Title: Ribosome profiling of porcine reproductive and respiratory syndrome virus reveals novel features of viral gene expression

Author names and affiliations: Georgia M. Cook[1], Katherine Brown[1], Pengcheng Shang[2,3], Yanhua Li[2,4], Lior Soday[1,5], Adam M. Dinan[1,6], Charlotte Tumescheit[1], A. P. Adrian Mockett[7], Ying Fang*[2,8], Andrew E. Firth*[1], Ian Brierley*[1]

[1]Department of Pathology, University of Cambridge, UK; [2]Department of Diagnostic Medicine and Pathobiology, Kansas State University, US; [3]Current address: Department of Pediatrics, School of Medicine, University of Pittsburgh, US; [4]Current address: College of Veterinary Medicine, Yangzhou Univeristy, Yangzhou, China; [5]Current address: Department of Chemistry, Molecular Sciences Research Hub, Imperial College London, London, UK; [6]Current address: Department of Medicine, University of Cambridge, UK; [7]Cambivac Ltd, Babraham Research Campus, Cambridge, UK; [8]Current address: Department of Pathobiology, University of Illinois at Urbana-Champaign, US

Keywords: arterivirus, porcine reproductive and respiratory syndrome virus, PRRSV, ribosome profiling, programmed ribosomal frameshifting, subgenomic mRNAs, open reading frames, nidovirus, RNA sequencing, transcriptome, translatome, gene expression regulation

Abstract

Porcine reproductive and respiratory syndrome virus (PRRSV) is an arterivirus which causes significant economic losses to the swine industry worldwide. Here, we use ribosome profiling (RiboSeq) and parallel RNA sequencing (RNASeq) to characterise the transcriptome and translatome of both species of PRRSV and analyse the host response to infection. We quantified viral gene expression over a timecourse of infection, and calculated the efficiency of programmed ribosomal frameshifting (PRF) at both sites on the viral genome. At the nsp2 frameshift site (a rare example of protein-stimulated frameshifting), –2 PRF efficiency increases over time, likely facilitated by accumulation of the PRF-stimulatory viral protein (nsp1β) during infection. This marks arteriviruses as the second example of temporally regulated PRF. Surprisingly, we also found PRF efficiency at the canonical ORF1ab frameshift site increases over time, in apparent contradiction of the common assumption that RNA structure-directed frameshift sites operate at a fixed efficiency. This has potential implications for the numerous other viruses with canonical PRF sites. Furthermore, we discovered several highly translated additional viral ORFs, the translation of which may be facilitated by multiple novel viral transcripts. For example, we found a 125-codon ORF overlapping nsp12, which is expressed as highly as nsp12 itself at late stages of replication, and is likely translated from novel subgenomic (sg) RNA transcripts that overlap the 3′ end of ORF1b. Similar transcripts were discovered for both PRRSV-1 and PRRSV-2, suggesting a potential conserved mechanism for temporal regulation of expression of the 3′-proximal region of ORF1b. In addition, we identified a highly translated, short upstream ORF (uORF) in the 5′ UTR, the presence of which is highly conserved amongst PRRSV-2 isolates. This is the first application of RiboSeq to arterivirus-infected cells, and reveals new features which add to the complexity of gene expression programmes in this important family of nidoviruses.

Introduction

Porcine reproductive and respiratory syndrome virus (PRRSV) is an enveloped, positive-sense, single-stranded RNA virus in the family *Arteriviridae* (order: *Nidovirales*)[1,2], and the aetiological agent of the disease from which it takes its name: porcine reproductive and respiratory syndrome (PRRS). Attempts to control PRRS by vaccination have had limited success[3] and it remains one of the most economically devastating diseases of swine, causing reproductive failure in adult sows and respiratory failure in young pigs, at an estimated cost of $664 million a year in the US alone[4,5]. The two lineages of PRRSV, formerly known as "European" (Type 1) and "North American" (Type 2) PRRSV, share just ~60% pairwise nucleotide similarity and were recently re-classified as two separate species, *Betaarterivirus suid 1* and *2* (viruses named PRRSV-1 and PRRSV-2)[5–7]. For ease of reference, PRRSV-1 is herein referred to as EU (European) and PRRSV-2 as NA (North American) PRRSV, although both lineages are observed worldwide[8].

The PRRSV genome (14.9–15.5 kb; Figure 1A) is 5′-capped, 3′-polyadenylated and directly translated following release into the cytoplasm[9]. Like most members of the order *Nidovirales*, PRRSV replication includes the production of a nested set of subgenomic (sg) RNAs by discontinuous transcription, where the viral RNA-dependent RNA polymerase (RdRp) jumps between similar sequences in the 3′-proximal region of the genome and the 5′ UTR, known as body and leader transcription regulatory sequences (TRSs), respectively[5,10]. These sgRNAs are 5′- and 3′-co-terminal and are translated to express the structural proteins encoded towards the 3′ end of the genome[5,10]. The 5′-proximal two thirds of the genome contains two long ORFs, ORF1a and ORF1b, with a −1 programmed ribosomal frameshift (PRF) site present at the overlap of the two ORFs[11,12]. Ribosomes that frameshift at this site synthesise polyprotein (pp)1ab, while the remainder synthesise pp1a, both of which are cleaved by viral proteases into several non-structural proteins (nsps)[5,13]. The proteins encoded by ORF1b include the RdRp and the helicase, and frameshifting at this site is thought to set the stoichiometry of these proteins relative to those encoded by ORF1a, a prevalent expression strategy in the *Nidovirales* order[14].

Canonical −1 PRF signals are characterised by two main features, a heptanucleotide "slippery" sequence (SS) which permits re-pairing of the codon:anticodon duplex in the new reading frame, separated by a 5–9 nucleotides (nt) spacer from a downstream RNA structure, often a pseudoknot. This is thought to present a "roadblock" which impedes ribosome processivity over the slippery sequence and stimulates frameshifting[15–20]. In the PRRSV genome, the ORF1ab frameshift signal comprises a U_UUA_AAC slippery sequence (where underscores delineate codons in the 0 frame) and a pseudoknot beginning 5 nt downstream[11,12]. The efficiency of −1 PRF at the PRRSV ORF1ab site has not been measured in the context of infection, but is thought to be around 15–20% based on assays using reporter constructs[21,22].

Recently, the region of the PRRSV genome encoding nsp2 was found to contain a second PRF signal (Figure 1A, inset, WT), conserved in all known arteriviruses except equine arteritis virus (EAV) and wobbly possum disease virus (WPDV)[23–26]. This PRF signal is unusual in that it stimulates both −1 and −2 PRF, enabling production of three variants of nsp2 and rendering it the first example of efficient −2 PRF in a eukaryotic system[23,24]. These three proteins share the N-terminal two-thirds of nsp2 (the 0-frame product), which encodes a papain-like protease (PLP) 2 domain – an ovarian tumour domain (OTU) superfamily protease with deubiquitinase (DUB) and deISGylase activity[27–32]. This has an immune antagonistic effect, and interferon (IFN)-β signalling inhibition has been demonstrated for all three variants of nsp2, most strongly for the frameshift products[32]. After the PRF site, nsp2 contains a multi-spanning transmembrane (TM) domain, thought to promote formation of double-membrane vesicles (DMVs) in the peri-nuclear region and anchor nsp2 to these membranes[33–35]. Ribosomes which undergo −2 PRF at this site translate 169 codons in the –2 frame to produce nsp2TF. This contains an alternative putative multi-spanning TM domain, thought to be responsible for targeting nsp2TF to the exocytic pathway, where it deubiquitinates the PRRSV structural proteins GP5 and M, preventing their degradation[23,36]. Nsp2N, the product of −1 PRF, is a truncated form of nsp2, which is generated following termination of translation at a −1-frame stop codon immediately downstream of the slippery sequence, and is predicted to be cytosolic[23,24].

A second unique feature of the nsp2 PRF site is its non-canonical nature. Rather than an RNA secondary structure, the stimulatory element is a complex of a cellular protein, poly(rC) binding protein (PCBP), and the viral protein nsp1β, bound at a C-rich motif (CCCANCUCC) located 10 nt downstream of the slippery sequence (G_GUU_UUU)[23–26]. How binding of this motif by the protein complex stimulates PRF is uncertain, but it may act as a roadblock analogous to the RNA structures of canonical PRF[24–26,37]. In contrast to RNA structure-directed PRF sites, which are commonly assumed to operate at a fixed efficiency, the *trans*-acting mechanism of PRF stimulation at the nsp2 site presents a potential mechanism for temporal regulation, as observed for cardioviruses – the only other known example of protein-stimulated PRF[38–43]. Frameshift efficiency in EU PRRSV-infected MARC-145 cells at 24 hpi was calculated as 20% for −2 PRF and 7% for −1 PRF[23], however this has not been measured over a timecourse of infection.

In recent years, both low- and high-throughput studies of nidoviruses have highlighted considerably greater complexity in both the transcriptome and translatome than is captured solely by the canonical transcripts and ORFs[12,44–51]. Here, we use ribosome profiling (RiboSeq), a deep-sequencing-based technique which generates a global snapshot of ongoing translation[52], in parallel with RNASeq, to probe viral and host gene expression over a timecourse of PRRSV infection. Host differential gene expression analysis revealed that many of the transcriptional changes upon infection were counteracted by reductions in translation efficiency, indicating a dampened host response, and highlighting the importance of looking beyond transcription when analysing gene expression. On the viral genome, our studies reveal for the first time, a significant increase in frameshift efficiency over the course of infection at the nsp2 −2 PRF site, highlighting arteriviruses as the second example of temporally regulated frameshifting during infection. In addition, we identify several novel viral ORFs, including a highly expressed upstream ORF (uORF), the presence of which is conserved amongst NA PRRSV isolates. In both species of PRRSV, related non-canonical sgRNAs overlapping ORF1b were identified and characterised. These likely facilitate the expression of several of the novel ORFs which overlap ORF1b, and the observation of increased ribosome density in the 3′-proximal region of ORF1b suggests

they may also function to temporally regulate expression of the 3′ region of ORF1b itself. This first application of RiboSeq to an arterivirus uncovers hidden layers of complexity in PRRSV gene expression that have implications for other important viruses.

Results

Viral transcription and translation over a timecourse of infection

PRRSV gene expression was investigated using three viruses: an EU PRRSV isolate based on the Porcilis® vaccine strain (MSD Animal Health; GenBank accession OK635576), NA PRRSV SD95-21 (GenBank accession KC469618.1), and a previously characterised mutant variant (NA PRRSV SD95-21 KO2) which bears silent mutations in the nsp2 PRF site slippery sequence and C-rich motif rendering it unable to bind PCBP, induce −1 or −2 PRF, or produce nsp2N or nsp2TF (Figure 1A, inset)[23–25,32]. MA-104 cells (*Chlorocebus sabaeus*) were infected with EU PRRSV at a multiplicity of infection (MOI) of ~1–3 and harvested at 8 hours post-infection (hpi) following a two-minute pre-treatment with the translation elongation inhibitor, cycloheximide (CHX). MARC-145 cells (a cell line derived from MA-104) were infected with NA PRRSV (WT or KO2 mutant) at MOI 5 or mock-infected and harvested at 3, 6, 9 or 12 hpi by flash-freezing without CHX pre-treatment. Cell lysates were used for ribosome profiling, in which RNase I was added to digest unprotected regions of RNA and ribosomes were purified to isolate ribosome-protected fragments (RPFs) of RNA (Figure 1B). In parallel, aliquots of the same lysates were subjected to alkaline hydrolysis to generate fragments of RNA for RNASeq. Amplicons were prepared, deep sequenced and reads aligned to host (*C. sabaeus*) and viral genomes (Supplementary Table 2) to characterise the transcriptome and translatome of infected cells.

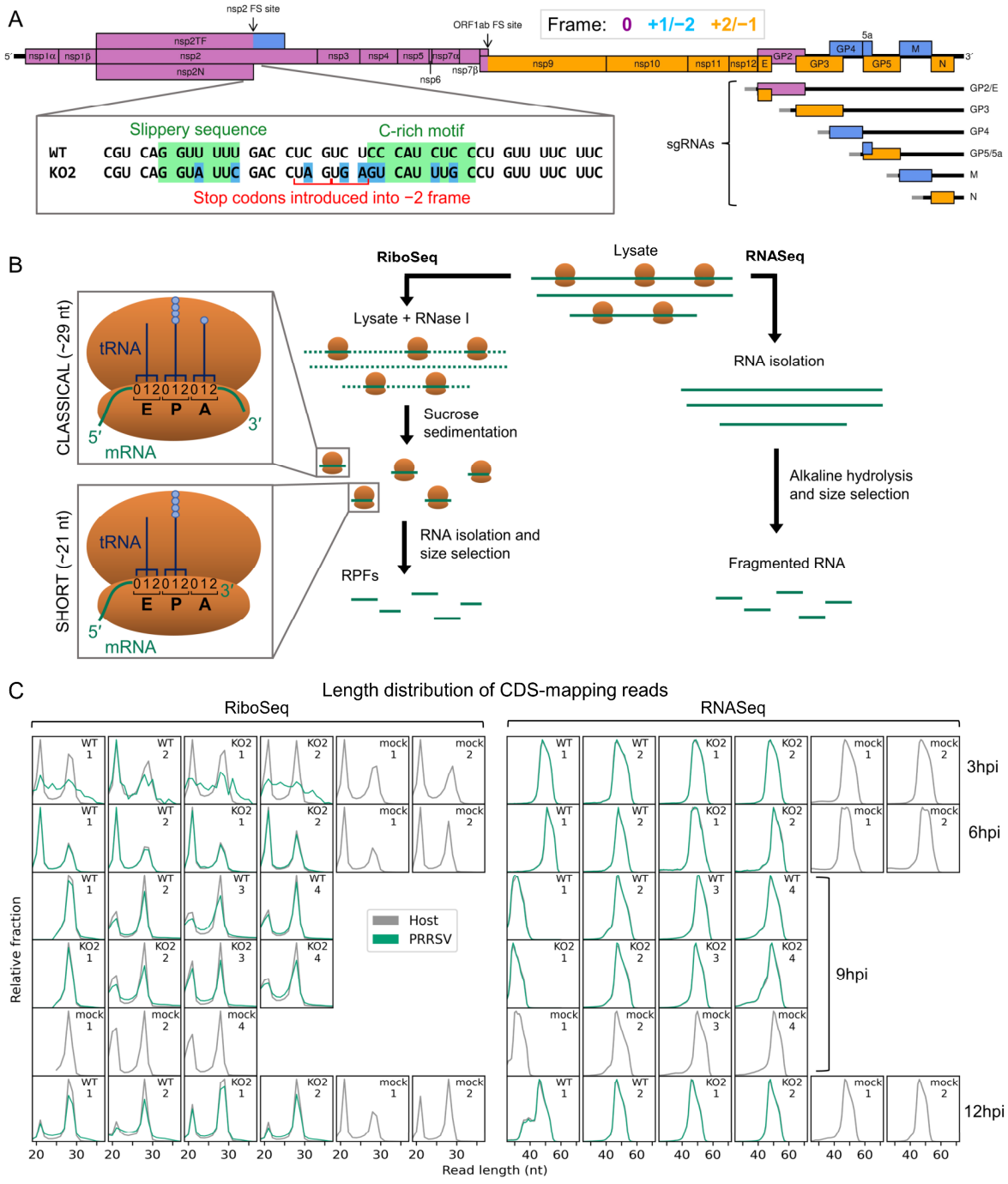


Figure 1. An overview of the experimental set-up and the quality of the datasets.

