA" if no annotated ORF was within these regions. Multiple TSSs located within 500 bp of ORFs were split into subsets; the "Primary" cluster subset contained either the highest-scoring CAGEfightR cluster or the highest-scoring manually annotated peak (when manual ORF corrections were necessary), and the highest peak coordinate was defined as the primary TSS (pTSS) for an ORF. Further clusters associated with these ORFs were classified as "nonprimary," with their highest peak as a nonprimary TSS (npTSS). If the strongest TSS location was intra-ORF, without any TSSs located upstream of the ORF, then the ORF was manually redefined as starting from the next start codon ATG downstream.

DESeq2 differential expression analysis of GRG genes. For analyzing differential expression with the CAGE-seq data set, a GFF was created with BEDtools extending from the pTSS coordinate, 25 bp upstream and 75 bp downstream; however, in cases of alternating pTSSs, this region was defined as 25 bp upstream of the most upstream pTSS and 75 bp downstream of the most downstream pTSS. HTSeq-count (102) was used to count reads mapping to genomic regions described above for both the RNA- and CAGE-seq sample data sets. The raw read counts were then used to analyze differential expression across these regions between the time points using DESeq2 (default normalization described by Love et al. [47]), and those regions showing changes with an adjusted *P* value (padj) of <0.05 were considered significant. A caveat of this early or late definition is that it is a binary definition of whether a gene is up- or downregulated between conditions (time points), relative to the background read depth of reads which map to the genome in question. Further analysis of ASFV genes used their characterized or predicted functions, from the VOCS tool database (<https://4virology.net/>) (103, 104) entries for the GRG genome.

Quantification of viral genome copies at different time points of infection. Porcine lung macrophages were seeded and infected as described above. Vero cells were similarly cultured in 6-well plates in Dulbecco's modified Eagle's medium (DMEM) supplemented with 10% fetal calf serum, 100 IU/mL penicillin, and 100 μ g/mL streptomycin; when semiconfluent, they were infected with MOI 5 of Ba71V. Immediately after infection (after a 1-h adsorption period, considered 0 hpi) or at 5 hpi and 16 hpi, the supernatant was removed and nucleic acids were extracted using the QIAmp viral RNA kit (Qiagen) and quantified using a NanoDrop spectrophotometer (Thermo Fisher Scientific). For quantification of viral genome copy equivalents, 50 ng of each nucleic acid sample was used in qPCR with primers and probe targeting the viral capsid gene B646L. As previously described (105), standard curve quantification qPCR was carried out on an Mx3005P system (Agilent Technologies) using the primers CTGCTCATGGTATCA ATCTTATCGA and GATACCACAAGAT(AG)GCCGT and probe 5'-(6-carboxyfluorescein [FAM])-CCACGGG AGGAATACCAACCCAGTG-3'-(6-carboxytetramethylrhodamine [TAMRA]).

Analysis of mRNA levels by RT-PCR and quantitative real-time PCR (qPCR). RNA from GRG- or Ba71V-infected macrophages, or Vero cells, or from uninfected cell controls, was collected at the different time points postinfection with TRIzol, as described above. RNA was reverse transcribed (800 ng RNA per sample) using a SuperScript III first-strand synthesis system for RT-PCR and random hexamers (Invitrogen). For PCR, cDNAs were diluted 1:20 with nuclease-free water, and 1 μ L each sample was amplified in a total volume of 20 μ L using Platinum green hot start PCR master mix (Invitrogen) and 200 nM each primer. Annealing temperatures were tested for each primer pair in gradient PCR to determine the one optimal for amplification.

Table S7a shows the primers used for each gene target, the amplicon size, PCR conditions, and NCBI accession numbers for sequences used for primer design. PCRs were then performed with limited cycles of amplification to have a semiquantitative comparison of transcript abundance between infection time points (by not reaching the maximum product amplification plateau). Amplification products were viewed using 1.5% agarose gel electrophoresis.

C315R transcript levels were assessed by qPCR, and housekeeping gene glyceraldehyde-3-phosphate dehydrogenase (GAPDH) expression was used for normalization. Primer details and the qPCR amplification program are shown in Table S7b (GAPDH primers used for Vero cells were previously published by Melchjorsen et al. [106]). Primers were used at 250 nM concentration with Brilliant III ultrafast SYBR green QPCR master mix (Agilent Technologies; 600882), 1 μ L cDNA in 20 μ L (1:20) total reaction volumes, and qPCRs were carried out in an Mx3005P system (Agilent Technologies). Similar amplification efficiencies (97 to 102%) for all primers had been observed upon amplification of serially diluted cDNA samples, and the relative expression at each time point of infection was calculated using the formula $2^{-\Delta\Delta CT}$ ($2^{-(CT_{GAPDH} - CT_{C315R})}$).

Preparation of supernatant and cell lysis extracts for ELISA and Western blot detection of host proteins. Lung macrophage cultures from two donor outbred pigs (same cells used for CAGEseq) were prepared in 6-well plates. Approximately 1.5×10^6 cells were seeded per well with 3 mL medium (RPMI with penicillin/streptomycin and 10% pig serum) and incubated at 37°C and 5% CO₂ overnight. Cultures were washed once with culture medium to remove nonadherent cells and inoculated with an MOI of 5 of ASFV-Georgia 2007/1 (or left uninfected as the control) and centrifuged 1 h at 600 $\times g$, 26°C (adsorption period). Supernatants from cell cultures were collected immediately after adsorption for obtaining the 0-hpi time point and stored at -70°C degrees until analysis. Adherent cells were washed twice with cold Dulbecco's phosphate-buffered saline (DPBS; Sigma) and then lysed with 0.12 mL/well cold RIPA buffer (Thermo Scientific) supplemented with protease inhibitors (Halt protease inhibitor cocktail; Thermo Scientific). For the 5-h and 16-h time points, the inoculum was removed after adsorption, and cells were washed twice in culture medium and returned to the incubator with 3 mL fresh medium per well for the specified times of infection. Supernatants and lysis volumes were collected similarly to the control. Supernatants were analyzed for the presence of CCL2 (porcine CCL2/MCP-1 ELISA kit, ES2RB; Invitrogen), CXCL8 (Quantikine ELISA, porcine IL-8/CXCL8 immunoassay, P8000; R&D) and TNF- α (Quantikine ELISA, porcine TNF- α immunoassay, PTA00; R&D) as recommended by the manufacturers. A volume of 25 μ L each lysate was analyzed by Western blotting for expression of ISG15 (anti-ISG15 antibody ab233071; Abcam; used at 1:1,000 dilution), γ -tubulin (anti-gamma tubulin antibody ab11321; Abcam; used at 1:1,000 dilution), and viral ASFV protein P30 (in-house mouse monoclonal antibody used at 1:500 dilution). Secondary antibodies used were goat anti-rabbit IgG H&L (horseradish peroxidase [HRP]) (ab205718; Abcam) and goat anti-mouse immunoglobulins/HRP (P0447; Dako), both at 1:2,000 dilution. Western blot membranes were revealed using Pierce ECL western blotting substrate (32106; Thermo Scientific). Band densities were quantified using ImageJ (107; <https://imagej.nih.gov/ij/>).