A) Main: Genome map of NA PRRSV (isolate SD95-21, GenBank accession KC469618.1). ORFs are coloured and offset on the y axis according to their frame relative to ORF1a (0: purple, no offset; +1/–2: blue, above axis; +2/–1: yellow, below axis). Subgenomic RNAs are shown beneath the full-length genomic RNA, with the region of 5′ UTR that is identical to the genomic 5′ UTR shown in grey (known as the "leader"). ORFs translated from each sgRNA are depicted as coloured boxes and named to the right. **Inset:** Nucleotide sequence at the nsp2 PRF site of the NA PRRSV viruses used in this study (SD95-21). Mutations made to disrupt PRF and/or expression of nsp2TF in the KO2 mutant virus are highlighted in blue. All mutations are synonymous with respect to the ORF1a amino acid sequence. **B) Main:** Key experimental steps

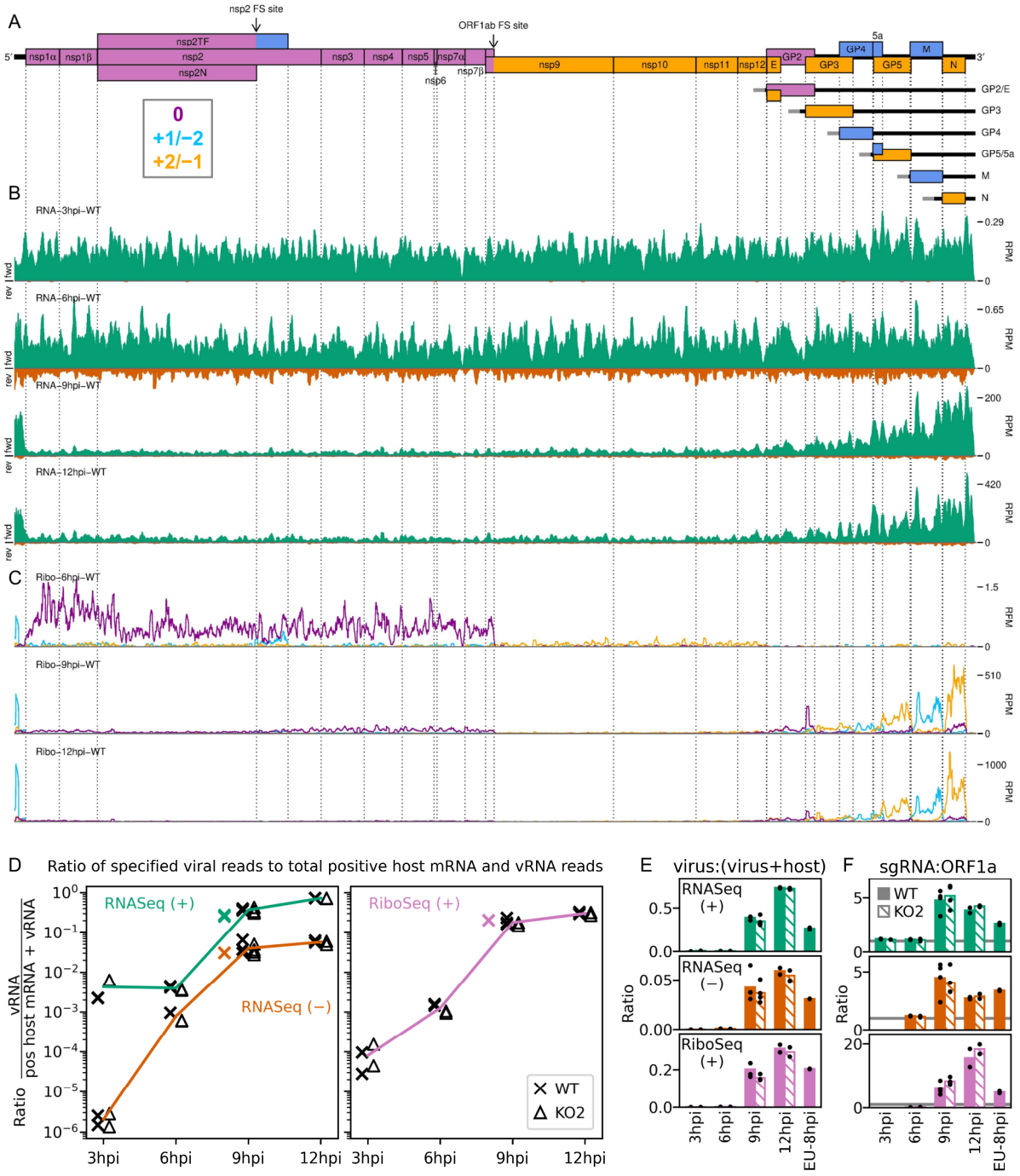
in preparation of RiboSeq libraries (left) and parallel RNASeq libraries (right). **Insets:** Schematics of ribosomes protecting classical length RPFs (A site occupied) and short RPFs (A site unoccupied), with numbers within the decoding centre indicating nucleotide positions within codons. **C)** Length distribution of positive-sense RiboSeq (left) and RNASeq (right) reads mapping within host (grey) or viral (green, mock excluded) CDSs in replicate libraries. For 9 hpi replicate 1 samples (RiboSeq and RNASeq), fragments of 25–34 nt were size-selected during the library preparation; for all other samples the minimum length selected for was 19 nt for RiboSeq and ~45 nt for RNASeq. Note that the RiboSeq library 9 hpi mock replicate three was discarded due to poor quality.

Quality control analyses were performed as previously described[49] (Figure 1C, Supplementary Figures 1–6). The length distribution of CDS-mapping RPFs is observed to peak at ~21 nt (where fragments of this length were purified) and at ~29 nt, with RPFs of these lengths thought to originate from, respectively, ribosomes with an empty A site or an A site occupied by aminoacyl-tRNA (Figure 1B, insets, Figure 1C, Supplementary Figure 6A)[52–56]. Interestingly, the proportion of "short" (19–24 nt) RPFs is significantly lower in the NA PRRSV-infected libraries than mock libraries at late timepoints (9 and 12 hpi grouped; $p = 0.03$ – see Methods), a phenotype that in yeast cells has been attributed to stress-induced phosphorylation of eukaryotic elongation factor (eEF) 2, leading to global inhibition of translation elongation[55]. The predominant distance between the 5′ end of an RPF and the P site of the ribosome is 12 nt in these datasets (Supplementary Figure 1, Supplementary Figure 6B), resulting in CDS-mapping RiboSeq reads showing clear triplet periodicity, known as "phasing", with the majority of RPF 5′ ends mapping to the first position within the codon, known as phase 0 (Supplementary Figure 1, Supplementary Figure 2, Supplementary Figure 6B and C). Together with the observed characteristic length distribution (Figure 1C, Supplementary Figure 6A), this indicates that a high proportion of these reads are genuine RPFs. In contrast, the length and 5′ end position of RNASeq reads is determined by alkaline hydrolysis and size selection, leading to a broader length distribution (Figure 1C, Supplementary Figure 6A) and lack of a clearly dominant phase (Supplementary Figure 1, Supplementary Figure 2, Supplementary Figure 6B and C). Virus CDS-mapping reads show a similar profile to host CDS-mapping reads (Figure 1C, Supplementary Figure 2, Supplementary Figure 6A and C), with the exception of 3 hpi NA PRRSV RiboSeq libraries, in which the background level of non-RPF contamination in the virus-mapping fraction appears to be high relative to the proportion of genuine RPFs, likely due to the low levels of viral translation at this timepoint. These libraries are therefore

181 excluded from all analyses except those in Figure 2D-F and Supplementary Figure 9A, where they
182 provide an upper bound. The subtle flattening of the length distribution and phase composition of virus-
183 mapping reads compared to host-mapping reads in some NA PRRSV RiboSeq libraries at late
184 timepoints (Figure 1C, Supplementary Figure 2) suggests that a small proportion of viral reads originate
185 from non-RPF sources, such as protection from RNase I digestion by viral ribonucleoprotein (RNP)
186 complex formation. This non-RPF fraction of the library (henceforth referred to as RNP contamination
187 although it could originate from several sources) is predominantly noticeable among reads mapping to
188 the ORF1b region of the viral genome (Supplementary Figures 3 and 4), where the read depth from
189 genuine translation is lowest. RiboSeq read lengths for which a high proportion of reads map to phase
190 0 were inferred to be least likely to have a high proportion of RNP contamination (Supplementary
191 Figure 4), and were selected for all NA PRRSV RiboSeq analyses henceforth, unless specified. RNP
192 contamination is not a relevant concern for RNASeq libraries (as proteins are enzymatically digested
193 before RNA purification) and it does not noticeably affect the EU PRRSV RiboSeq libraries, nor RPFs
194 mapping to the host transcriptome (Supplementary Figures 5 and 6). Overall, we inferred that these
195 datasets have a high proportion of RiboSeq reads representing genuine RPFs, and where RNP
196 contamination is evident in lowly translated regions of the viral genome its effects will likely be
197 ameliorated by stratification of read lengths.

198 Having confirmed data quality, we moved on to analyse virus replication over the timecourse by plotting
199 RNASeq and RiboSeq read densities at each position on the viral genome (Figure 2, Figure 3,
200 Supplementary Figures 7 and 8). RNASeq plots revealed a predictable pattern of PRRSV replication
201 and transcription, with low read levels at 3 hpi, likely corresponding to input genomes, evidence for
202 genome replication at 6 hpi, with the appearance of negative-sense reads (Figure 2, Supplementary
203 Figure 7), and high-level synthesis of subgenomic mRNAs at later time points. The observed profile of
204 general virus translation was also consistent with expectation. At 3 hpi (plot not shown), a small number
205 of genuine RPFs were observed (see above) indicating that translation of the NA PRRSV genome is
206 just beginning to reach the level detectable by RiboSeq under these conditions. At 6 hpi, translation of

ORF1ab is robustly detectable and comprises the majority of viral translation (Figure 2, Supplementary Figure 8; mean sgRNA:ORF1a RPF density ratio 0.08), consistent with the lack of significant sgRNA production at this timepoint. At 9 hpi, translation of sgRNAs dominates the landscape, and viral translation represents a sizeable proportion of ongoing translation in the cell (Figure 2, Supplementary Figure 8). Consistent with this, viral nsp1β expression at 9 hpi is clearly detectable by western blotting (Figure 3D and E) and other studies have shown robust expression of viral replicase proteins and viral RNA (vRNA) replication at this timepoint[57,58]. Positive-sense vRNA continues to accumulate between 9 and 12 hpi, although accumulation of the negative-sense counterpart appears to reach a plateau, and at both timepoints, production and translation of sgRNAs is highly favoured over gRNA (Figure 2, Supplementary Figures 7 and 8). This likely represents a transition towards virion formation, for which the main components required are positive-sense gRNA and structural proteins, expressed from sgRNAs. At all timepoints, a large RiboSeq peak in the NA PRRSV 5′ UTR is seen (Figure 2, Supplementary Figure 8), which results from translation of a novel upstream ORF (uORF), discussed below. With the exception of this highly expressed uORF, the transcriptional and translational profile of EU PRRSV at 8 hpi is similar to that of NA PRRSV at 9 hpi, although the production and translation of sgRNAs relative to ORF1a is slightly lower (Figure 2, Figure 3). In all RiboSeq libraries, we noted a variable proportion of negative-sense reads that mapped to the viral genome; however, they do not display the characteristic length distribution or phasing of genuine RPFs (Supplementary Figure 9), suggesting they originate from other sources (discussed above). They are therefore excluded from plots and analyses hereafter.



227

228 **Figure 2. An overview of viral transcription and translation over a timecourse of infection.**

229 **A)** Genome map of NA PRRSV, reproduced from Figure 1A. **B)** RNASeq read densities in reads
230 per million mapped reads (RPM) on the WT viral genome, after application of a 45-nt running
231 mean filter, from cells harvested over a timecourse of 3–12 hpi. Positive-sense reads are plotted
232 in green (above the horizontal axis), negative-sense in orange (below the horizontal axis). The
233 WT libraries with the best RiboSeq quality control results were selected for this plot (3 hpi
234 replicate one, 6 hpi replicate two, 9 hpi replicate four, 12 hpi replicate one), with further replicates
235 and KO2 libraries shown in Supplementary Figure 7. **C)** RiboSeq read densities on the WT viral
236 genome from the counterpart libraries to B. Reads were separated according to phase (0: purple,

–2/+1: blue, –1/+2: yellow), and densities plotted after application of a 15-codon running mean filter. Only read lengths identified as having minimal RNP contamination (indicated in Supplementary Figure 4) were used to generate this plot. Further replicates and KO2 libraries are shown in Supplementary Figure 8. **D)** Ratio of virus-mapping reads to [positive-sense host mRNA- plus positive-sense vRNA-mapping reads]. Virus-mapping reads in the numerator were split into the following categories: positive-sense RNASeq (green), negative-sense RNASeq (orange), and positive-sense RiboSeq (purple). All read lengths were used. The line graphs represent the mean ratios for each category for NA PRRSV, calculated from WT and KO2 data combined. The individual datapoints are also plotted, with WT (cross) offset to the left and KO2 (triangle) offset to the right to aid visualisation. EU PRRSV (8 hpi) ratios are plotted as individual datapoints represented by crosses in the category colour. The RiboSeq (+) 3 hpi timepoint is plotted here to represent the upper limit of the NA PRRSV virus fraction at this timepoint as quality control plots for this fraction of these libraries indicates that they do not contain a high proportion of genuine RPFs, so the true ratio at this timepoint is likely lower. **E)** Data from D represented on a linear scale. Here, data from WT (solid bars) and KO2 (hatched bars) are plotted separately, and individual datapoints are plotted as black circles. **F)** Ratio of the density of sgRNA-mapping reads to ORF1a-mapping reads. All read lengths were used, and densities were calculated as reads per kilobase per million mapped reads (RPKM) of reads from each category in E. Note that the nsp2 frameshift site and downstream region were excluded from the ORF1a region (coordinates of both regions given in Supplementary Table 1). RiboSeq 3 hpi libraries were excluded, and negative-sense RNASeq was omitted from the plot at 3 hpi due to the number of reads being insufficient for robust assessment of the ratio. Categories arranged and plot constructed as in E, with a grey line indicating a ratio of one.

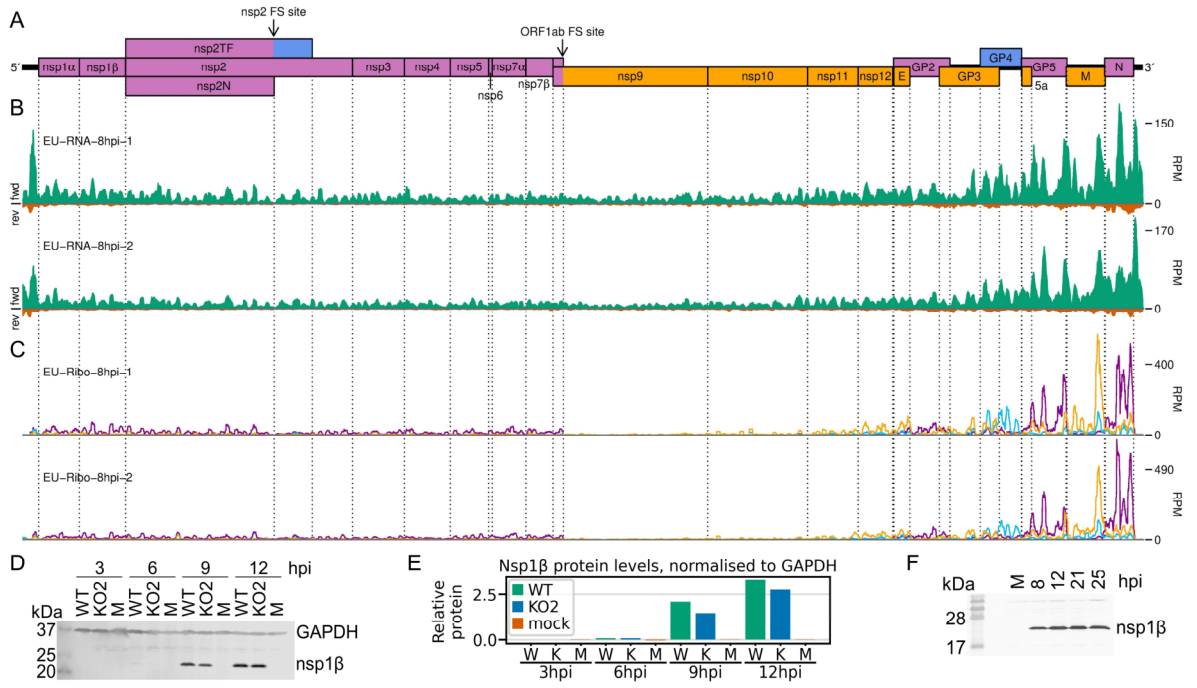


Figure 3. Transcription and translation of the EU PRRSV genome and western blots of nsp1β.

A) Genome map of the EU PRRSV strain used in this study (GenBank accession OK635576). Genome map constructed as in Figure 1A, with subgenomic RNAs omitted for space considerations. **B)** RNASeq read densities on the EU PRRSV genome. Plot constructed as in

Figure 2B. **C)** RiboSeq read densities on the EU PRRSV genome. Plot constructed as in Figure 2C, except for the selection of read lengths to include – in this case, read lengths showing good phasing were selected for inclusion (indicated in Supplementary Figure 6D). **D)** Western blot of lysates used for NA PRRSV ribosome profiling (replicate one samples) with antibodies to viral protein nsp1β (23 kDa) and cellular protein GAPDH (36 kDa) as a loading control. **E)** Quantification of the western blot from panel D to determine the level of nsp1β relative to GAPDH. W = WT, K = KO2, M = mock. **F)** Western blot of nsp1β expression in MA-104 cells infected with EU PRRSV, harvested over a timecourse of 8–25 hpi. M = mock.

Characterisation of the PRRSV transcriptome

As described above, discontinuous transcription by the viral RdRp is an integral part of the nidoviral life cycle. Recent RNASeq studies have revealed considerable complexity in nidoviral transcriptomes beyond the canonical transcripts, including the discovery of numerous novel sgRNAs[46–49,51,59]. Thus far, only a handful of non-canonical transcripts have been discovered for PRRSV, through relatively low-throughput methods. Here, we characterise the PRRSV transcriptome in more detail by examining novel junctions in RNASeq reads aligned to the genome using STAR[60]. Borrowing terminology from the process of splicing (which is unrelated to discontinuous transcription), we refer to the 5′-most and 3′-most positions of the omitted region (where "5′-most" and "3′-most" refer to orientation with respect to the positive-sense genome) as, respectively, the "donor" and "acceptor" sites, and their joining site as the "junction". Neighbouring junctions were merged to account for the potential ambiguity in assigning the exact junction site of discontinuous transcription events, and merged junctions were filtered to keep only those present in more than one replicate within each timepoint (where WT and KO2 were treated as equivalent) to generate one set of junctions per timepoint (Supplementary Tables 3–7; see Methods for details). Junctions for which the donor site spans the TRS (TRS-spanning) are expected to give rise to sgRNAs, while the remaining junctions (non-TRS-spanning) are herein termed "deletions" (unless specified).

Consistent with the trends identified in the general transcriptome analysis (Figure 2), junction-spanning reads attributed to sgRNAs do not pass the filters for detection at early timepoints, but are abundant at 9 and 12 hpi, where they make up the vast majority of viral junction-spanning reads (Figure 4, Figure

5). Canonical sgRNAs are the most abundant transcripts, although reasonably abundant transcript variants are present, which differ only in the length of 5′ UTR between the acceptor site and the CDS start, and are expected to produce the same protein. A study on another arterivirus, simian haemorrhagic fever virus (SHFV), suggests such transcripts may contribute to refining the overall stoichiometry of structural proteins[45]. For the N transcript, NA PRRSV isolates VR-2332 and tw91 have each been shown to have a (different) abundant secondary transcript variant[12,61]. Both of these are observed in our NA PRRSV dataset, although usage of the secondary body TRS found in VR-2332 (beginning at position 14,875 on the SD95-21 genome) was much more frequent than that of tw91, consistent with the fact that SD95-21 is more closely related to VR-2332. This more abundant secondary transcript, herein termed N-short, has a 5′ UTR 114 nt shorter than that of the NA PRRSV primary transcript (herein termed N-long, with body TRS beginning at position 14,761), presenting a potential opportunity for differential translation regulation. If such regulation exists, it is unlikely to be temporal, as the ratio of N-long to N-short remains constant, at approximately 6:1, between 9 and 12 hpi. Any such regulation would also likely be isolate-dependent, as the N-short body TRS is not completely conserved amongst NA PRRSV isolates, and species-dependent, as the N-long body TRS is neither highly conserved nor highly utilised in EU PRRSV, for which N-short is ~60-fold more abundant than any other N transcript and its body TRS is absolutely conserved.

In addition to the numerous novel sgRNAs predicted to express full-length structural proteins, we found that most canonical sgRNAs have transcript variants with body TRSs downstream of the start codon, which are expected to express truncated forms of the structural proteins (Figure 4B, Figure 5). One of these was also observed for VR-2332 PRRSV: the “5-1” transcript variant[12], which is thought to express a truncated form of GP5, and is present in our NA PRRSV dataset at ~1.7% of the abundance of the primary GP5 transcript (based on the number of junction-spanning reads at the donor site). Similar GP5 transcript variants were observed in SHFV, and mutagenesis studies suggest that the truncated GP5 may be beneficial for viral fitness[45], raising the possibility that the putative truncated forms of this and other PRRSV structural proteins could be functional.

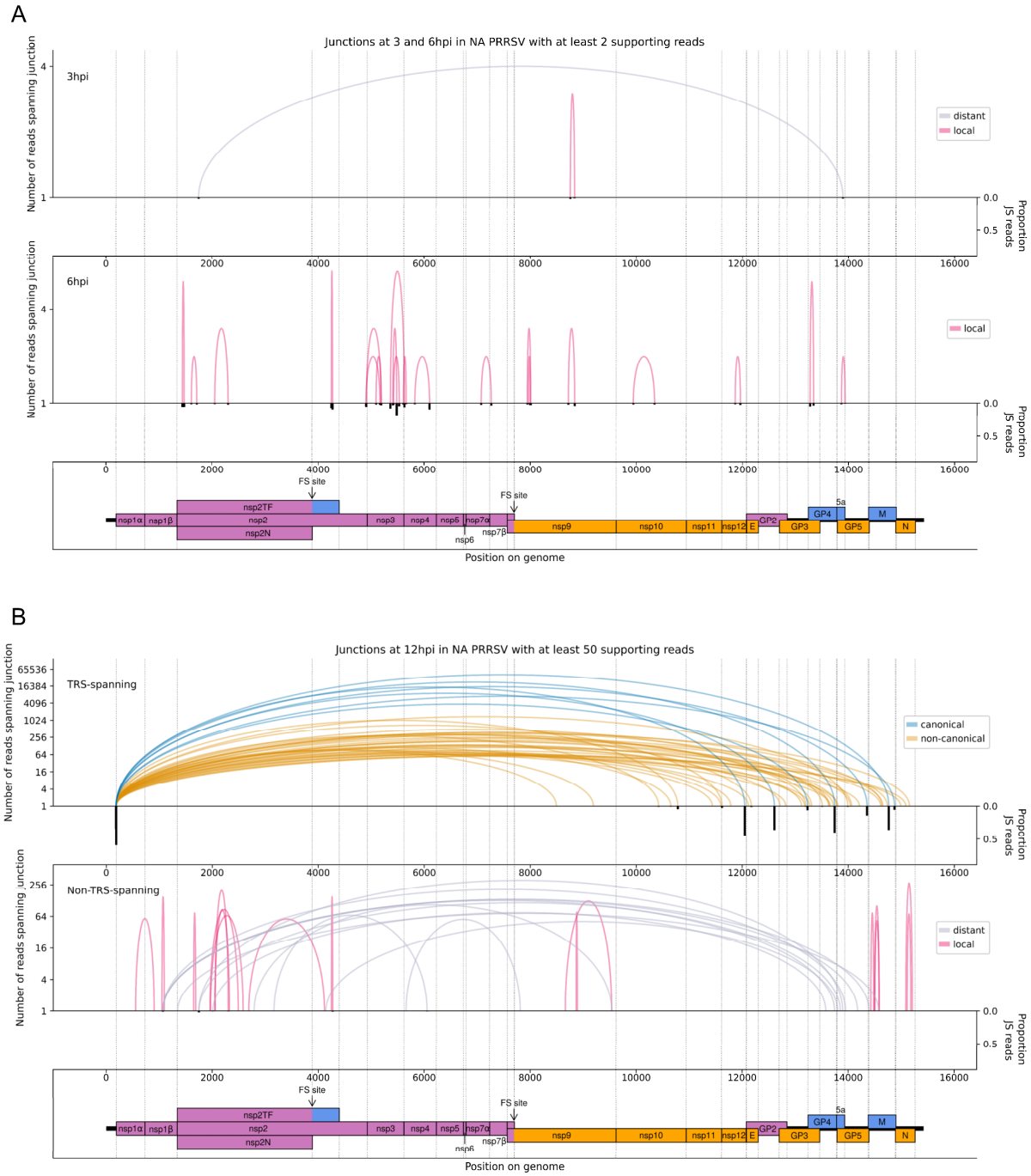


Figure 4. The NA PRRSV transcriptome at A) 3 and 6 hpi and B) 12 hpi.

A) Sashimi plot of junctions in the NA PRRSV dataset at early timepoints during infection. The number of reads spanning each junction is indicated by the highest point of its arc (note the logarithmic scale of the y axis) and represents the total number of reads spanning the junction in all libraries from the specified timepoint combined. Only junctions for which this number is ≥ 2 are plotted. Beneath the sashimi plot is an inverted bar chart (black) of the proportion of reads at each donor and acceptor site that span the junction of interest, calculated as junction-spanning / (junction-spanning + continuously aligned to reference genome) and plotted on a linear scale (see Methods for details). A list of all junctions (from all timepoints) and their associated proportions of junction-spanning (JS) reads is shown in Supplementary Tables 3–7. At both timepoints, no

junctions where the donor site overlapped the leader TRS passed the minimum read-count threshold for plotting. Internal deletions, in which the donor site does not overlap the leader TRS, are coloured according to whether the deletion is distant (> 2000 nt deleted, grey) or local (≤ 2000 nt deleted, red). **B) Upper:** Sashimi plot of junctions for which the donor site overlaps the leader TRS, with the major junction for each canonical sgRNA shown in blue (including both N-long and N-short for the N sgRNA), and other junctions ("non-canonical") shown in orange. Both upper and lower panels were constructed as in panel A except that the threshold for inclusion of junctions was adjusted to ≥ 50 supporting reads. **Lower:** Sashimi plot of junctions representing internal deletions.

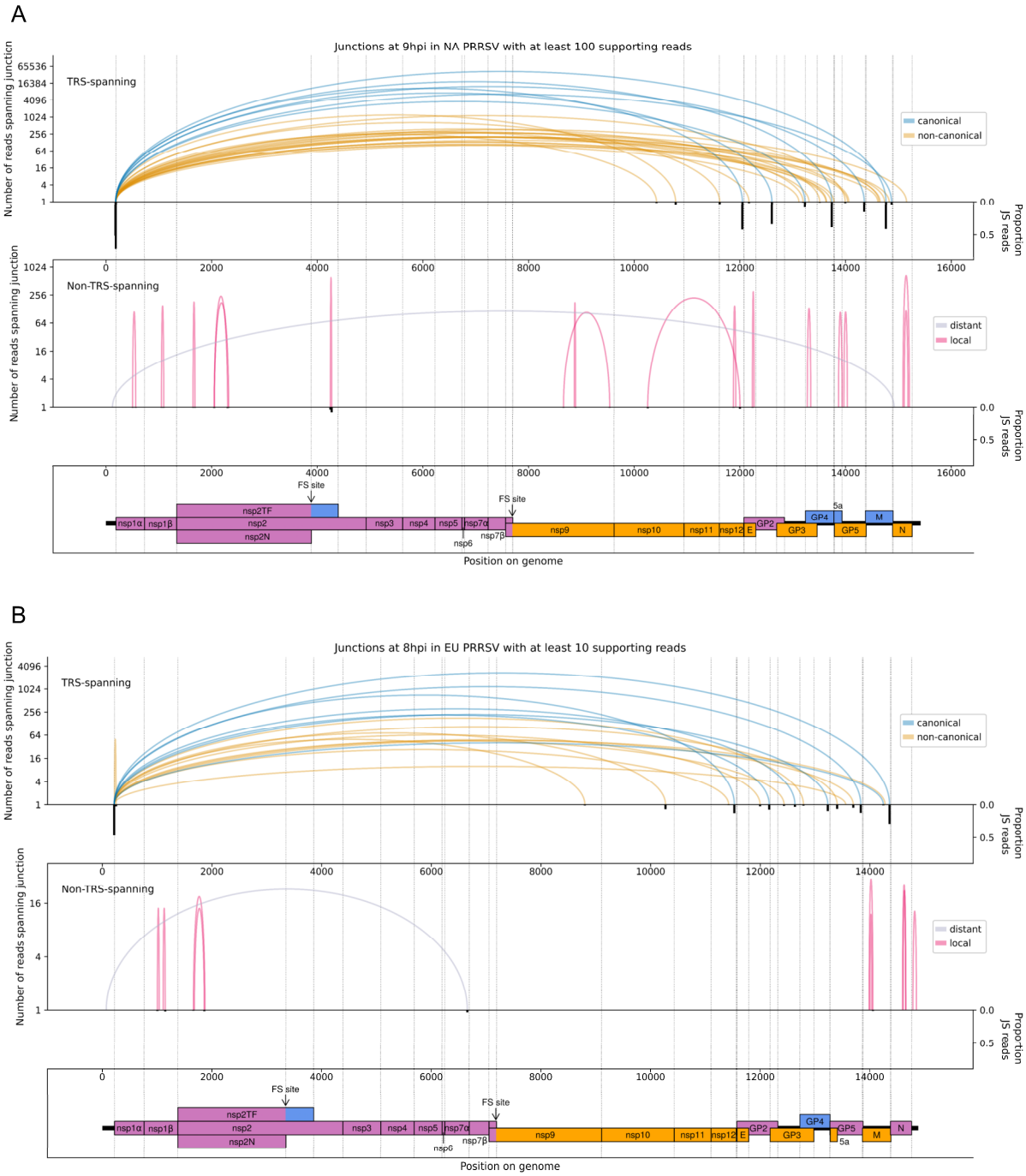


Figure 5. The NA and EU PRRSV transcriptomes at 9 and 8 hpi, respectively.

A) Sashimi plots of junctions for NA PRRSV at 9 hpi. Plots constructed as in Figure 4B, but with the threshold for inclusion of junctions adjusted to ≥ 100 junction-spanning reads in total from all 9 hpi libraries (note that eight libraries were analysed at this timepoint compared to four at other timepoints). **B)** Sashimi plots of junctions for EU PRRSV at 8 hpi. Plots constructed as in Figure 4B, but with the threshold for inclusion of junctions adjusted to ≥ 10 supporting reads (note that only two libraries were analysed and shorter read lengths are expected to lead to fewer identifiably junction-spanning reads).

In addition to the transcript variants for the structural proteins, a small number of non-canonical sgRNAs were discovered in both NA and EU PRRSV which have acceptor sites within ORF1b, herein termed ORF1b sgRNAs (Figure 4B, Figure 5). This was unexpected as ORF1b is thought to be expressed only from gRNA, which is much less abundant than sgRNAs at late timepoints and is relatively inefficiently translated (see below). ORF1b sgRNAs, even of low abundance relative to canonical sgRNAs, could therefore have a significant effect on the expression of polyprotein products. This is explored further below.

Deletions, in which the leader TRS is not the donor site, tend to have fewer junction-spanning reads than sgRNAs, but nonetheless may influence gene expression. Many of these likely represent defective interfering (DI) RNAs; however, several of the long-range deletions in the NA PRRSV 12 hpi dataset bear similarity to "heteroclite" sgRNAs, a family of non-canonical transcripts found in several NA PRRSV isolates[44,62]. Heteroclite sgRNA formation is thought to be directed by short (2–12 nt) regions of similarity between the donor site, located within ORF1a, and the acceptor site, usually located within the ORFs encoding structural proteins[44,62,63]. These transcripts can be packaged into virions but, unlike classical DI RNAs, they do not appear to interfere with canonical gRNA or sgRNA production and are present in a wide range of conditions, including low MOI passage and samples directly isolated from the field[44,62]. In our datasets, the most abundant deletion at 12 hpi is identical to the junction that forms the "S-2" heteroclite sgRNA for VR-2332 PRRSV, from which a fusion of the first 520 amino acids of ORF1a (nsp1α, nsp1β and part of nsp2) and the last 11 amino acids of 5a is thought to be expressed[44,62]. At 12 hpi, 2.3% of reads at the donor site span this junction, which indicates this is a relatively minor transcript relative to gRNA; however, this could be enough to affect gene expression, for example it is only ~5-fold lower than the corresponding percentage for the major GP4 junction, the canonical sgRNA with the fewest junction-spanning reads. Although this junction is not present above the limit of detection at 6 hpi, it is observed at 9 hpi (Supplementary Table 5; total read counts below the threshold for inclusion in Figure 5A) and at 3 hpi (Figure 4A, upper), consistent with this transcript being packaged into virions[44,62]. No transcripts resembling heteroclite sgRNAs were detected for EU PRRSV,

377 although it is possible such transcripts might be observed if a later timepoint was sampled and/or longer
378 RNASeq inserts were generated, as the shorter read lengths purified for these libraries (and NA PRRSV
379 9 hpi replicate one) are less amenable to detection of junctions.

380 The numerous novel transcripts described in this section not only present opportunities for regulation
381 of the known PRRSV proteins, but also highlight considerable flexibility in the transcriptome, which
382 provides a platform for expression of truncated protein variants and novel ORFs. Nonetheless, it is
383 likely that many of the lowly abundant novel transcripts are simply an unavoidable consequence of a
384 viral replication complex that has evolved to facilitate discontinuous transcription as an essential
385 component of the viral life cycle.