ASFV promoter motif analysis. DESeq2 results were used to categorize ASFV genes into two simple subclasses—early (87 genes downregulated from early to late infection) and late (the 78 upregulated from early to late infection). These characterized gene pTSSs were then pooled with the nORF pTSSs, and sequences upstream and downstream of the pTSSs were extracted from the GRG genome in FASTA format using BEDtools. Sequences 35 bp upstream of and including the pTSSs were analyzed using MEME software (<http://meme-suite.org>) (108), searching for 5 motifs with a maximum width of 20 nt and 27 nt, respectively (other settings at default). The input for MEME motif searches included sequences upstream of 134 early pTSSs (87 genes and 47 nORFs) for early promoter searching, while 234 late pTSSs (78 genes and 156 nORFs) were used to search for late promoters. For analysis of conserved motifs upstream of the five clusters described in Fig. 6a and b, sequences were extracted in the same manner as described above but grouped according to their cluster. MEME motif searches were carried out for sequences in each cluster, searching for 3 motifs, 5 to 36 bp in length, with zero or one occurrence per sequence (“zoops” mode).

Identification of TSSs by rapid amplification of cDNA ends (5' RACE). For 5' RACE of GRG genes DP146L, pNG4, and CP204L, we designed the gene-specific primers (GSP) shown in Table S7c and used the kit “5' RACE system for rapid amplification of cDNA ends” (Invitrogen), according to the manufacturer's instructions. Briefly, 150 ng RNA from either 5-hpi or 16-hpi macrophages (one of the replicate RNA samples used for CAGE-seq) was used for cDNA synthesis with GSP1 primers, followed by degradation of the mRNA template with RNase mix and column purification of the cDNA. A homopolymeric tail was added to the cDNA 3' ends with terminal deoxynucleotidyl transferase, which allowed PCR amplification with an abridged anchor primer (AAP) from the 5' RACE kit and a nested GSP2 primer. A second PCR was performed over an aliquot of the previous, with 5' RACE abridged universal amplification primer (AUAP), and an additional nested primer, GSP3, except for pNG4 where GSP2 was reused due to the small predicted size of the amplicon. Platinum green hot start PCR master mix (Invitrogen) was used for PCR, and products were run in 2% agarose gel electrophoresis (see Table S7c for expected sizes). Efficient recovery of cDNA from the purification column requires a product of at least 200 bases, and therefore, due to the small predicted size of pNG4 transcripts, its GSP1 primer was extended at the 5' end with an irrelevant nonannealing sequence of an extra 50 nt in order to create a longer recoverable product.

CAGE-seq analysis for the *Sus scrofa* genome. Analyses of TSS-mapping, gene expression, and motif searching with CAGE-seq reads mapped to the *Sus scrofa* 11.1 genome were carried out by DNAFORM (Yokohama, Kanagawa, Japan). The 5' ends of CAGE-seq reads were utilized as input for the Reclu pipeline (109) with a cutoff of 0.1 RPM and irreproducible discovery rate of 0.1. A total of 37,159 CAGE-seq peaks could be identified, of which around half (16,720) match unique CAGE peaks previously identified by Robert et al. (64) (i.e., within 100 nt of any of them). 9,384 TSSs were mapped to the nearest *S. scrofa* 11.1 Ensembl (110) protein-coding (out of 21,288) were annotated *de novo* from the CAGE-defined TSSs (Table S4).

Protein-coding genes with annotated TSSs (9,384 out of 21,288) were then subjected to differential expression analysis. CAGE-seq reads were summed up over all TSSs assigned to a gene and compared between two time points using edgeR (111) at a maximum false-discovery rate of 0.05. The full list of host genes with annotated promoters together with their estimated expression levels is provided in Table S5. Gene set enrichment analysis was performed with the DAVID 6.8 bioinformatics resources (112), using the best BLASTP (110, 113, 114) human hits (from the UniProt [115] reference human proteome). The 9,331 genes with human homologs were used as a background, and functional annotations of the four major expression response groups (late/early up/downregulated genes) were clustered in DAVID 6.8 using medium classification stringency. MEME motif searches were conducted for promoters of four differentially regulated subsets of host genes, as defined in Fig. 11a. Promoter sequences were extended 1,000 bp upstream and 200 bp downstream of TSSs, searched with MEME (max. 10 motifs, max. 100 bp long, on a given strand only, zero or one site per sequence, $E < 0.01$), and then compared against known vertebrate DNA motifs with Tomtom (P value < 0.01).

Data availability. The raw sequencing data are available on the Sequence Read Archive (SRA) database under BioProject number [PRJNA739166](https://www.ncbi.nlm.nih.gov/bioproject/PRJNA739166). This also includes CAGE-seq data aligned to the ASFV-GRG (GenBank accession number [FR682468.1](https://www.ncbi.nlm.nih.gov/nuccore/FR682468.1) *Sus scrofa* ([GCF_000003025.6](https://www.ncbi.nlm.nih.gov/nuccore/GCF_000003025.6)) genomes (see Materials and Methods, above) in BAM format.

SUPPLEMENTAL MATERIALS

Supplemental material is available online only.

SUPPLEMENTAL FILE 1, XLSX file, 0.1 MB.

SUPPLEMENTAL FILE 2, XLSX file, 0.1 MB.

SUPPLEMENTAL FILE 3, XLSX file, 0.02 MB.

SUPPLEMENTAL FILE 4, XLSX file, 3 MB.

SUPPLEMENTAL FILE 5, XLSX file, 1.8 MB.

SUPPLEMENTAL FILE 6, XLSX file, 0.5 MB.

SUPPLEMENTAL FILE 7, XLSX file, 0.02 MB.

SUPPLEMENTAL FILE 8, PDF file, 0.9 MB.

ACKNOWLEDGMENTS

The research at University College London (UCL) was funded in part by the Wellcome Trust (WT 207446/Z/17/Z and WT 108877/B/15/Z). For the purpose of open access, we have applied a CC BY public copyright license to any author accepted manuscript version arising from this submission. Research at The Pirbright Institute is funded by the Biotechnology and Biological Sciences Research Council (BBS/E/I/0007030 and BBS/E/I/0007031).

We are grateful to all members of the RNAP lab and Tine Arnvig for critical reading of the manuscript.

REFERENCES

- Gogin A, Gerasimov V, Malogolovkin A, Kolbasov D. 2013. African swine fever in the North Caucasus region and the Russian Federation in years 2007–2012. *Virus Res* 173:198–203. <https://doi.org/10.1016/j.virusres.2012.12.007>.
- Zhou X, Li N, Luo Y, Liu Y, Miao F, Chen T, Zhang S, Cao P, Li X, Tian K, Qiu H-J, Hu R. 2018. Emergence of African swine fever in China, 2018. *Transbound Emerg Dis* 65:1482–1484. <https://doi.org/10.1111/tbed.12989>.
- Alonso C, Borca M, Dixon L, Revilla Y, Rodriguez F, Escribano JM, ICTV Report Consortium. 2018. ICTV virus taxonomy profile: Asfarviridae. *J Gen Virol* 99:613–614. <https://doi.org/10.1099/jgv.0.001049>.
- Koonin EV, Yutin N. 2010. Origin and evolution of eukaryotic large nucleocytoplasmic DNA viruses. *Intervirology* 53:284–292. <https://doi.org/10.1159/000312913>.
- Yutin N, Koonin EV. 2012. Hidden evolutionary complexity of nucleocytoplasmic large DNA viruses of eukaryotes. *Virology* 439:161. <https://doi.org/10.1016/j.virusres.2012.09.014>.
- Broyles SS. 2003. Vaccinia virus transcription. *J Gen Virol* 84:2293–2303. <https://doi.org/10.1099/vir.0.18942-0>.
- Alejo A, Matamoros T, Guerra M, Andrés G. 2018. A proteomic atlas of the African swine fever virus particle. *J Virol* 92:JV1.01293-18. <https://doi.org/10.1128/JVI.01293-18>.
- Salas ML, Kuznar J, Viñuela E. 1981. Polyadenylation, methylation, and capping of the RNA synthesized in vitro by African swine fever virus. *Virology* 113:484–491. [https://doi.org/10.1016/0042-6822\(81\)90176-8](https://doi.org/10.1016/0042-6822(81)90176-8).
- Rodriguez JM, Salas ML. 2013. African swine fever virus transcription. *Virus Res* 173:15–28. <https://doi.org/10.1016/j.virusres.2012.09.014>.
- Cackett G, Matelska D, Sýkora M, Portugal R, Malecki M, Bähler J, Dixon L, Werner F. 2020. The African swine fever virus transcriptome. *J Virol* 94 <https://doi.org/10.1128/JVI.00119-20>.
- Iyer LM, Balaji S, Koonin EV, Aravind L. 2006. Evolutionary genomics of nucleocytoplasmic large DNA viruses. *Virus Res* 117:156–184. <https://doi.org/10.1016/j.virusres.2006.01.009>.
- Yutin N, Wolf YI, Raoult D, Koonin EV. 2009. Eukaryotic large nucleocytoplasmic DNA viruses: clusters of orthologous genes and reconstruction of viral genome evolution. *Virology* 392:223. <https://doi.org/10.1016/j.virusres.2009.06.023>.
- Grimm C, Bartuli J, Boettcher B, Szalay AA, Fischer U. 2021. Structural basis of the complete poxvirus transcription initiation process. *Nat Struct Mol Biol* 28:779–788. <https://doi.org/10.1038/s41594-021-00655-w>.
- Rodriguez JM, Moreno LT, Alejo A, Lacasta A, Rodriguez F, Salas ML. 2015. Genome sequence of African swine fever virus BA71, the virulent parental strain of the nonpathogenic and tissue-culture adapted BA71V. *PLoS One* 10:e0142889. <https://doi.org/10.1371/journal.pone.0142889>.
- Yañez RJ, Rodríguez JM, Nogal ML, Yuste L, Enríquez C, Rodríguez JF, Viñuela E. 1995. Analysis of the complete nucleotide sequence of African swine fever virus. *Virology* 208:249–278. <https://doi.org/10.1006/viro.1995.1149>.
- Dixon LK, Chapman DAG, Netherton CL, Upton C. 2013. African swine fever virus replication and genomics. *Virus Res* 173:3–14. <https://doi.org/10.1016/j.virusres.2012.10.020>.
- Chapman DAG, Tcherepanov V, Upton C, Dixon LK. 2008. Comparison of the genome sequences of non-pathogenic and pathogenic African swine fever virus isolates. *J Gen Virol* 89:397–408. <https://doi.org/10.1099/vir.0.83343-0>.
- Forth J, Forth L, King J, Groza O, Hübner A, Olesen A, Höper D, Dixon L, Netherton C, Rasmussen T, Blome S, Pohlmann A, Beer M. 2019. A deep-sequencing workflow for the fast and efficient generation of high-quality African swine fever virus whole-genome sequences. *Viruses* 11:846. <https://doi.org/10.3390/v11090846>.
- Chapman DAG, Darby AC, da Silva M, Upton C, Radford AD, Dixon LK. 2011. Genomic analysis of highly virulent Georgia 2007/1 isolate of African swine fever virus. *Emerg Infect Dis* 17:599–605. <https://doi.org/10.3201/eid1704.101283>.
- Farlow J, Donduashvili M, Kokhreizidze M, Kotorashvili A, Vepkhvadze NG, Kotaria N, Gulbani A. 2018. Intra-epidemic genome variation in highly pathogenic African swine fever virus (ASFV) from the country of Georgia. *Virology* 515:190. <https://doi.org/10.1186/s12985-018-1099-z>.
- Mazur-Panasiuk N, Woźniakowski G, Niemczuk K. 2019. The first complete genomic sequences of African swine fever virus isolated in Poland. *Sci Rep* 9:3–5. <https://doi.org/10.1038/s41598-018-36823-0>.
- Granberg F, Torresi C, Oggiano A, Malmberg M, Iscaro C, De Mia GM, Belák S. 2016. Complete genome sequence of an African swine fever virus isolate from Sardinia, Italy. *Genome Announc* 4:1220–1236. <https://doi.org/10.1128/genomeA.01220-16>.
- Wang Z, Jia L, Li J, Liu H, Liu D. 2019. Pan-genomic analysis of African swine fever virus. *Virology* 524:662–665. <https://doi.org/10.1007/s12250-019-00173-6>.
- Rowlands RJ, Michaud V, Heath L, Hutchings G, Oura C, Vosloo W, Dwarka R, Onashvili T, Albina E, Dixon LK. 2008. African swine fever virus isolate, Georgia, 2007. *Emerg Infect Dis* 14:1870–1874. <https://doi.org/10.3201/eid1412.080591>.
- Zhao D, Liu R, Zhang X, Li F, Wang J, Zhang J, Liu X, Wang L, Zhang J, Wu X, Guan Y, Chen W, Wang X, He X, Bu Z. 2019. Replication and virulence in pigs of the first African swine fever virus isolated in China. *Emerg Microbes Infect* 8:438–447. <https://doi.org/10.1080/22221751.2019.1590128>.
- Zani L, Forth JH, Forth L, Nurmoja I, Leidenberger S, Henke J, Carlson J, Breidenstein C, Viltrop A, Höper D, Sauter-Louis C, Beer M, Blome S. 2018. Deletion at the 5'-end of Estonian ASFV strains associated with an attenuated phenotype. *Sci Rep* 8:1–11. <https://doi.org/10.1038/s41598-018-24740-1>.
- Gallardo C, Nurmoja I, Soler A, Delicado V, Simón A, Martín E, Pérez C, Nieto R, Arias M. 2018. Evolution in Europe of African swine fever genotype II viruses from highly to moderately virulent. *Vet Microbiol* 219:70–79. <https://doi.org/10.1016/j.vetmic.2018.04.001>.
- Pershin A, Shevchenko I, Igolkin A, Zhukov I, Mazloum A, Aronova E, Vlasova N, Shevtsov A. 2019. A long-term study of the biological properties of ASF virus isolates originating from various regions of the Russian Federation in 2013–2018. *Vet Sci* 6:99. <https://doi.org/10.3390/vetsci6040099>.
- Sun E, Zhang Z, Wang Z, He X, Zhang X, Wang L, Wang W, Huang L, Xi F, Huangfu H, Tsegay G, Huo H, Sun J, Tian Z, Xia W, Yu X, Li F, Liu R, Guan Y, Zhao D, Bu Z. 2021. Emergence and prevalence of naturally occurring lower virulent African swine fever viruses in domestic pigs in China in 2020. *Sci China Life Sci* 64:752–765. <https://doi.org/10.1007/s11427-021-1904-4>.
- Imbery J, Upton C. 2017. Organization of the multigene families of African swine fever virus. *Finefocus* 3:155–170. <https://doi.org/10.33043/FF.3.2.155-170>.
- Netherton CL, Connell S, Benfield CTO, Dixon LK. 2019. The genetics of life and death: virus-host interactions underpinning resistance to African swine fever, a viral hemorrhagic disease. *Front Genet* 10:402. <https://doi.org/10.3389/fgene.2019.00402>.
- Reis AL, Abrams CC, Goatley LC, Netherton C, Chapman DG, Sanchez-Cordon P, Dixon LK. 2016. Deletion of African swine fever virus interferon inhibitors from the genome of a virulent isolate reduces virulence in