386 **Characterising the PRRSV translatome**

387 To characterise the viral translatome, RiboSeq reads were mapped to the host and viral genomes using
388 STAR, which formed the input for PRICE[64]. PRICE detected 14 novel NA PRRSV ORFs and eight
389 novel EU PRRSV ORFs (Figure 6, Supplementary Figures 10 and 11, Supplementary Table 8). An
390 additional NA PRRSV library, which had been harvested after CHX pre-treatment, was also inspected,
391 as CHX pre-treatment may emphasise initiation peaks (albeit less efficiently than specific initiation
392 inhibitors such as harringtonine)[52,65]. For NA PRRSV (Figure 6A), four of the novel ORFs overlap the
393 ORFs encoding the structural proteins and may be expressed from the array of non-canonical sgRNAs
394 discovered in this part of the genome. Most of the other novel ORFs overlap ORF1b and are likely
395 expressed from the novel ORF1b sgRNAs described above, consistent with the fact that their translation
396 is predominantly observed at late timepoints (Figure 6B, Supplementary Figure 10). Some of these
397 ORFs are highly translated – for example the 125-codon NA PRRSV nsp12-iORF is translated at a level
398 similar to nsp12 at 12 hpi (Figure 6C). To test whether these novel ORFs in either virus are subject to
399 purifying selection (an indicator of functionality), we analysed synonymous site conservation within
400 the known functional viral ORFs (Figure 6A and Supplementary Figure 11A). Overlapping functional
401 elements are expected to place additional constrains on evolution at synonymous sites, leading to local

peaks in synonymous site conservation. While such peaks were observed in the regions where the known viral ORFs overlap (and also within the M ORF and at the 5′ end of ORF1a), no large conservation peaks were observed in the vicinity of the novel, translated overlapping ORFs, indicating their functional relevance is debatable.

As mentioned earlier, we also identified a uORF in the NA PRRSV 5′ UTR (Figure 6D), which is highly expressed at all timepoints (Figure 2C, blue peak). At only ten amino acids, the peptide expressed from this uORF is unlikely to be functional, and the ORF is truncated or extended in a small proportion of isolates (Figure 6E). However, the presence of a uORF in this position is highly conserved in NA PRRSV (Figure 6E–G), with the initiator AUG conserved in 558/564 available sequences, and relatively efficient[66] non-canonical initiation codons (GUG, AUA or ACG) in the remainder. This suggests the uORF may have advantages for viral fitness, for example by modulating translation of other ORFs.

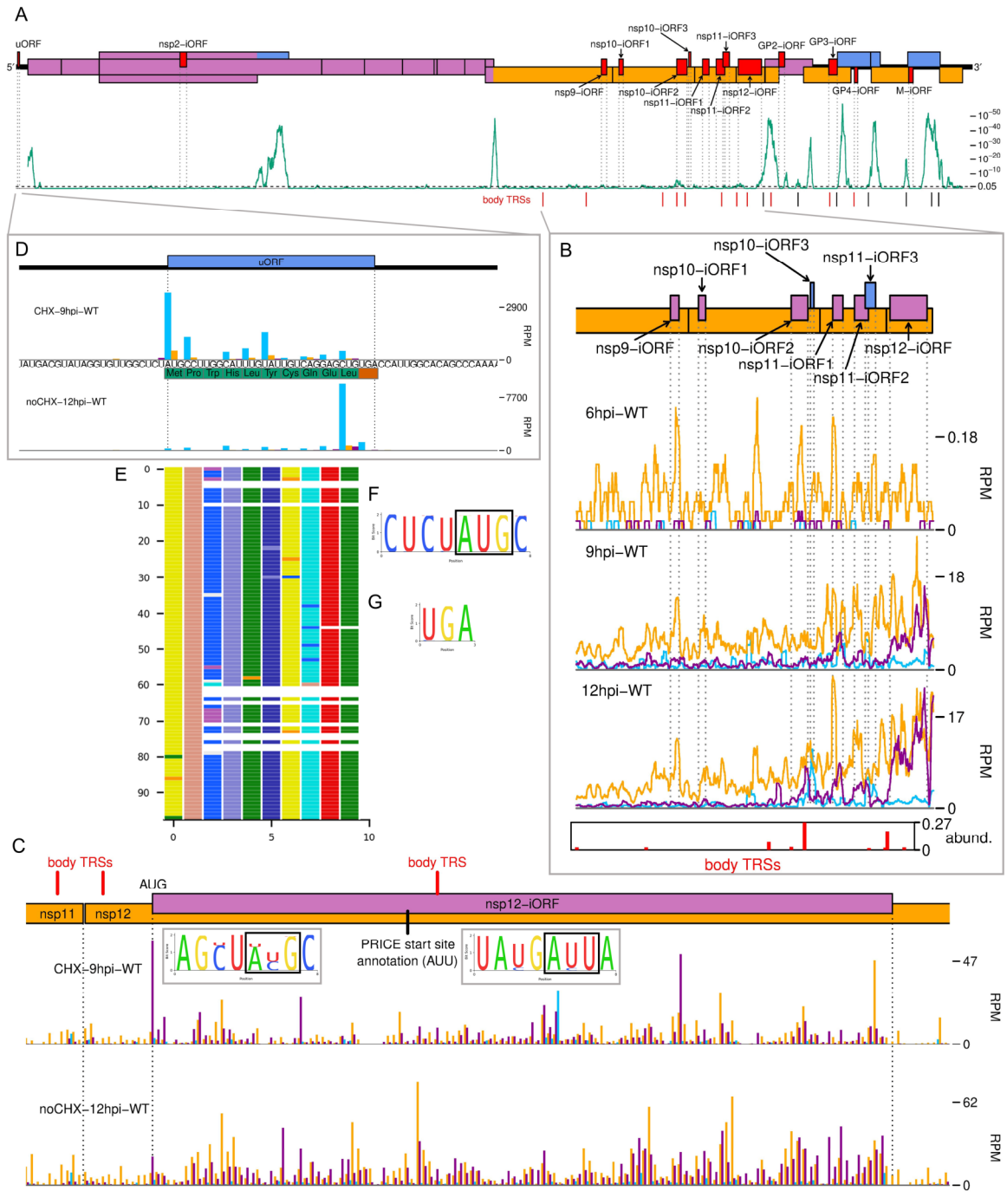


Figure 6. The NA PRRSV translatome.

A) Locations of novel ORFs in the NA PRRSV genome. The genome map of canonical ORFs is reproduced, without labels, and novel ORFs identified by PRICE are overlaid in red, offset on the y axis according to frame relative to ORF1a. Below this is a SYNPLOT2[67] analysis of synonymous site conservation in the canonical protein-coding regions, based on 137 NA PRRSV genomes representative of NA PRRSV diversity (see Methods). The green line represents the probability (over a 25-codon sliding window) that the observed conservation could occur under a null model of neutral evolution at synonymous sites; conservation peaks are indicative of overlapping functional elements. Locations of selected body TRSs are indicated below, with known body TRSs shown in black and novel body TRSs in red, displaying only the major body

TRSs for ORF1b sgRNAs and/or novel sgRNAs expected to facilitate expression of a novel ORF. **B)** Translation of novel ORFs overlapping ORF1b. Reads mapping to the NA PRRSV genome between the ORF1b sgRNA 1 body TRS and the end of ORF1b are shown, separated according to phase and plotted after application of a 15-codon running mean filter. Only read lengths identified as having minimal RNP contamination (indicated in Supplementary Figure 4) were used to generate this plot. Novel ORFs in this region are indicated on the genome map, coloured and offset according to frame relative to ORF1a. Positions of moderately frequently used body TRSs (at least 44 junction-spanning reads) are indicated by red bars at the bottom of the plot. The height of each red bar is scaled according to the number of junction-spanning reads for that body TRS (across all 12 hpi RNASeq libraries combined), relative to the canonical sgRNA with the fewest junction-spanning reads (GP4, major junction only). "abund." = abundance (of junction-spanning reads relative to GP4). The libraries displayed are those in Figure 2C, with remaining replicates and KO2 libraries in Supplementary Figure 10. **C) Main:** RPF distribution on the region of the NA PRRSV genome predicted to contain nsp12-iORF. RPFs are coloured according to phase and plotted without application of a sliding window, using only read lengths identified as having minimal RNP contamination (indicated in Supplementary Figure 12F for CHX-9hpi-WT and Supplementary Figure 4 for noCHX-12hpi-WT, which is replicate one). The positions of body TRSs with ≥ 50 junction-spanning reads at 12 hpi are indicated by a red bar at the top of the genome map. The final initiation codon predicted by PRICE is an AUU codon, indicated by a black line. However, the observed RPF profiles are more consistent with the N-terminally extended ORF annotated in this plot, for which ribosomes would initiate at the upstream AUG, which was designated as the start site by PRICE for the "Candidate location" before application of start site selection algorithms (coordinates in Supplementary Table 8). **Insets:** Conservation of the context of the predicted nsp12-iORF initiation codon for the "Candidate location" (left) and final "Location" (right), based on 661 available sequences for full NA PRRSV genomes. The putative initiator codons are indicated by black boxes. **D)** Distribution of RPFs mapping to the region of the NA PRRSV 5′ UTR containing the uORF. Plot constructed as in panel C, with the genome sequence in this region, and the uORF amino acid sequence, underlaid. Note that CHX pre-treated libraries typically have heightened initiation peaks, while libraries harvested without CHX pre-treatment have heightened termination peaks. **E)** Predicted amino acid sequences of the uORF from 98 PRRSV genomes representative of NA PRRSV diversity. The uORF CDS nucleotide sequence was extracted from a multiple sequence alignment and frame 0 was translated. Each row represents one sequence, with each coloured rectangle representing an amino acid (logo plots and alignment visualisations made using CIAlign[68]). Gaps indicate translation is predicted to have terminated due to a stop codon. One sequence out of these 98 (KY348852) has a 28-codon extension to the ORF which is not depicted. **F and G)** Conservation of F) the initiation context and G) the stop codon for the NA PRRSV uORF, based on 661 NA PRRSV sequences. Sequences were filtered to take only those spanning the entire feature of interest with no gaps, leaving F) 564 and G) 598 sequences in the alignment used for the logo plots. The initiator AUG is indicated by a black box. The initiation context of this ORF is weak, as defined by the absence of a G at position +4 or a A/G at position –3 relative to the A of the AUG, but the sequence is highly conserved.

Quantification of viral gene expression

Next, we quantified viral transcription and translation to better understand PRRSV gene expression profiles and determine the contribution of the novel transcripts and ORFs. RiboSeq read density (in reads per kilobase per million mapped reads [RPKM]) was calculated using the PRICE output, and transcript abundance was quantified based on the number of junction-spanning reads (in RPM). ORFs were paired with the transcripts from which they are most likely expressed (Figure 7), and translation efficiency (TE) calculated as the RiboSeq read density divided by the transcript abundance. Although TE calculations can be confounded by factors such as differences in translation speed, such effects are expected to average out over longer regions such as whole ORFs. To facilitate this calculation for the uORF, "transcript abundance" for the 5′ leader was determined as though this was a separate transcript (see Methods).

Consistent with the results shown in Figure 2, gRNA is by far the most abundant viral transcript at 6 hpi, after which there is a marked shift towards sgRNA production at 9 hpi (Figure 8A and B, Supplementary Figure 13). Between 9 and 12 hpi, the proportion of gRNA increases slightly; however, this may be partly related to changes in abundance of heteroclite sgRNAs, which are not discriminated from gRNA in this analysis, and are investigated separately below. The relative abundance of each sgRNA remains fairly stable between 9 and 12 hpi, consistent with findings for SHFV and MHV[45,49], and non-canonical sgRNAs make up a relatively small proportion of the viral transcriptome. The results for negative-sense transcripts broadly mirror the positive-sense results, although negative-sense gRNA is proportionally more abundant (Figure 8A, Supplementary Figure 13). The overall transcriptional profile of EU PRRSV resembles that of NA PRRSV at 9 hpi, although gRNA is more abundant and there are some differences in the relative proportions of canonical sgRNAs (Figure 8A and B, Supplementary Figure 13).

Analysis of RiboSeq datasets revealed a similar trend to the RNASeq analysis of NA PRRSV, with ORF1a translation predominating at 6 hpi, while sgRNA translation dominates at 9 hpi (Figure 8C,

493 Supplementary Figure 14). ORF1a translation declines by 12 hpi, despite the increase in transcript
494 abundance, perhaps representing the sequestration of gRNA through genome packaging, reducing the
495 pool available for translation. Consistent with this, ORF1a, which has relatively low TE throughout
496 infection, is the only canonical NA PRRSV ORF for which TE decreases over time, decreasing from
497 ~20 at 6 and 9 hpi to 4.4 at 12 hpi (Supplementary Figure 15; explored in more detail below). Strikingly,
498 the uORF is among the most highly translated NA PRRSV ORFs at all timepoints (Figure 8C,
499 Supplementary Figure 14), although this may be somewhat inflated by the heightened termination peak
500 having a proportionally greater effect on RPKM for a small ORF such as this. Its high expression at 6
501 hpi indicates that the uORF is expressed from gRNA, as this is by far the most abundant viral transcript
502 at this timepoint (Figure 8B). The increase in the ratio of uORF compared to ORF1a translation at later
503 timepoints (Figure 8C, Supplementary Figure 14), when sgRNAs become abundant, suggests that the
504 uORF is also translated from the sgRNAs. This is supported by the TE results (Supplementary Figure
505 15), which show an increase in uORF TE between 9 and 12 hpi, and a concomitant increase in TE for
506 all canonical sgRNA ORFs (except GP2 which remains stable), consistent with increased TE of
507 sgRNAs affecting both the uORF and the main ORFs.

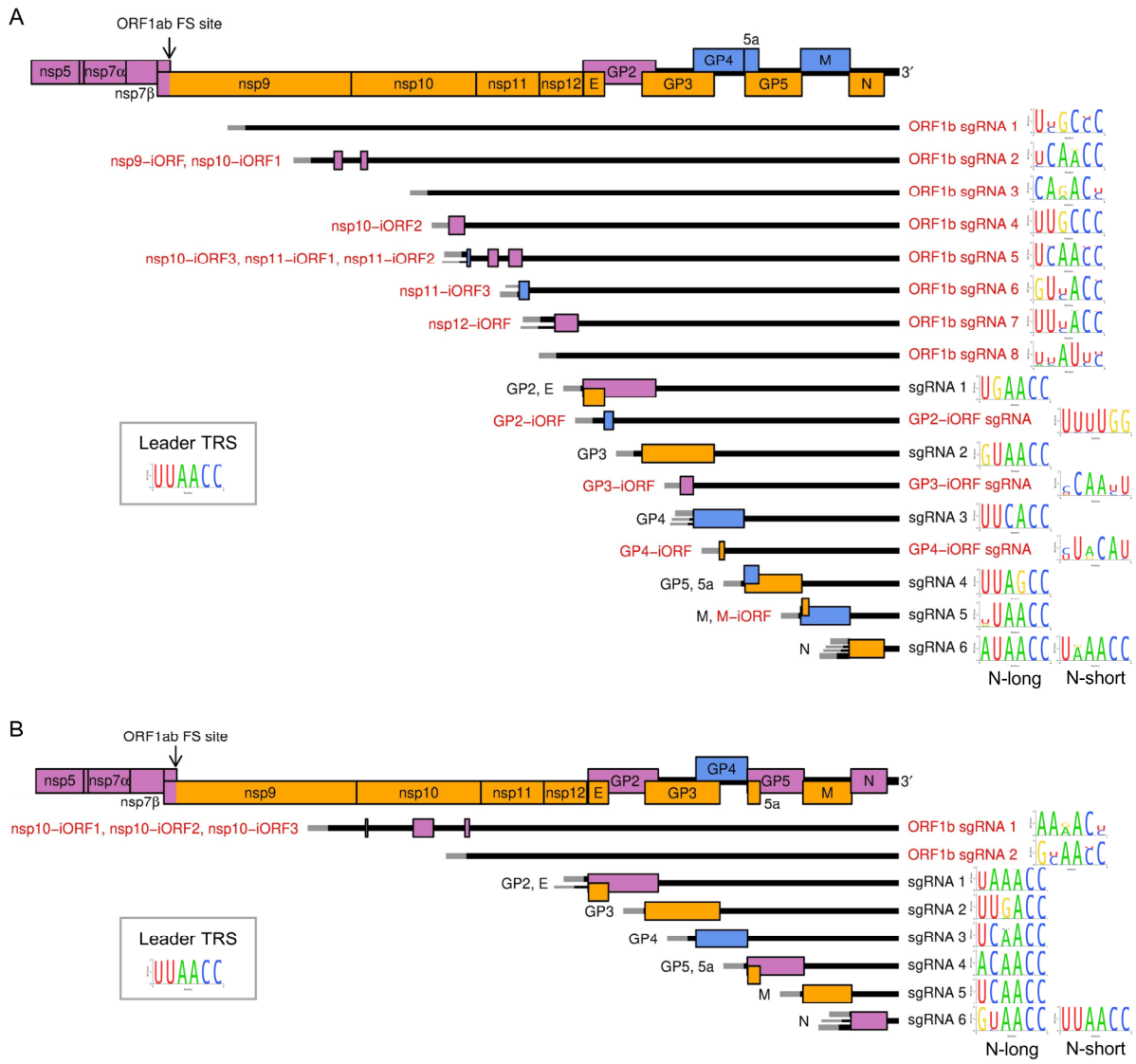


Figure 7. Subgenomic mRNA transcripts and ORFs included in viral gene expression analysis of A) NA PRRSV and B) EU PRRSV.