- domestic pigs and induces a protective response. *Vaccine* 34:4698–4705. <https://doi.org/10.1016/j.vaccine.2016.08.011>.
33. O'Donnell V, Risatti GR, Holinka LG, Krug PW, Carlson J, Velazquez-Salinas L, Azzinaro PA, Gladue DP, Borca MV. 2017. Simultaneous deletion of the 9GL and UK genes from the African swine fever virus Georgia 2007 isolate offers increased safety and protection against homologous challenge. *J Virol* 91:e01760–16. <https://doi.org/10.1128/JVI.01760-16>.
 34. Li D, Zhang J, Yang W, Li P, Ru Y, Kang W, Li L, Ran Y, Zheng H. 2021. African swine fever virus protein MGF-505-7R promotes virulence and pathogenesis by inhibiting JAK1- and JAK2-mediated signaling. *J Biol Chem* 297:101190. <https://doi.org/10.1016/j.jbc.2021.101190>.
 35. Li D, Liu Y, Qi X, Wen Y, Li P, Ma Z, Liu Y, Zheng H, Liu Z. 2021. African swine fever virus MGF-110-9L-deficient mutant has attenuated virulence in pigs. *Virol Sin* 36:187–195. <https://doi.org/10.1007/s12250-021-00350-6>.
 36. Keßler C, Forth JH, Keil GM, Mettenleiter TC, Blome S, Karger A. 2018. The intracellular proteome of African swine fever virus. *Sci Rep* 8:14714. <https://doi.org/10.1038/s41598-018-32985-z>.
 37. Randall RE, Goodbourn S. 2008. Interferons and viruses: an interplay between induction, signalling, antiviral responses and virus countermeasures. *J Gen Virol* 89:1–47. <https://doi.org/10.1099/vir.0.83391-0>.
 38. Afonso CL, Piccone ME, Zaffuto KM, Neilan J, Kutish GF, Lu Z, Balinsky CA, Gibb TR, Bean TJ, Zsak L, Rock DL. 2004. African swine fever virus multi-gene family 360 and 530 genes affect host interferon response. *J Virol* 78:1858–1864. <https://doi.org/10.1128/JVI.78.4.1858-1864.2004>.
 39. Neilan JG, Zsak L, Lu Z, Kutish GF, Afonso CL, Rock DL. 2002. Novel swine virulence determinant in the left variable region of the African swine fever virus genome. *J Virol* 76:3095–3104. <https://doi.org/10.1128/jvi.76.7.3095-3104.2002>.
 40. Golding JP, Goatley L, Goodbourn S, Dixon LK, Taylor G, Netherton CL. 2016. Sensitivity of African swine fever virus to type I interferon is linked to genes within multigene families 360 and 505. *Virology* 493:154–161. <https://doi.org/10.1016/j.virol.2016.03.019>.
 41. Mosser DM, Edwards JP. 2008. Exploring the full spectrum of macrophage activation. *Nat Rev Immunol* 8:958–969. <https://doi.org/10.1038/nri2448>.
 42. Zhu JJ, Ramanathan P, Bishop EA, O'Donnell V, Gladue DP, Borca MV. 2019. Mechanisms of African swine fever virus pathogenesis and immune evasion inferred from gene expression changes in infected swine macrophages. *PLoS One* 14:e0223955. <https://doi.org/10.1371/journal.pone.0223955>.
 43. Jaing C, Rowland RRR, Allen JE, Certoma A, Thissen JB, Bingham J, Rowe B, White JR, Wynne JW, Johnson D, Gaudreault NN, Williams DT. 2017. Gene expression analysis of whole blood RNA from pigs infected with low and high pathogenic African swine fever viruses. *Sci Rep* 7:10115. <https://doi.org/10.1038/s41598-017-10186-4>.
 44. Fan W, Cao Y, Jiao P, Yu P, Zhang H, Chen T, Zhou X, Qi Y, Sun L, Liu D, Zhu H, Liu W, Hu R, Li J. 2021. Synergistic effect of the responses of different tissues against African swine fever virus. *Transbound Emerg Dis* <https://doi.org/10.1111/tbed.14283>.
 45. Ju X, Li F, Li J, Wu C, Xiang G, Zhao X, Nan Y, Zhao D, Ding Q. 2021. Genome-wide transcriptomic analysis of highly virulent African swine fever virus infection reveals complex and unique virus host interaction. *Vet Microbiol* 261:109211. <https://doi.org/10.1016/j.vetmic.2021.109211>.
 46. Yang B, Shen C, Zhang D, Zhang T, Shi X, Yang J, Hao Y, Zhao D, Cui H, Yuan X, Chen X, Zhang K, Zheng H, Liu X. 2021. Mechanism of interaction between virus and host is inferred from the changes of gene expression in macrophages infected with African swine fever virus CN/GS/2018 strain. *Virol J* 18:186. <https://doi.org/10.1186/s12985-021-01654-5>.
 47. Love MI, Huber W, Anders S. 2014. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol* 15:550. <https://doi.org/10.1186/s13059-014-0550-8>.
 48. Hammond JM, Kerr SM, Smith GL, Dixon LK. 1992. An African swine fever virus gene with homology to DNA ligases. *Nucleic Acids Res* 20:2667–2671. <https://doi.org/10.1093/nar/20.11.2667>.
 49. Reis AL, Goatley LC, Jabbar T, Sanchez-Cordon PJ, Netherton CL, Chapman DAG, Dixon LK. 2017. Deletion of the African swine fever virus gene DP148R does not reduce virus replication in culture but reduces virus virulence in pigs and induces high levels of protection against challenge. *J Virol* 91:e01428-17. <https://doi.org/10.1128/JVI.01428-17>.
 50. Yang Z, Martens CA, Bruno DP, Porcella SF, Moss B. 2012. Pervasive initiation and 3'-end formation of poxvirus postreplicative RNAs. *J Biol Chem* 287:31050–31060. <https://doi.org/10.1074/jbc.M112.390054>.
 51. Frouco G, Freitas FB, Coelho J, Leitão A, Martins C, Ferreira F. 2017. DNA-binding properties of African swine fever virus pA104R, a histone-like protein involved in viral replication and transcription. *J Virol* 91:e02498-16. <https://doi.org/10.1128/JVI.02498-16>.
 52. Conesa A, Madrigal P, Tarazona S, Gomez-Cabrero D, Cervera A, McPherson A, Szczesniak MW, Gaffney DJ, Elo LL, Zhang X, Mortazavi S. 2016. A survey of best practices for RNA-seq data analysis. *Genome Biol* 17:13. <https://doi.org/10.1186/s13059-016-0881-8>.
 53. García-Escudero R, Viñuela E. 2000. Structure of African swine fever virus late promoters: requirement of a TATA sequence at the initiation region. *J Virol* 74:8176–8182. <https://doi.org/10.1128/JVI.74.17.8176-8182.2000>.
 54. Bailey TL, Boden M, Buske FA, Frith M, Grant CE, Clementi L, Ren J, Li WW, Noble WS. 2009. MEME SUITE: tools for motif discovery and searching. *Nucleic Acids Res* 37:W202–W208. <https://doi.org/10.1093/nar/gkp335>.
 55. Gupta S, Stamatoyannopoulos JA, Bailey TL, Noble W. 2007. Quantifying similarity between motifs. *Genome Biol* 8:R24. <https://doi.org/10.1186/gb-2007-8-2-r24>.
 56. Rodríguez JM, Salas ML, Viñuela E. 1996. Intermediate class of mRNAs in African swine fever virus. *J Virol* 70:8584–8589. <https://doi.org/10.1128/JVI.70.12.8584-8589.1996>.
 57. Kim DE, Chivian D, Baker D. 2004. Protein structure prediction and analysis using the Robetta server. *Nucleic Acids Res* 32:W526–W531. <https://doi.org/10.1093/nar/gkh468>.
 58. Vydellingum S, Baylis SA, Bristow C, Smith GL, Dixon LK. 1993. Duplicated genes within the variable right end of the genome of a pathogenic isolate of African swine fever virus. *J Gen Virol* 74:2125–2130. <https://doi.org/10.1099/0022-1317-74-10-2125>.
 59. Zhang J, Zhang Y, Chen T, Yang JJ, Yue H, Wang L, Zhou X, Qi Y, Han X, Ke J, Wang S, Yang J, Miao F, Zhang S, Zhang F, Wang Y, Li M, Hu R. 2021. Deletion of the L7L-L11L genes attenuates ASFV and induces protection against homologous challenge. *Viruses* 13:255. <https://doi.org/10.3390/v13020255>.
 60. Krug PW, Holinka LG, O'Donnell V, Reese B, Sanford B, Fernandez-Sainz I, Gladue DP, Arzt J, Rodriguez L, Risatti GR, Borca MV. 2015. The progressive adaptation of a Georgian isolate of African swine fever virus to Vero cells leads to a gradual attenuation of virulence in swine corresponding to major modifications of the viral genome. *J Virol* 89:2324–2332. <https://doi.org/10.1128/JVI.03250-14>.
 61. Kelley LA, Mezulis S, Yates CM, Wass MN, Sternberg MJE. 2015. The Phyre2 web portal for protein modeling, prediction and analysis. *Nat Protoc* 10:845–858. <https://doi.org/10.1038/nprot.2015.053>.
 62. Liu H, Li L, Voss C, Wang F, Liu J, Li SSC. 2015. A comprehensive immunoreceptor phosphotyrosine-based signaling network revealed by reciprocal protein-peptide array screening. *Mol Cell Proteomics* 14:1846–1858. <https://doi.org/10.1074/mcp.M115.047951>.
 63. Gabler F, Nam SZ, Till S, Mirdita M, Steinegger M, Söding J, Lupas AN, Alva V. 2020. Protein sequence analysis using the MPI Bioinformatics toolkit. *Curr Protoc Bioinforma* 72:e108. <https://doi.org/10.1002/cpbi.108>.
 64. Robert C, Kapetanovic R, Beraldi D, Watson M, Archibald AL, Hume DA. 2015. Identification and annotation of conserved promoters and macrophage-expressed genes in the pig genome. *BMC Genomics* 16 <https://doi.org/10.1186/s12864-015-2111-2>.
 65. Benjamini Y, Hochberg Y. 1995. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J R Stat Soc Ser B* 57:289–300.
 66. Ganchi PA, Sun SC, Greene WC, Ballard DW. 1993. A novel NF-kappa B complex containing p65 homodimers: implications for transcriptional control at the level of subunit dimerization. *Mol Cell Biol* 13:7826–7835.
 67. Dixon LK, Abrams CC, Bowick G, Goatley LC, Kay-Jackson PC, Chapman D, Liverani E, Nix R, Silk R, Zhang F. 2004. African swine fever virus proteins involved in evading host defence systems. *Vet Immunol Immunopathol* 100:117–134. <https://doi.org/10.1016/j.vetimm.2004.04.002>.
 68. Powell PP, Dixon LK, Parkhouse RM. 1996. An IkappaB homolog encoded by African swine fever virus provides a novel mechanism for downregulation of proinflammatory cytokine responses in host macrophages. *J Virol* 70:8527–8533. <https://doi.org/10.1128/JVI.70.12.8527-8533.1996>.
 69. Granja AG, Sánchez EG, Sabina P, Fresno M, Revilla Y. 2009. African swine fever virus blocks the host cell antiviral inflammatory response through a direct inhibition of PKC-theta-mediated p300 transactivation. *J Virol* 83:969–980. <https://doi.org/10.1128/JVI.01663-08>.
 70. Nogal ML, González de Buitrago G, Rodríguez C, Cubelos B, Carrascosa AL, Salas ML, Revilla Y. 2001. African swine fever virus IAP homologue inhibits caspase activation and promotes cell survival in mammalian cells. *J Virol* 75:2535–2543. <https://doi.org/10.1128/JVI.75.6.2535-2543.2001>.
 71. Takeya T, Hanafusa H. 1982. DNA sequence of the viral and cellular src gene of chickens. II. Comparison of the src genes of two strains of avian