Canonical transcripts and ORFs are labelled in black, novel ones in red. The genome map from nsp5 onwards is reproduced above for comparison. The leader (grey) is treated as a separate transcript for the purposes of these analyses, and the NA PRRSV uORF putatively expressed from it was omitted from these plots for clarity. Where more than one 5′ UTR is depicted for some mRNAs this indicates that multiple merged junctions were detected that likely give rise to transcripts from which the same ORF(s) are translated. In these cases, the alternative transcripts were considered as one species in the gene expression analysis, and junction-spanning read counts for the junctions were combined. To the right of each transcript, the consensus sequence of the body TRS used to generate the major transcript variant (indicated by the thicker UTR) is plotted, based on A) 661 NA PRRSV or B) 120 EU PRRSV genome sequences. For ease of identification, both N-long and N-short are depicted as major transcripts for N. In addition to these sgRNAs and ORFs, ORF1a and all novel ORFs not depicted here were included in the analysis and designated as expressed from the gRNA transcript.

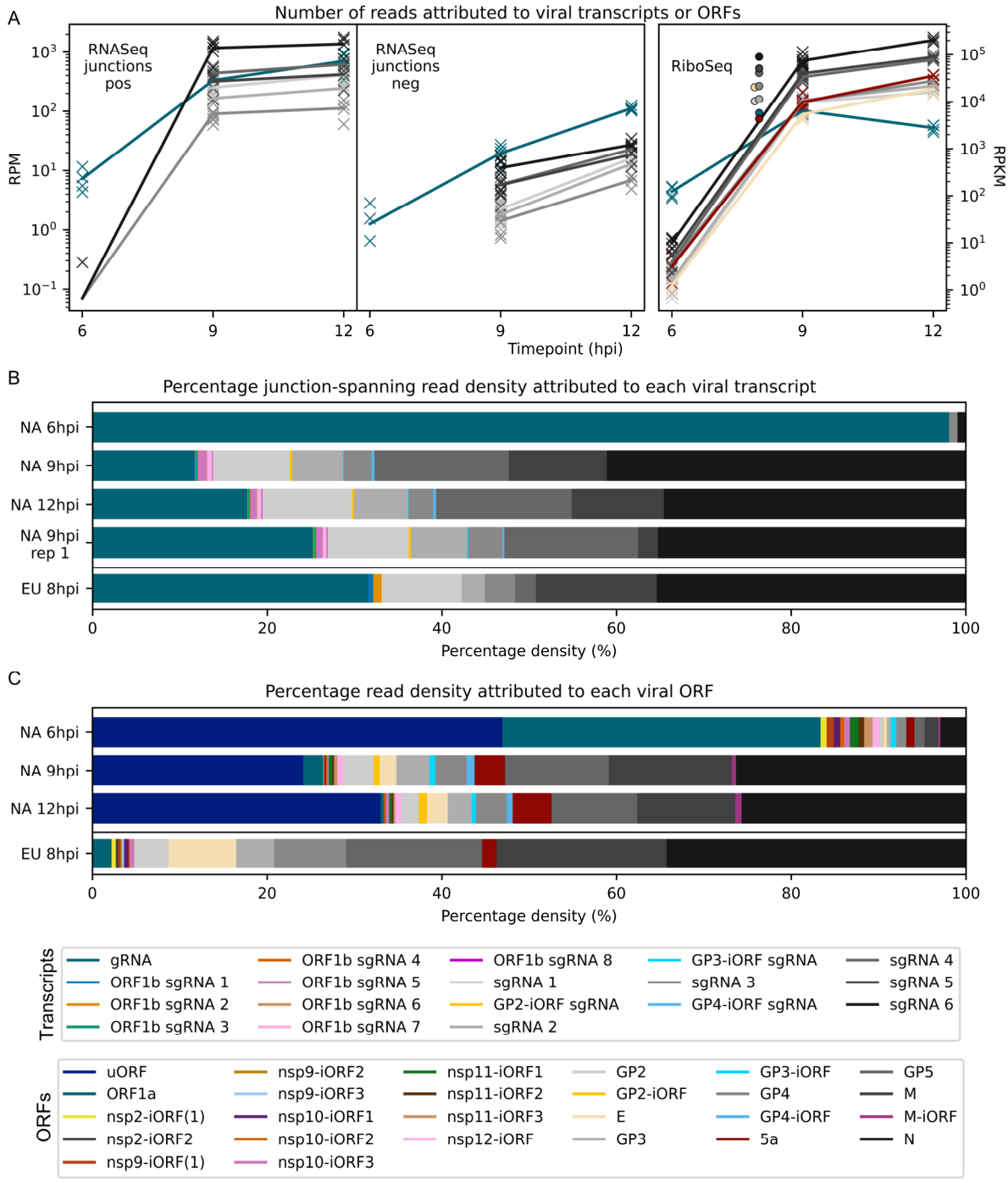


Figure 8. Viral transcript abundance and total translation of viral ORFs.

A) Left: Junction-spanning read density attributed to canonical viral transcripts. The left-most panel shows the results for positive-sense reads (pos), with negative-sense reads (neg) on the right. Mean values are indicated by the line graph, with individual data points plotted as crosses. Note that, due to the shorter RNA fragment lengths selected for EU libraries and for NA 9 hpi replicate one libraries, the junction-spanning read counts are not comparable to the remaining NA PRRSV libraries. For all junction-spanning read analyses in which averages were taken, replicate one libraries were excluded from the "9 hpi" group and analysed separately (in this case see Supplementary Figure 13, for these and the EU PRRSV results). **Right:** RiboSeq read density

attributed to canonical viral ORFs, based on the PRICE read count values. Mean values for EU PRRSV are plotted as filled circles, with individual data points omitted for clarity and some circles offset on the x axis to aid visualisation. ORF1b is omitted from this and several other plots in this section, and investigated separately in Figure 9A-C. Investigation of variations in transcription and translation within ORF1a are given in Figure 9D; for all other analyses in this section, "ORF1a" refers to the region designated by PRICE, which begins at genomic coordinates 2249 (NA) and 1212 (EU) and extends to the end of ORF1a. Similarly, "gRNA" transcript abundance is calculated ignoring the existence of putative heteroclite sgRNAs, except for where this is investigated in Figure 9D. The legend for colours in all panels is displayed beneath panel C. **B)** Percentage of the viral transcriptome represented by each transcript. Transcript abundances (estimated from junction-spanning read RPM values) within each library were converted to a percentage of the total RPM of all transcripts plotted for that library. Mean percentages for each group were calculated, treating WT and KO2 as equivalent. The leader, which is treated as a separate transcript for future junction-spanning read analyses, was omitted from this percentage calculation. **C)** Percentage of the viral translatome represented by each ORF. Plot constructed as in panel B, with percentages representing the RPKM value of each ORF (calculated from the PRICE output) as a proportion of the total sum of RPKM values of all ORFs plotted. Note that the novel ORFs detected on the EU PRRSV genome were named according to the same convention as for NA PRRSV novel ORFs, but equivalent names does not indicate that they are equivalent ORFs.

Except for the absence of a uORF, the relative translation levels of EU PRRSV ORFs are similar to those in NA PRRSV, although with less translation of 5a (Figure 8C, Supplementary Figure 14). This may reflect the different relative arrangements of GP5 and 5a for these two isolates, with 5a beginning 5 nt downstream of the beginning of GP5 for EU PRRSV and 10 nt upstream for NA PRRSV. TE values for EU PRRSV are slightly higher than those for NA PRRSV (Supplementary Figure 15); however, this may be influenced by reduced accuracy of transcript abundance quantification due to the shorter read lengths of the EU libraries.

Novel ORFs make up a relatively small proportion of total viral translation (Figure 8C). Nonetheless, they may represent a significant contribution to the viral proteome – for example, the novel ORFs overlapping the end of ORF1b have a similar density of ribosomes as ORF1a at 12 hpi (Figure 8C, Supplementary Figure 14). These overlapping ORFs are not subject to noticeable purifying selection (Figure 6A, Supplementary Figure 11A), indicating they are unlikely to produce functional proteins. This raises the possibility that their translation is tolerated as a side effect of ORF1b sgRNA production, which may primarily function to regulate expression of ORF1b. This is supported by the observed step increases in ORF1b-phase RiboSeq density after some of the ORF1b sgRNA body TRSs at late timepoints (Figure 9A and B), a trend confirmed by quantification of this read density in the regions

between these body TRSs (Figure 9C). At 6 hpi, when no ORF1b sgRNAs are detected, read density remains reasonably constant throughout ORF1b, while at later timepoints, as ORF1b sgRNA expression increases, a pattern of increasing density towards the 3′ end of ORF1b emerges, with the 3′-most regions more highly translated than ORF1a (Figure 9C). For NA PRRSV, the greatest step increases are observed after the ORF1b sgRNA 2, 5 and 7 body TRSs (Figure 9A and C) – the only non-canonical sgRNAs in Figure 7 which have just a single mismatch in the body TRS compared to the leader TRS. These body TRSs are also well-conserved, particularly the final two Cs, identified as the most highly conserved part of the canonical sgRNA body TRS consensus in this and other studies[12,61] (Figure 7). This raises the likelihood that such body TRSs may also produce ORF1b sgRNAs in other isolates of NA PRRSV. Furthermore, although the body TRSs for the EU PRRSV ORF1b sgRNAs are less well-conserved within the species (Figure 7), they are located at very similar positions on the genome compared to the NA PRRSV ORF1b sgRNA 2 and 5 body TRSs, which correlate with two of the greatest increases in ORF1b-phase RiboSeq read density for NA PRRSV. Indeed, the EU PRRSV ORF1b sgRNA 2 body TRS is in a genomic location exactly equivalent to that of NA PRRSV ORF1b sgRNA 5, and both body TRSs have only a single mismatch compared to the leader TRS, despite this not being a requirement for maintaining the amino acid identities at this position. The conservation of these features of ORF1b sgRNAs between these two highly divergent arterivirus species suggests there may be a selective advantage in their production, which could result from temporal modulation of the stoichiometry of nsps 10–12.

Similarly, the heteroclite sgRNAs have the potential to modulate the stoichiometry of ORF1a. To examine this, the RNASeq read density in ORF1ab was partitioned between gRNA and heteroclite sgRNAs (a distinction not made in the junction-spanning read analysis) using a “decumulation” procedure introduced in Irigoyen *et al.*[49], and RiboSeq read density in three regions of ORF1a was determined (Figure 9D). NA PRRSV RiboSeq read density upstream of the major (S-2) heteroclite junction is considerably higher than in the downstream regions, with the highest ratio of heteroclite:ORF1a (before FS) translation being reached at 12 hpi, consistent with the increased ratio

of heteroclite:gRNA RNASeq density at this timepoint (Figure 9D). This supports the hypothesis that the N-terminal region of ORF1a can be independently translated from heteroclite sgRNAs (besides from gRNA as part of pp1a/ab) during infection, which could function to increase the ratio of nsp1α and nsp1β compared to the other nsps (Figure 9D). Consistent with the previous analysis (Supplementary Figure 15), the TE of ORF1a decreases over time for both regions upstream of the nsp2 PRF site (Figure 9D). TE in the region after the PRF site does not follow the same trend, likely due to the unexpected increase in RiboSeq read density after the nsp2 frameshift site at 9 and 12 hpi. This is contrary to expectation, as ribosomal frameshifting into nsp2TF should decrease the ribosome density downstream of the nsp2TF stop codon. The reason for this is unclear; perhaps it is a consequence of expressed non-canonical transcripts below the threshold of detection, or biological and/or technical biases. Despite the absence of detectable EU PRRSV heteroclite sgRNAs in the junction-spanning read analysis (Figure 5B), analogous calculations were performed to investigate heteroclite sgRNA and ORF1a expression in EU PRRSV (Figure 9D), revealing RNASeq and RiboSeq outcomes consistent with the presence of translated heteroclite sgRNAs (Figure 9D). These transcripts could potentially be present below the threshold of detection for the junction-spanning read analysis pipeline. Taken together, these results demonstrate that the non-canonical transcripts discovered in this study provide a potential mechanism to temporally regulate the stoichiometry of the polyprotein components, which may reflect changing requirements for the different non-structural proteins throughout infection.

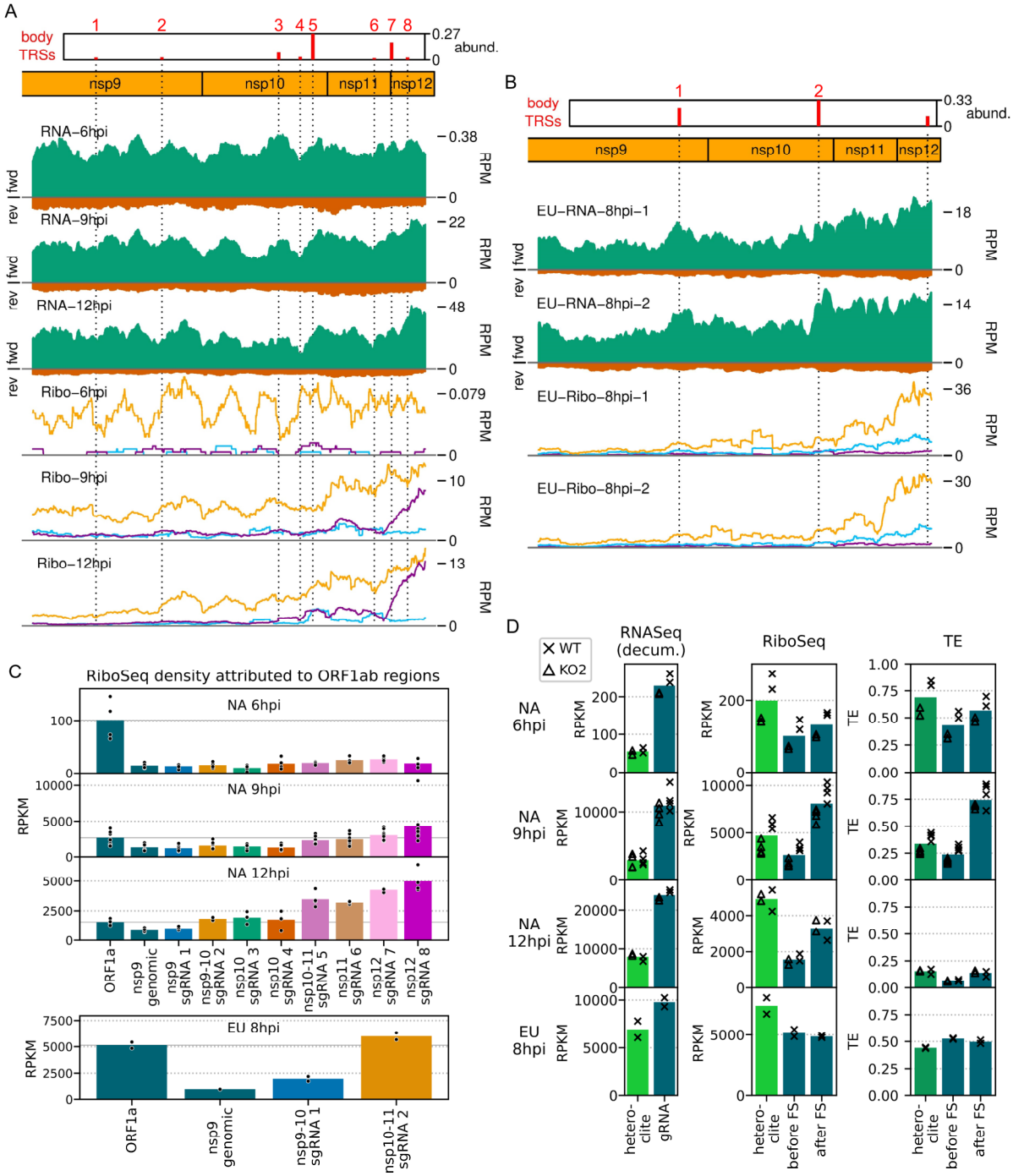


Figure 9. Translation of specific regions of ORF1a and ORF1b.