- sarcoma virus and the cellular homolog. *J Virol* 44:12–18. <https://doi.org/10.1128/JVI.44.1.12-18.1982>.
72. Kaneko T, Stogios PJ, Ruan X, Voss C, Evdokimova E, Skarina T, Chung A, Liu X, Li L, Savchenko A, Ensminger AW, Li SS-C. 2018. Identification and characterization of a large family of superbinding bacterial SH2 domains. *Nat Commun* 9:4549. <https://doi.org/10.1038/s41467-018-06943-2>.
 73. Liu Y, Li Y, Xie Z, Ao Q, Di D, Yu W, Lv L, Zhong Q, Song Y, Liao X, Song Q, Wang H, Chen H. 2021. Development and in vivo evaluation of MGF100-1R deletion mutant in an African swine fever virus Chinese strain. *Vet Microbiol* 261:109208. <https://doi.org/10.1016/j.vetmic.2021.109208>.
 74. Vuono E, Ramirez-Medina E, Pruitt S, Rai A, Silva E, Espinoza N, Zhu J, Velazquez-Salinas L, Gladue DP, Borca MV. 2020. Evaluation in swine of a recombinant Georgia 2010 African swine fever virus lacking the i81 gene. *Viruses* 13:39. <https://doi.org/10.3390/v13010039>.
 75. Camacho A, Viñuela E. 1991. Protein p22 of African swine fever virus: an early structural protein that is incorporated into the membrane of infected cells. *Virology* 181:251–257. [https://doi.org/10.1016/0042-6822\(91\)90490-3](https://doi.org/10.1016/0042-6822(91)90490-3).
 76. Netherton C, Rouiller I, Wileman T. 2004. The subcellular distribution of multigene family 110 proteins of African swine fever virus is determined by differences in C-terminal KDEL endoplasmic reticulum retention motifs. *J Virol* 78:3710–3721. <https://doi.org/10.1128/jvi.78.7.3710-3721.2004>.
 77. Ramirez-Medina E, Vuono E, Pruitt S, Rai A, Silva E, Espinoza N, Zhu J, Velazquez-Salinas L, Borca MV, Gladue DP. 2021. Development and in vivo evaluation of a MGF110-1L deletion mutant in African swine fever strain Georgia. *Viruses* 13:286. <https://doi.org/10.3390/v13020286>.
 78. Cackett G, Sýkora M, Werner F. 2020. Transcriptome view of a killer: African swine fever virus. *Biochem Soc Trans* 48:1569–1581. <https://doi.org/10.1042/BST20191108>.
 79. Quintas A, Pérez-Núñez D, Sánchez EG, Nogal ML, Hentze MW, Castelló A, Revilla Y. 2017. Characterization of the African swine fever virus decapping enzyme during infection. *J Virol* 91:e00990-17. <https://doi.org/10.1128/JVI.00990-17>.
 80. Kunsch C, Ruben SM, Rosen CA. 1992. Selection of optimal kappa B/Rel DNA-binding motifs: interaction of both subunits of NF-kappa B with DNA is required for transcriptional activation. *Mol Cell Biol* 12:4412–4421. <https://doi.org/10.1128/mcb.12.10.4412-4421.1992>.
 81. Gómez del Moral M, Ortuño E, Fernández-Zapatero P, Alonso F, Alonso C, Ezquerro A, Domínguez J. 1999. African swine fever virus infection induces tumor necrosis factor alpha production: implications in pathogenesis. *J Virol* 73:2173–2180. <https://doi.org/10.1128/JVI.73.3.2173-2180.1999>.
 82. Roy S, Schmeier S, Arner E, Alam T, Parihar SP, Ozturk M, Tamgue O, Kawaji H, de Hoon MJL, Itoh M, Lassmann T, Carninci P, Hayashizaki Y, Forrest ARR, Bajic VB, Guler R, Brombacher F, Suzuki H, Fantom Consortium. 2015. Redefining the transcriptional regulatory dynamics of classically and alternatively activated macrophages by deepCAGE transcriptomics. *Nucleic Acids Res* 43:6969–6982. <https://doi.org/10.1093/nar/gkv646>.
 83. Castelló A, Quintas A, Sánchez EG, Sabina P, Nogal M, Carrasco L, Revilla Y. 2009. Regulation of host translational machinery by African swine fever virus. *PLoS Pathog* 5:e1000562. <https://doi.org/10.1371/journal.ppat.1000562>.
 84. Sánchez EG, Quintas A, Nogal M, Castelló A, Revilla Y. 2013. African swine fever virus controls the host transcription and cellular machinery of protein synthesis. *Virus Res* 173:58–75. <https://doi.org/10.1016/j.virusres.2012.10.025>.
 85. Barrado-Gil L, Del Puerto A, Muñoz-Moreno R, Galindo I, Cuesta-Geijo MA, Urquiza J, Nistal-Villán E, de Motes CM, Alonso C. 2020. African swine fever virus ubiquitin-conjugating enzyme interacts with host translation machinery to regulate the host protein synthesis. *Front Microbiol* 11:622907. <https://doi.org/10.3389/fmicb.2020.622907>.
 86. Gschwandtner M, Derler R, Midwood KS. 2019. More than just attractive: how CCL2 influences myeloid cell behavior beyond chemotaxis. *Front Immunol* 10:2759. <https://doi.org/10.3389/fimmu.2019.02759>.
 87. Xia C, Braunstein Z, Toomey AC, Zhong J, Rao X. 2018. S100 proteins as an important regulator of macrophage inflammation. *8:1908*. *Front Immunol* <https://doi.org/10.3389/fimmu.2017.01908>.
 88. Yang Z, Bruno DP, Martens CA, Porcella SF, Moss B. 2010. Simultaneous high-resolution analysis of vaccinia virus and host cell transcriptomes by deep RNA sequencing. *Proc Natl Acad Sci U S A* 107:11513–11518. <https://doi.org/10.1073/pnas.1006594107>.
 89. Olasz F, Tombácz D, Torma G, Csabai Z, Moldován N, Dörmő Á, Prászkál I, Mészáros I, Magyar T, Tamás V, Zádori Z, Boldogkői Z. 2020. Short and long-read sequencing survey of the dynamic transcriptomes of African swine fever virus and the host cells. *Front Genet* 11:758. <https://doi.org/10.3389/fgene.2020.00758>.
 90. Torma G, Tombácz D, Csabai Z, Moldován N, Mészáros I, Zádori Z, Boldogkői Z. 2021. Combined short and long-read sequencing reveals a complex transcriptomic architecture of African swine fever virus. *Viruses* 13:579. <https://doi.org/10.3390/v13040579>.
 91. Yang Z, Bruno DP, Martens CA, Porcella SF, Moss B. 2011. Genome-wide analysis of the 5' and 3' ends of vaccinia virus early mRNAs delineates regulatory sequences of annotated and anomalous transcripts. *J Virol* 85:5897–5909. <https://doi.org/10.1128/JVI.00428-11>.
 92. Gershon PD, Moss B. 1990. Early transcription factor subunits are encoded by vaccinia virus late genes. *Proc Natl Acad Sci U S A* 87:4401–4405. <https://doi.org/10.1073/pnas.87.11.4401>.
 93. Li J, Broyles SS. 1993. Recruitment of vaccinia virus RNA polymerase to an early gene promoter by the viral early transcription factor. *J Biol Chem* 268:2773–2780. [https://doi.org/10.1016/S0021-9258\(18\)53841-2](https://doi.org/10.1016/S0021-9258(18)53841-2).
 94. Patikoglou GA, Kim JL, Sun L, Yang SH, Kodadek T, Burley SK. 1999. TATA element recognition by the TATA box-binding protein has been conserved throughout evolution. *Genes Dev* 13:3217–3230. <https://doi.org/10.1101/gad.13.24.3217>.
 95. Andrews S. 2010. FastQC: a quality control tool for high throughput sequence data. <https://www.bioinformatics.babraham.ac.uk/projects/fastqc/>.
 96. Langmead B, Salzberg SL. 2012. Fast gapped-read alignment with Bowtie 2. *Nat Methods* 9:357–359. <https://doi.org/10.1038/nmeth.1923>.
 97. Kim D, Langmead B, Salzberg SL. 2015. HISAT: a fast spliced aligner with low memory requirements. *Nat Methods* 12:357–360. <https://doi.org/10.1038/nmeth.3317>.
 98. Quinlan AR, Hall IM. 2010. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* 26:841–842. <https://doi.org/10.1093/bioinformatics/btq033>.
 99. RStudio Team. 2016. RStudio: integrated development for R. RStudio, Inc., Boston, MA.
 100. Huber W, Carey VJ, Gentleman R, Anders S, Carlson M, Carvalho BS, Bravo HC, Davis S, Gatto L, Girke T, Gottardo R, Hahne F, Hansen KD, Irizarry RA, Lawrence M, Love MI, MacDonald J, Obenchain V, Oleś AK, Pagès H, Reyes A, Shannon P, Smyth GK, Tenenbaum D, Waldron L, Morgan M. 2015. Orchestrating high-throughput genomic analysis with Bioconductor. *Nat Methods* 12:115–121. <https://doi.org/10.1038/nmeth.3252>.
 101. Thodberg M, Thieffry A, Vitting-Seerup K, Andersson R, Sandelin A. 2019. CAGEfightR: analysis of 5'-end data using R/Bioconductor. *BMC Bioinformatics* 20:487. <https://doi.org/10.1186/s12859-019-3029-5>.
 102. Anders S, Pyl PT, Huber W. 2015. HTSeq: a Python framework to work with high-throughput sequencing data. *Bioinformatics* 31:166–169. <https://doi.org/10.1093/bioinformatics/btu638>.
 103. Upton C, Slack S, Hunter AL, Ehlers A, Roper RL, Rock DL. 2003. Poxvirus orthologous clusters: toward defining the minimum essential poxvirus genome. *J Virol* 77:7590–7600. <https://doi.org/10.1128/jvi.77.13.7590-7600.2003>.
 104. Tu SL, Upton C. 2019. Bioinformatics for analysis of poxvirus genomes. *Methods Mol Biol* 2023:29–62. https://doi.org/10.1007/978-1-4939-9593-6_2.
 105. King DP, Reid SM, Hutchings GH, Grierson SS, Wilkinson PJ, Dixon LK, Bastos ADS, Drew TW. 2003. Development of a TaqMan PCR assay with internal amplification control for the detection of African swine fever virus. *J Virol Methods* 107:53–61. [https://doi.org/10.1016/s0166-0934\(02\)00189-1](https://doi.org/10.1016/s0166-0934(02)00189-1).
 106. Melchjorsen J, Kristiansen H, Christiansen R, Rintahaka J, Matikainen S, Paludan SR, Hartmann R. 2009. Differential regulation of the OASL and OAS1 genes in response to viral infections. *J Interf Cytokine Res* 29:199–207. <https://doi.org/10.1089/jir.2008.0050>.
 107. Schneider CA, Rasband WS, & Eliceiri KW. 2012. NIH Image to ImageJ: 25 years of image analysis. *Nature Methods*, 9(7), 671–675. <https://doi.org/10.1038/nmeth.2089>.
 108. Bailey TL, Elkan C. 1994. Fitting a mixture model by expectation maximization to discover motifs in biopolymers. *Proceedings Int Conf Intell Syst Mol Biol* 2:28–36.
 109. Ohmiya H, Vitezic M, Frith MC, Itoh M, Carninci P, Forrest ARR, Hayashizaki Y, Lassmann T, FANTOM Consortium. 2014. RECLU: a pipeline to discover reproducible transcriptional start sites and their alternative regulation using capped analysis of gene expression (CAGE). *BMC Genomics* 15:269. <https://doi.org/10.1186/1471-2164-15-269>.
 110. Yates AD, Achuthan P, Akanni W, Allen J, Allen J, Alvarez-Jarreta J, Amode MR, Armean IM, Azov AG, Bennett R, Bhai J, Billis K, Boddu S, Marugán JC, Cummins C, Davidson C, Dodiya K, Fatima R, Gall A, Giron CG, Gil L, Grego T, Haggerty L, Haskell E, Hourlier T, Izuogu OG, Janacek