A) Distribution of RNASeq (upper) and RiboSeq (lower) reads mapping to the ORF1b region of the NA PRRSV genome. Plots constructed as in Figure 2B and Figure 6B, respectively, with the application of a 213-nt running mean filter. Dotted lines indicate body TRS positions, with junction-spanning read abundances supporting body TRSs reproduced from Figure 6B, and the designated ORF1b sgRNA number indicated above in red. For RNASeq, all read lengths were used, and for RiboSeq, read lengths identified as having minimal RNP contamination were used. The libraries displayed are those in Figure 2B and C, with remaining replicates and KO2 libraries in Supplementary Figure 16. **B)** Distribution of RNASeq (upper) and RiboSeq (lower) reads

mapping to the ORF1b region of the EU PRRSV genome. For RNASeq, all read lengths were used, and for RiboSeq, read lengths with good phasing were used. Plot constructed as in panel A, with junction-spanning read abundances supporting body TRSs reproduced from Supplementary Figure 11B. The body TRS annotated at the end of nsp12 does not represent an ORF1b sgRNA, but is expected to produce an alternative transcript for GP2. **C)** RiboSeq read density attributed to different regions of ORF1b. ORF1b was divided into regions based on the positions of the ORF1b sgRNAs, and RiboSeq density of reads in-phase with ORF1b was determined. All sgRNA numbers in the x axis labels refer to ORF1b sgRNAs. RiboSeq density in ORF1a (the "before FS" region from panel D) is included for comparison and its mean value is indicated by a solid grey line. Plot constructed as in Supplementary Figure 13, using a linear scale. **D)** Gene expression in different regions of ORF1a. Transcript abundance for gRNA was calculated by determining RNASeq read density (RPKM) in the region between the major heteroclite (S-2) junction and the nsp2 PRF site. This was subtracted from the density between the beginning of ORF1a and the major heteroclite junction to "decumulate" (decum.) the density in this region and estimate the abundance of the heteroclite transcripts (where all heteroclite sgRNAs contribute read density but the normalisation for length is based only on S-2). RiboSeq read density was calculated in the region between the beginning of ORF1a and the major heteroclite sgRNA junction ("heteroclite"), the region between this junction and the nsp2 PRF site ("before FS"), and the region between the nsp2 and ORF1ab frameshift sites ("after FS"). Although no junctions were detected for putative heteroclite sgRNAs in the EU dataset, regions were designated analogously to NA PRRSV, for comparison. For TE of the "heteroclite" region, the denominator was both gRNA and (decumulated) heteroclite sgRNA combined. Plot constructed as in Supplementary Figure 13, using a linear scale, and with WT and KO2 values indicated by crosses and triangles, respectively.

Investigation of PRF on the viral genome

Another key mechanism by which the stoichiometry of the polyprotein components is controlled is PRF. The ORF1ab frameshift site facilitates a reduction in the ratio of nsp9–12 compared to the upstream proteins[11,12], whereas frameshifting at the nsp2 site produces three variants of nsp2 and causes a proportion of ribosomes to terminate before reaching nsp3[23–25,32]. The occurrence of both frameshift events is evident on the WT NA and EU PRRSV genomes from the changes in phasing after the PRF sites (Figure 10A and B, Supplementary Figures 17-20).

We began by quantifying the efficiency of frameshifting at the nsp2 site. Commonly, from profiling data, frameshift efficiency is calculated using the ratio of the read density upstream of the PRF site compared to downstream, where density is expected to be lower due to termination of either the 0-frame or the transframe ORF[43,49–51]. However, at the NA PRRSV nsp2 PRF site, ribosome drop-off at the end of nsp2N and nsp2TF is not evident (Figure 10A, Supplementary Figure 17), with an increase in RiboSeq read density after the frameshift site, as discussed above (Figure 9D). This increase is not seen

in the counterpart RNASeq libraries (Supplementary Figure 21) and is not related to frameshifting, as it also occurs in the KO2 mutant, in which nsp2 frameshifting is prevented.

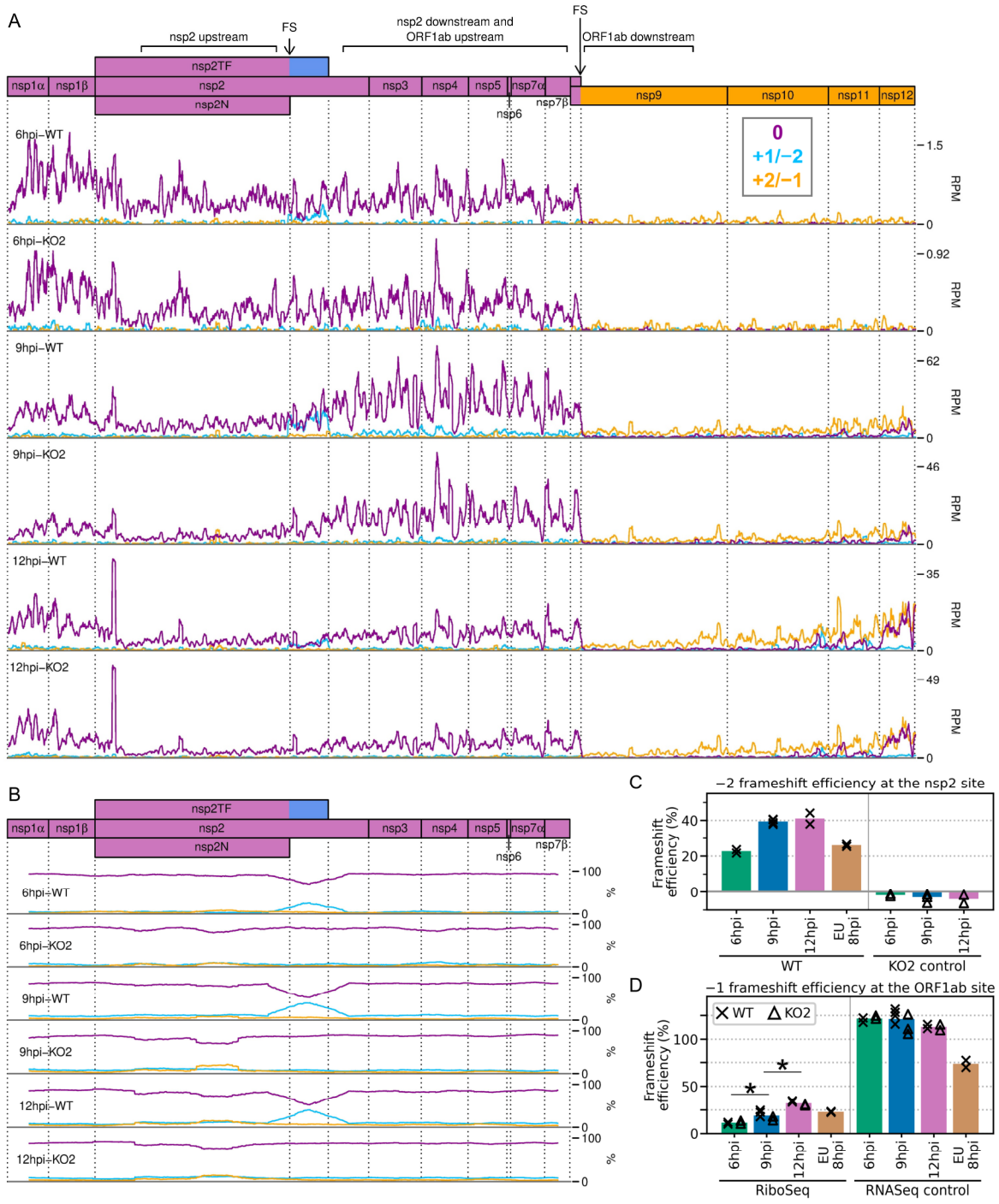


Figure 10. Frameshifting on the PRRSV genome.

A) Distribution of RiboSeq reads in each phase in the ORF1ab region of the NA PRRSV genome. Plot constructed as in Figure 2C. Regions defined as "upstream" and "downstream" in the frameshift efficiency calculations for the nsp2 and ORF1ab sites are annotated above the genome map. Only read lengths identified as having minimal RNP contamination (indicated in

Supplementary Figure 4) were used to generate this plot. Replicates shown are noCHX-Ribo-6hpi-WT-2, noCHX-Ribo-6hpi-KO2-2, noCHX-Ribo-9hpi-WT-4, noCHX-Ribo-9hpi-KO2-3, noCHX-Ribo-12hpi-WT-1 and noCHX-Ribo-12hpi-KO2-1, with remaining replicates in Supplementary Figure 17. The heightened peak shortly after the beginning of nsp2 corresponds to ribosomes with proline codons, which are known to be associated with ribosomal pausing[69–71], in both the P and A sites (P site genomic coordinates 1583–1585). The S-2 heteroclite junction is shortly downstream (genomic coordinate 1747), and is excluded from the nsp2 upstream region. Similarly, the ORF1ab downstream region ends upstream of the body TRS for ORF1b sgRNA 1. **B)** Percentage of RiboSeq reads in each phase across the ORF1a region of the NA PRRSV genome. Reads were separated according to phase, and a 183-codon running mean filter applied to avoid any instances of 0 across ORF1a (excluding the half-window at each end). From this, the percentage of reads in each phase at each codon was calculated. Replicates shown are those from panel A, with remaining replicates in Supplementary Figure 18. Only read lengths with minimal RNP contamination were used to generate this plot. **C)** Bar chart of −2 PRF efficiency at the nsp2 site, calculated based on the differences in phasing in the upstream and transframe regions (data from Supplementary Figure 20). Bars represent the mean results for each group, with individual datapoints overlaid as crosses (WT) and triangles (KO2). The KO2 libraries provide a negative control (expected value ~0%). Only read lengths identified as having minimal RNP contamination (NA PRRSV) or good phasing (EU PRRSV) were used to perform these calculations. **D)** Percentage frameshift efficiency at the ORF1ab site, calculated based on differences in read density upstream and downstream of the frameshift site. Plot constructed as in panel C, with WT and KO2 scatter points offset on the x axis to aid visualisation. The right-hand panel shows the results of applying these calculations to RNASeq reads as a control, for which the expected result is ~100%. Only RiboSeq read lengths with minimal RNP contamination (NA PRRSV) or, for comparability, good phasing (EU PRRSV) were used to perform these calculations. For the RNASeq control all read lengths were used.

Initially, drawing on our previous work on cardioviruses[40,43], we attempted to estimate frameshift efficiencies at the nsp2 site by dividing the RiboSeq profile for the WT virus by that of the KO2 mutant, to factor out differences in translation speed and/or biases introduced during library preparation (see Methods for details). Using the resulting quotient profile, we then compared densities upstream of the nsp2 frameshift site and downstream of the nsp2TF stop codon in order to calculate the combined −2/−1 frameshift efficiency at different timepoints. However, the nsp2-site frameshift efficiencies calculated using this method were quite variable (Supplementary Figure 22A and B). This may be due to the modest level of frameshifting at this site (see below) meaning ribosomal drop-off is low relative to the level of non-frameshift translation, besides the extra complications of (temporally dependent) heteroclite and noncanonical sgRNA production in PRRSV. This is in contrast to cardioviruses, where the frameshift efficiencies reach ~80%[40,43] and there is only a single transcript species (full-length gRNA), or frameshifting at the ORF1ab site, where it is only frameshifted ribosomes rather than non-frameshifted ribosomes that contribute to downstream RiboSeq density.

Therefore, we instead quantified −2 PRF efficiency at the nsp2 site by comparing the proportion of reads in each phase in the upstream and transframe regions (see Methods for details). This led to much greater reproducibility between replicates, and revealed that −2 PRF efficiency significantly increases, from 23% at 6 hpi to 39% at 9 hpi, at which point it reaches a plateau (Figure 10C, Supplementary Figure 22; $p < 0.0005$ based on bootstrap resampling). Although these calculations could be systematically biased if translation were slower in one frame than the other (for example due to sub-optimal codon usage resulting from maintaining two overlapping ORFs), such bias would not be expected to change systematically over the timecourse of infection, and therefore the observed trend should be robust. This is only the second known example of temporally regulated PRF (after cardioviruses[40]), and supports a model of increasing −2 PRF efficiency as nsp1β, the viral protein responsible for stimulating PRF at this site, accumulates and then similarly starts to plateau at 9 hpi (Figure 3D and E). The −2 PRF efficiency on the EU PRRSV genome at 8 hpi was estimated to be 26%, which is similar to the 20% value determined by ^{35}S-Met radiolabelling of MARC-145 cells infected with the EU PRRSV isolate SD01-08 and harvested at 24 hpi (MOI 0.1)[23]. The efficiency of EU PRRSV −2 PRF at 8 hpi (26%) is significantly lower than the NA PRRSV efficiency at 9 hpi (39%; $p < 0.0005$ based on bootstrap resampling). This likely reflects differences between the two viruses as opposed to the difference in timepoints, as EU nsp1β has already accumulated by 8 hpi (Figure 3F), and gene expression analyses suggest the 8 hpi EU PRRSV samples and 9 hpi NA PRRSV samples have progressed to a similar stage of infection (for example, see Figure 2). Although these higher levels (~39%) of −2 PRF have not previously been measured in the context of viral infection, nsp2-site frameshift efficiencies of up to ~50% have been previously recorded in various reporter systems[23,24,37], confirming that this site is capable of facilitating the highly efficient −2 PRF observed here.

Frameshift efficiency at the arterivirus ORF1ab site has not previously been determined in the context of infection, although previous studies using transfected reporter constructs for PRRSV[22] (in yeast) and EAV[21] (in HeLa cells) estimated −1 PRF efficiency as 16% and 15–20%, respectively. We set out to quantify its efficiency in the context of PRRSV infection. Ribosomal drop-off is clearly evident at the

734 ORF1ab −1 PRF site for both NA and EU PRRSV, corresponding to ribosomes which do not frameshift
735 encountering the ORF1a stop codon shortly downstream of the frameshift site (Figure 10A,
736 Supplementary Figure 17). We quantified the ratio of RiboSeq read density in the region downstream
737 of the PRF site compared to that upstream to calculate frameshift efficiency (Figure 10D). PRF
738 efficiency at RNA structure-directed sites is commonly assumed to be fixed; however, surprisingly, −1
739 PRF efficiency at this site also increased over the course of infection, from 11% at 6 hpi to 19% at 9
740 hpi (p value from two-tailed Mann-Whitney U test = 8.5×10^{-3}), and further increased to 32% at 12 hpi
741 ($p = 8.5\times10^{-3}$). The same trend was not observed in the RNASeq libraries (Figure 10D), which were
742 processed as a negative control, indicating it does not result from shared technical biases or an increase
743 in non-canonical transcripts facilitating translation of ORF1b (note that all detected ORF1b sgRNAs
744 are excluded from the regions used). The ORF1ab −1 PRF efficiency on the EU PRRSV genome at 8
745 hpi was 23%, which is similar to the calculated efficiency for NA PRRSV at 9 hpi (Figure 10D). This
746 is consistent with the replicase components being required at similar stoichiometries at this stage of
747 infection for these two viruses.

748 Ribosomal pausing over the slippery sequence is considered to be an important mechanistic feature of
749 PRF[18,19,72,73], although it has been difficult to detect robustly on WT slippery sequences using ribosome
750 profiling[49–51]. To determine whether ribosomal pausing occurs over the nsp2 slippery sequence, we
751 plotted the RPF distribution on the WT genome in this region, and compared this to the KO2 genome
752 to control for shared biases (Supplementary Figure 23A). This revealed a peak on the WT genome,
753 derived predominantly from 21-nt reads and corresponding to ribosomes paused with P site over the
754 slippery sequence (G-G<u>UU-U</u>UU, P site pause location underlined, hyphens delineate 0-frame codons)
755 (Supplementary Figure 23A-C). The peak is not present on the KO2 genome, nor does it overlap the
756 point mutations of KO2, indicating that the differences between the WT and KO2 profiles do not result
757 from technical biases (Supplementary Figure 23D), however its origin is unclear. The frameshift-
758 associated pause is thought to occur at a late stage of the translocation event which begins with the
759 GUU codon in the P site[20,73,74]. The positioning of the short-read peak one nt downstream of this could

be due to an unusual frameshift-intermediate conformation (for example a hyper-rotated state[75,76]) protecting a shorter region of mRNA in the exit tunnel. However, short RPFs are thought to originate from post-translocation ribosomes without an aminoacyl-tRNA in the A site[55,56], suggesting this peak could instead represent ribosomes pausing while decoding the first codon of nsp2TF, which would not be translated on the KO2 genome. This UUG (Leu) codon is normally not expected to be slow to decode, as it is well-adapted to the cellular tRNA pool (Supplementary Figure 23A, heatmap); however, other factors, such as incomplete dissociation of the PCBP/nsp1β complex shortly downstream, could hinder decoding after frameshifting. We found no convincing evidence of ribosomal pausing at the ORF1ab −1 PRF site, and it will be interesting to see whether future ribosome profiling studies which capture the population of short RPFs find similar pauses at other sites.