- SH, Juettemann T, Kay M, Lavidas I, Le T, Lemos D, Martinez JG, Maurel T, McDowall M, McMahon A, Mohanan S, Moore B, Nuhn M, Oheh DN, Parker A, Parton A, Patricio M, Sakthivel MP, Abdul Salam AI, Schmitt BM, Schuilenburg H, Sheppard D, Sycheva M, Szuba M, et al. 2020. *Nucleic Acids Res* 48:D682–8. <https://doi.org/10.1093/nar/gkz966>.
111. Robinson MD, McCarthy DJ, Smyth GK. 2010. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* 26:139–140. <https://doi.org/10.1093/bioinformatics/btp616>.
112. Huang DW, Sherman BT, Lempicki RA. 2009. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat Protoc* 4:44–57. <https://doi.org/10.1038/nprot.2008.211>.
113. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. 1990. Basic local alignment search tool. *J Mol Biol* 215:403–410. [https://doi.org/10.1016/S0022-2836\(05\)80360-2](https://doi.org/10.1016/S0022-2836(05)80360-2).
114. Wheeler DL, Church DM, Federhen S, Lash AE, Madden TL, Pontius JU, Schuler GD, Schriml LM, Sequeira E, Tatusova TA, Wagner L. 2003. Database resources of the National Center for Biotechnology. *Nucleic Acids Res* 31:28–33. <https://doi.org/10.1093/nar/gkg033>.
115. UniProt Consortium. 2019. UniProt: a worldwide hub of protein knowledge. *Nucleic Acids Res* 47:D506–D55. <https://doi.org/10.1093/nar/gky1049>.
116. Ramírez F, Dündar F, Diehl S, Grüning BA, Manke T. 2014. deepTools: a flexible platform for exploring deep-sequencing data. *Nucleic Acids Res* 42:W187–W11. <https://doi.org/10.1093/nar/gku365>.
117. Crooks GE, Hon G, Chandonia JM, Brenner SE. 2004. WebLogo: a sequence logo generator. *Genome Res* 14:1188–1190. <https://doi.org/10.1101/gr.849004>.
118. Grant CE, Bailey TL, Noble WS. 2011. FIMO: scanning for occurrences of a given motif. *Bioinformatics* 27:1017–1018. <https://doi.org/10.1093/bioinformatics/btr064>.
119. McCarthy DJ, Chen Y, Smyth GK. 2012. Differential expression analysis of multifactor RNA-Seq experiments with respect to biological variation. *Nucleic Acids Res* 40:4288–4297. <https://doi.org/10.1093/nar/gks042>.