Host differential gene expression at 12 hpi

Finally, we interrogated our datasets to investigate the host transcriptional and translational response to NA PRRSV infection. Although several analyses of host differential transcription have been performed previously[32,77–83], changes in the host translatome in response to PRRSV infection have not been determined. Here, we characterise infection-induced changes in host transcription and TE by analysing our 12 hpi libraries using xtail[84] and DESeq2[85].

First, we compared the WT libraries against mock (Figure 11A, Supplementary Table 9) and the KO2 libraries against mock (Figure 11B, Supplementary Table 10). Similarly to other studies[79,86,87], we found transcription of genes related to regulation of the cell cycle (amongst other GO terms) to be perturbed by WT PRRSV infection (Supplementary Table 9, sheet: 'GO_TS_up'; GO term GO:0051726 ~26-fold enriched in transcriptionally up-regulated genes). However, a comparison of transcriptional fold changes with those of TE reveals that the majority of transcriptionally up-regulated genes in WT or KO2 compared to mock are down-regulated in terms of TE (Figure 11A and B, top-left quadrants of right-column panels). Such an effect has previously been described as "translational buffering" and is expected to result in little to no change in protein abundance[88], suggesting that many

785 of the observed transcriptional changes make only a minor contribution to the host response to infection.
786 This is consistent with the observation that many of the GO terms enriched amongst the transcriptionally
787 up-regulated genes are also enriched in the translationally down-regulated genes (Supplementary Table
788 9, 'GO_TE_down' sheet). Comparisons between RNA and TE fold changes further reveal many genes
789 with large fold changes of TE and little to no change at the transcriptional level (Figure 11A and B,
790 right-column panels, points in centre-left and centre-right regions), suggesting that translational
791 regulation may be a greater contributor to the host response than transcriptional changes. This is
792 supported by the fact that several GO terms (such as those related to lipid binding and the extra-cellular
793 matrix) are enriched amongst the lists of translationally regulated genes and not in the genes
794 transcriptionally regulated in the opposing direction (Supplementary Tables 9 and 10).

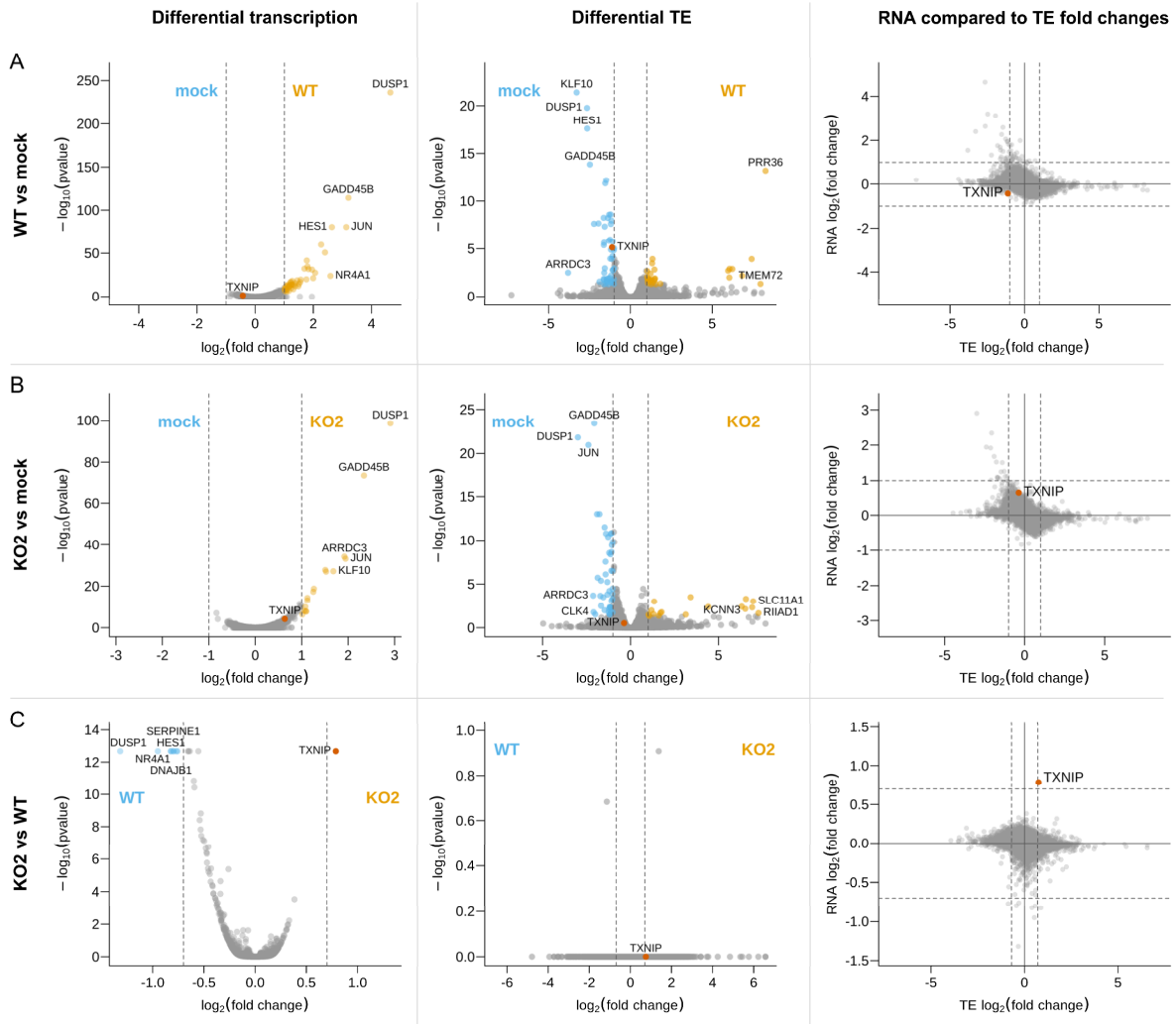


Figure 11. Host differential gene expression at 12 hpi.

Differences in transcription **(left)** and translation efficiency **(centre)** were determined using DESeq2 and xtail, respectively. Volcano plots show relative changes in pair-wise comparisons between the 12 hpi libraries ($n = 2$ biological replicates per condition): WT and mock **(top)**, KO2 and mock **(centre)** or KO2 and WT **(bottom)**. The y axis shows the false discovery rate (FDR)-corrected p values. Genes with FDR-corrected p values ≤ 0.05 and $\log_2$(fold change) magnitudes greater than 1 (WT vs mock and KO2 vs mock) or 0.7 (KO2 vs WT; thresholds in each case indicated by grey dashed lines) were considered differentially expressed and are coloured orange (up-regulated) or blue (down-regulated) in the volcano plots. Where gene names were available, those of the top five significantly up- or down-regulated genes with the greatest fold changes are annotated, and TXNIP is annotated in red on all plots. For all genes where both RNA and TE fold changes were determinable, these were compared **(right)**, irrespective of p value. The full results of these differential expression analyses, including lists of GO terms enriched in each set of differentially expressed genes, are in Supplementary Table 9, Supplementary Table 10 and Supplementary Table 11.

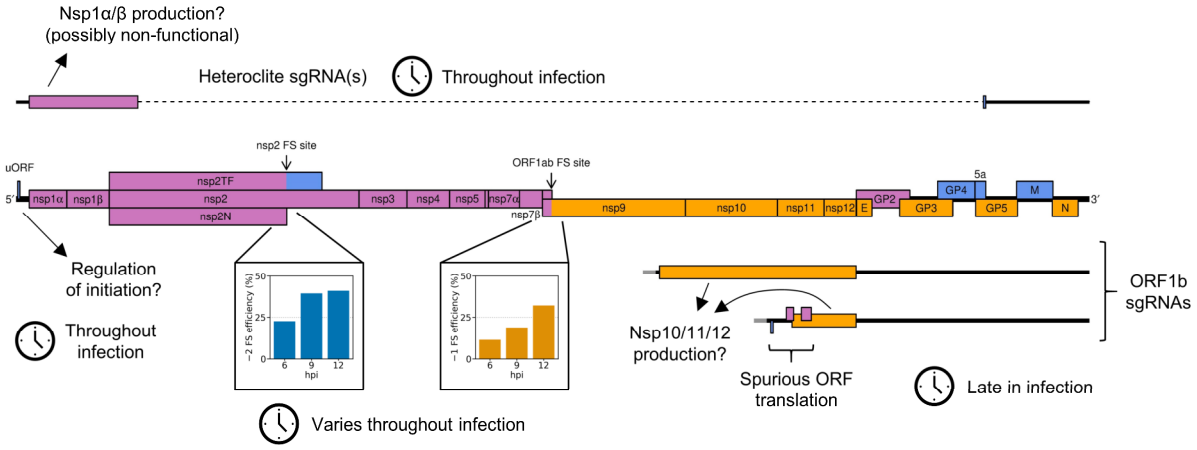
We moved on to compare the host response to infection with WT PRRSV to that of KO2 PRRSV, to investigate the effects of the nsp2 frameshift products, nsp2TF and nsp2N (Figure 11C, Supplementary Table 11). As we expected relatively small differences in the gene expression programmes activated by the two viruses, we lowered the magnitude of the $\log_2$(fold change) required to qualify as a "differentially expressed" gene (DEG) from 1 (in comparisons of infected vs mock) to 0.7 to increase sensitivity. Many of the DEGs in the comparison between WT and KO2 also appear in the list of DEGs in the WT vs mock comparison. These genes have fold changes in the same direction in both the WT vs mock and KO2 vs mock analyses, but those in the KO2 analysis are of a lower magnitude, likely representing the slightly slower replication kinetics of the mutant virus[23,32] as opposed to meaningful differences in the host response. An exception to this is thioredoxin interacting protein (TXNIP, formerly known as vitamin D_3 up-regulated protein 1; Figure 11, red). TXNIP is significantly more highly transcribed in KO2 than WT infection [$\log_2$(fold change) = 0.79, $p = 2.1 \times 10^{-13}$; Figure 11C, left column] and, although it is below our thresholds for qualification as a DEG in these analyses, it is transcriptionally regulated in opposing directions in the WT vs mock [$\log_2$(fold change) = −0.43, p = 0.055] and KO2 vs mock [$\log_2$(fold change) = 0.64, $p = 7.0 \times 10^{-5}$] comparisons (Figure 11A and B, left column), suggesting the difference is not related to replication kinetics. No genes generated significant p values for the KO2 vs WT TE analysis (Figure 11C, middle column), in which the distribution of p values was conservative, likely due to the similarity between the two datasets. Nonetheless, TXNIP clearly stands out in the comparison of RNA and TE fold changes as it is both more highly transcribed and more efficiently translated in KO2 than WT (Figure 11C, right column), further supporting the conclusion that increased TXNIP expression is a notable feature of KO2 infection.

The mechanism by which the presence of nsp2TF/nsp2N could lead to reduced TXNIP expression in WT infection is unclear. The frameshift products share a PLP2 protease domain with the 0-frame product, nsp2, although the DUB/deISGylase activity of this domain is most potent in nsp2N[32]. The frameshift proteins also have different sub-cellular distributions to nsp2[23,36], which may grant them

access to proteins involved in the transcriptional activation of TXNIP, allowing them to interfere with this signalling pathway (for example by de-ubiquitinating its components). While the mechanism remains elusive, there are several reasons why this down-regulation may be beneficial to PRRSV. TXNIP is a key protein in metabolism and redox homeostasis[89,90], regulates cell survival/apoptosis via apoptosis signal regulating kinase 1 (ASK1)[91], and triggers NLRP3 inflammasome activation in monocytes and innate immune cells[92,93]. TXNIP largely exerts its functions by binding and inhibiting thioredoxin[94], an antioxidant which is central to redox signalling and homeostasis[95,96], and reduced cellular levels of TXNIP lead to lower concentrations of reactive oxygen species (ROS)[97,98]. ROS are known to be induced by PRRSV infection[99,100], and lead to apoptosis[100,101], which would likely be detrimental to viral replication. Further, ROS, and by extension thioredoxin and TXNIP, have particular significance for the physiology and function of macrophages[102,103], the primary target for PRRSV *in vivo*[104]. As well as being integral to phagocytosis[102,103], ROS exert a complex, context-dependent effect on macrophage polarisation[105,106], and high levels of thioredoxin have been found to favour polarisation of M2 macrophages[107]. These produce less anti-viral cytokines, and are less suppressive of PRRSV replication, than their M1 counterparts[108]. Therefore, reducing cellular TXNIP concentrations during WT PRRSV infection may be favourable by allowing thioredoxin to function uninhibited and preventing excessive ROS concentrations, with potential implications for macrophage physiology and polarisation. Indeed, it has already been suggested that PRRSV infection induces a skew towards M2 polarisation, although the mechanism was uncharacterised[108].

Discussion

Here, we describe a high resolution analysis of PRRSV replication through ribosome profiling and RNASeq. In addition to confirming and extending the findings of previous transcriptomic analyses, we define the PRRSV translatome and identify strategies of gene expression that may permit the virus to exert translational control during the replication cycle (Figure 12).



862

863 **Figure 12. Schematic summary of non-canonical mechanisms of PRRSV gene expression**
864 **regulation supported by this study.**

865

866 Global analysis of the host response to PRRSV infection at 12 hpi reveals that many of the observed
867 changes in transcript abundance are offset at the translational level, indicating that changes in translation
868 efficiency of host mRNAs may play a dominant role in the response to infection. This phenomenon of
869 host transcriptional responses being counteracted by opposing changes in TE has also been observed in
870 response to SARS-CoV-2 infection, where it was attributed to inhibition of mRNA export from the
871 nucleus, preventing translation[109,110]. This activity is known to be associated with coronavirus nsp1,
872 which inhibits nuclear export by interacting with the nuclear export factor NXF1 in SARS-CoV or the
873 nuclear pore complex component Nup93 in SARS-CoV-2[111,112]. Similarly, PRRSV nsp1β causes
874 imprisonment of host mRNAs in the nucleus, by binding Nup62 to cause the disintegration of the
875 nuclear pore complex[113,114]. This inhibition of host mRNA nuclear export has been found to reduce
876 synthesis of interferon-stimulated genes[114], and mutations in nsp1β which ablate this activity lead to
877 reduced viral load and increased neutralising antibody titres in pigs[115]. These findings suggest that
878 nsp1β-mediated nuclear export inhibition may be responsible for the translational repression seen in our
879 host differential gene expression analyses, which, analogously to conclusions drawn for SARS-CoV-
880 2[109,110], may be a key mechanism by which PRRSV evades the host response to infection.

On the viral genome, numerous translated novel ORFs were discovered, including a short but highly expressed uORF in the NA PRRSV 5′ UTR. The presence of this uORF is very well-conserved amongst NA PRRSV isolates, with an AUG in this position in 558/564 available NA PRRSV sequences, and a non-canonical initiation codon (GUG/AUA/ACG) present in the remainder. In many contexts, uORFs have been shown to regulate translation of the main ORF downstream[116] and, interestingly, uORFs have been found in many nidovirus genomes. For example, recent ribosome profiling studies of the coronaviruses MHV, IBV and SARS-CoV-2 revealed translation of uORFs initiating within the 5′ leader[49–51,117]. Intriguingly, the extent of translation initiation on the SARS-CoV-2 leader uORF differs on different viral transcripts[51,117], suggesting that the same could be true of the NA PRRSV uORF, and presenting a potential mechanism by which the uORF could differentially regulate translation of the canonical ORFs downstream. Looking to the *Arteriviridae* family, EAV has a highly conserved AUG within the leader (present in 94/95 available genome sequences), at genomic coordinates 14–16, predicted to permit expression of a 37-amino-acid leader peptide[118]. *In vitro* translation of the N transcript demonstrated that the uORF is translated, and found it down-regulated translation of the downstream N ORF[119]. Mutations which disrupted this AUG or altered its predicted initiation efficiency were detrimental to viral fitness and led to rapid reversion to the WT sequence[119], although the uORF was not essential for virus viability[119,120]. Upstream ORFs in other arteriviruses have not been characterised, although it was noted that the SHFV 5′ leader contains a putative 13-codon uORF at genomic coordinates 35–73[121], which is conserved in all 37 available full genome sequences. This SHFV uORF is of a similar length and position to the NA PRRSV uORF (ten codons; genomic coordinates 24–36), with the uORFs in these two viruses ending, respectively, 126 nt and 128 nt upstream of the leader TRS. The similarity between these two putative uORFs suggests a possible conserved function, for example conferring resistance to eIF2α-phosphorylation-mediated translation inhibition (as observed for some cellular uORFs[122]), or affecting re-initiation efficiency on the downstream ORF. The latter could help to modulate the ratio of overlapping ORFs (GP2:E and/or GP5:5a) in which the downstream AUG is thought to be accessed by leaky scanning. However, only three out of the 100 EU PRRSV genome sequences with any 5′ UTR have AUGs other than at the

908 extreme 5′ end of the leader, and we do not detect robust uORF translation in the EU PRRSV isolate
909 used in this study, indicating that any putative function of the uORF is not conserved across the entire
910 family.

911 Our analysis of canonical PRRSV ORFs over a 12 hour timecourse revealed that expression of many
912 of these ORFs is controlled by additional mechanisms, at both the transcriptional and translational level,
913 beyond what was previously appreciated. A key observation is that −2 PRF at the NA PRRSV nsp2
914 PRF site is both highly efficient and temporally regulated. At 6 hpi its efficiency is 23%, and this
915 increases to ~40% at 9 and 12 hpi, likely due to accumulation of the frameshift-stimulatory viral protein,
916 nsp1β. Such regulation may be a selective advantage for PRRSV by directing ribosomes to translate
917 proteins which are most beneficial at each stage of infection, optimising the use of cellular resources.
918 At early timepoints, lower nsp2 frameshift efficiency means more ribosomes continue to translate the
919 remainder of pp1a or pp1ab, which encode components of the replication and transcription complex
920 (RTC), which may be more important for establishing infection than translation of the accessory protein
921 nsp2TF. Later in the replication cycle, higher −2 PRF efficiency likely corresponds to an increased
922 requirement for nsp2TF to prevent degradation of GP5 and M, which are expressed from ~8 hpi and are
923 essential for virion assembly[36,123]. Further, nsp2TF is a more potent innate immune suppressor than
924 nsp2, and down-regulates expression of swine leukocyte antigen class I (swine MHC class I)[32,124], which
925 may become critical later in infection as viral proteins and double-stranded RNA accumulate due to
926 viral translation and replication. The only other known example of temporally regulated frameshifting
927 is provided by cardioviruses[40], which encode the only other known protein-stimulated PRF site[38–43].
928 This suggests that temporal regulation may emerge as a common feature of *trans*-activated PRF sites,
929 as more non-canonical PRF sites are discovered in future.

930 RNA structure-directed frameshift sites are commonly assumed to operate at a fixed ratio, due to the
931 lack of *trans*-acting factors involved; however, we found that −1 PRF efficiency at the ORF1ab site
932 increased over time, from 11% to 32%. As opposed to representing specific "regulation" of PRF, we
933 suggest that this is due to changes in gRNA translation conditions as infection progresses. Such changes

could result from activation of pathways that globally regulate translation, such as the unfolded protein response (which is known to be activated by PRRSV infection[125]) or potential phosphorylation of eEF2 (discussed above). Additionally, changes in the localisation or availability of gRNA for translation could result in changing ribosome density as infection progresses, and decreases in ribosome load have been shown to increase −1 PRF efficiency in some studies[126,127]. The mechanism responsible for this effect is not well characterised, although it has been suggested that frameshift-stimulatory RNA structures are more likely to have time to re-fold in between ribosomes if the transcript is more sparsely occupied[126,127]. Consistent with this hypothesis, we find a trend of decreasing gRNA TE over time, although this analysis may be confounded by, for example, the inability to discern translatable gRNA from that undergoing packaging. Whether the observed changes in PRF efficiency represent a selective advantage for PRRSV, or whether they are simply incidental, is unclear. The expected result is an increase in the ratio of ORF1b products to ORF1a products over time. This could be advantageous for PRRSV, for example if there is a greater requirement for nsp2 and nsp3, which promote DMV formation[34], early in infection to establish a protective environment for viral replication, followed by a later preference for producing more of the RdRp (nsp9) and helicase (nsp10) to promote replication itself. Alternatively, this could simply reflect that PRRSV can tolerate a reasonably wide range of ORF1a:ORF1b stoichiometries. Nonetheless, the finding that changes in the translational landscape during infection affect PRF efficiency is relevant to many other RNA viruses, and future studies may reveal temporal changes in PRF efficiency at other frameshift sites, such as those in coronaviruses and HIV. For MHV and SARS-CoV-2, previous results suggest ORF1ab frameshift efficiency may increase over time[49,117,128], but temporal dependence was not assessed in detail, nor statistical significance determined, highlighting this as an interesting area for future coronavirus research.

In addition to changes in the ratio of ORF1b to ORF1a translation, we observed temporal changes in the relative translation of different regions within ORF1b, with increasing translation of the 3′-proximal region as infection progresses. This may result from translation of non-canonical sgRNAs, which we term ORF1b sgRNAs, in which the body TRS is within ORF1b. If the putative translated proteins are

processed to produce functional nsps, this would be expected to increase the stoichiometry of nsps 10–12 compared to nsp9, and alter the relative stoichiometries of nsp10, nsp11 and nsp12. There are several possible reasons this could be beneficial to viral fitness. Although the stoichiometry of the arterivirus RTC is unknown, there is some evidence that the stoichiometry of the coronavirus replication complex varies, containing either one or two copies of the helicase for each copy of the holo-RdRp[129]. This highlights the possibility that the composition of the PRRSV RTC could change over time, for example if extra copies of nsp10, 11 or 12 are supplied from ORF1b sgRNA translation (as well as the potential contribution of increased ORF1ab frameshift efficiency). This provides a potential mechanism of regulating viral replication, for example by altering the ratio of gRNA to sgRNA production (as observed in Figure 2 and Figure 8), as both nsp10 (the helicase) and nsp12 are thought to be involved in promoting sgRNA transcription[129–134]. Nsp11 (NendoU) is an endoribonuclease found in many nidoviruses, which has broad substrate specificity *in vitro*[135], and is also an innate immune antagonist[136–140]. Its expression outside the context of infection is highly toxic[139,141], leading to the suggestion that its restricted perinuclear localisation during infection is important to prevent its expression becoming "suicidal" for the virus[9,137]. Therefore, it may be beneficial to maintain relatively low levels of nsp11 early in infection, and increase production after the optimal microenvironment for its localisation has formed. However, such possibilities are clearly speculative at present.

Interestingly, ORF1b sgRNAs have been found in a number of other nidoviruses. Our results are highly consistent with a previous study on SHFV, in which several ORF1b sgRNAs were detected, which were predicted to produce in-frame portions of the ORF1b polyprotein, or in one case a novel overlapping ORF[45]. Quantitative mass spectrometry provided support for translation of both categories of ORF1b sgRNA, and showed that peptides from nsp11 and nsp12 were 1.2- and 3.1-fold more abundant (respectively) than those from ORF1a-encoded nsp8[45]. ORF1b sgRNAs were also found in lactate dehydrogenase elevating virus (predicted to express the C-terminal 200 amino acids of ORF1b)[142], SARS-CoV-2, HCoV-229E and equine torovirus[46,48,51,59]. Whether this has evolved by virtue of conferring a selective advantage, or whether it is a neutral consequence of the promiscuous

discontinuous transcription mechanism, this suggests that ORF1b sgRNAs are a conserved feature of the nidovirus transcriptome. Further characterisation of these non-canonical transcripts would be highly informative, to determine potential initiation sites and ascertain whether any in-frame products are functional.

Another group of non-canonical transcripts with the potential to modulate polyprotein stoichiometry comprises those termed "heteroclite sgRNAs" by Yuan *et al.*[44,62]. These transcripts result from large internal deletions between regions of sequence similarity in nsp1β/nsp2 and the canonical sgRNA ORFs[44,62], and have been found in several isolates of NA PRRSV[44,62] (including in the present work), with similar types of transcripts found in coronaviruses[46,47,59,143]. Translation of these transcripts is supported by our ribosome profiling results (Figure 9D), *in vitro* experiments using PRRSV reporter constructs[62], and ribosome profiling and mass spectrometry studies of SARS-CoV-2-infected cells[47,51,144–146]. However, it remains to be determined whether the resultant proteins are appropriately cleaved to generate functional nsps. We detected heteroclite sgRNAs as early as 3 hpi, consistent with the finding that they are packaged into PRRSV virions[44,62], and our results suggest they are present throughout infection. If they do produce functional proteins, this could serve as a mechanism to increase the levels of nsp1α and nsp1β, generally considered the most potent innate immune suppressors encoded by PRRSV[113–115,147–151], from early timepoints onwards to evade immune activity.

Conclusion

This work is the first application of ribosome profiling to an arterivirus, and has revealed a complex complement of PRRSV gene expression strategies, several of which permit stoichiometric modulation of the polyprotein proteins. At the level of translational control, the nsp2 −2 PRF site was found to be a rare example of temporally regulated frameshifting, while the finding that −1 PRF efficiency at the ORF1a/1b overlap also increased over time challenges the paradigm that RNA structure-directed frameshift sites operate at a fixed efficiency. At the transcriptional level, numerous non-canonical

transcripts were characterised, some of which bear similarities to those found in other nidoviruses. Among these transcripts, the ORF1b sgRNAs and the heteroclite sgRNAs encode portions of the polyprotein and may provide an additional method of regulating the stoichiometry of its components. Further, some ORF1b sgRNAs likely facilitate the surprisingly high levels of translation we observed for several novel ORFs overlapping ORF1b. Although there is no evidence that these overlapping ORFs produce functional proteins, the lability in the translatome that is afforded by the heterogeneous transcriptome potentially paves the way for similar ORFs to gain functions and become fixed in the viral population. Further expanding the PRRSV translatome, a short but highly translated uORF was discovered in the NA PRRSV 5′ UTR, the presence of which is highly conserved. This presents another opportunity for regulation of viral translation, potentially allowing adaptation in response to infection-induced cellular stress. This is the most comprehensive analysis of PRRSV gene expression to date, and presents new paradigms for understanding arterivirus gene expression and the wider field of programmed ribosomal frameshifting, with potential ramifications for a range of viruses.

Materials and methods

Cells and viruses

For EU PRRSV infections, MA-104 cells were infected with a PRRSV strain derived from the Porcilis® vaccine strain (MSD Animal Health; GenBank accession OK635576). Confluent 6 cm^2 dishes of MA-104 cells were infected at a multiplicity of infection (MOI) within the range 1–3. At 0 (mock), 8, 12, 21 and 25 hpi, cycloheximide (Sigma-Aldrich) was added to the medium (final concentration 100 µg/ml) and incubated for 2 min. Cells were rinsed with 5 ml of ice cold PBS, placed on ice, and 400 µl lysis buffer (20 mM Tris pH 7.5, 150 mM NaCl, 5 mM $MgCl_2$, 1 mM DTT, 1% Triton X-100, 100 µg/ml CHX and 25 U/ml TURBO™ DNase [Life Technologies]) added drop-wise. Cells were scraped off the plate and triturated 10 times with a 26-G needle, before cell debris was removed by centrifugation (13,000 g, 4°C, 20 min) and the supernatant harvested and stored at −70°C.

For NA PRRSV infections, confluent 10 cm^2 dishes of MARC-145 cells were infected with NA PRRSV isolate SD95-21 (GenBank accession KC469618.1), a previously characterised mutant (KO2) based on this background[23,24,32], or mock-infected. At the time of harvesting for the timecourse samples, cells were washed with warm PBS and snap-frozen in liquid nitrogen. For the CHX pre-treated library (CHX-9hpi-WT), an additional CHX pre-treatment step (100 µg/ml, 2 min) was included directly prior to harvesting, and cells were washed with ice cold PBS containing 100 µg/ml CHX before snap-freezing. Snap-frozen dishes were transferred to dry ice and 400 µl lysis buffer added. The dish was transferred to ice to defrost, and cells were scraped and processed as described above.

Western blotting

Samples were resolved by 15% SDS-PAGE and transferred to 0.2 µm nitrocellulose membranes. Membranes were blocked with 5% Marvel milk powder (milk) dissolved in PBS (1 h, 25°C). Primary antibodies were diluted in 5% milk, PBS, 0.1% Tween-20 and incubated with membranes (1 h, 25°C). After three washes in PBS, 0.1% Tween-20, membranes were incubated with IRDye fluorescent antibodies in 5% milk, PBS, 0.1% Tween-20 (1 h, 25°C). Membranes were washed three times in PBS, 0.1% Tween-20 and rinsed in PBS prior to fluorescent imaging with an Odyssey CLx platform (LI-COR). Antibodies used were mouse monoclonal anti-NA-nsp1β (1/1000) and anti-EU-nsp1β (1/500), mouse IgM monoclonal anti-GAPDH (1/20,000, clone G8795, Sigma-Aldrich), goat anti-mouse IRDye 800 CW (1/10,000, LI-COR), and goat anti-mouse IgM (µ chain specific) IRDye 680RD (1/10,000, LI-COR).

Ribosome profiling library preparation

Ribosome profiling and RNASeq libraries were prepared as described in Irigoyen *et al.*[49] using a protocol derived from Refs. [52,65,152,153], with the following modifications. For RiboSeq libraries, RNase I (Ambion) was added to final concentration 2.7, 3, 5, 4.17, 3.33, 2.5, or 2.5 U/µl for 3 hpi, 6 hpi, 9 hpi replicate 1, 9 hpi replicates 2–4, 12 hpi, CHX-9hpi-WT, and EU libraries, respectively, and SUPERase-

In RNase inhibitor (Invitrogen) scaled accordingly (amounts adjusted to improve phasing). The range of fragment sizes selected during the first polyacrylamide gel purification was 25–34 nt (9 hpi replicate 1 RiboSeq and RNASeq libraries, all EU libraries, CHX-9hpi-WT), 19–80 nt (9 hpi replicate 2 RiboSeq libraries), 19–34 nt (all other RiboSeq libraries), or ~50 nt (all other RNASeq libraries), and subsequent gel slices were adjusted accordingly. The greater length range of 9 hpi replicate 2 RiboSeq libraries was selected to investigate the potential presence of long fragments protected by disomes or stable structures/complexes at frameshift sites (analyses not included herein due to inconclusive results). Ribosomal RNA depletion was carried out solely using the RiboZero Gold Human/Mouse/Rat kit (Illumina). Adapter sequences were based on the TruSeq small RNA sequences (Illumina) and, for most libraries, additional randomised bases were added to the end destined for ligation: seven randomised bases on both adapters for all NA PRRSV libraries, no randomised bases on EU PRRSV replicate 1, 14 randomised bases on the 3′ adapter for EU PRRSV replicate 2. Randomised bases allow identification of PCR duplicates and are expected to reduce technical biases. Libraries were deep sequenced as a single-end run on the NextSeq 500 platform (Illumina), or a paired-end run using a Mid Output v2 kit (150 cycles: 2×75) for 9 hpi replicate 2 RiboSeq libraries. Data are available on ArrayExpress under accession numbers E-MTAB-10621, E-MTAB-10622 and E-MTAB-10623.

Core analysis pipeline

For single-end libraries, fastx_clipper (FASTX Toolkit version 0.0.14, parameters: -Q33 -l 33 -c -n -v) was used to trim the universal adapter sequence from reads and to discard adapter-only reads, non-clipped reads, and “too-short” reads (inferred original fragment lengths shorter than the minimum intended length experimentally purified – see library preparation description for lengths). Adapter dimers were counted using the grep command line utility and added to the adapter-only read count. For paired-end libraries, adapter trimming, read pair merging and removal of adapter-only reads was carried out using LeeHom[154] (v.1.1.5) with the --ancient-dna option specified (as the expected fragment lengths of such DNA are in the same range as ours). Pairs of reads which LeeHom was unable to merge were

put in the "non-clipped" category for the purposes of library composition analysis, and "too-short" reads were removed using awk. For libraries prepared using adapters with randomised bases, PCR duplicates were removed using awk, and seqtk (version 1.3) was used to trim the randomised bases from the reads. Bowtie1 (version 1.2.3)[155] was used to map reads to host and viral genomes using parameters "-v n_mismatches --best", where n_mismatches was one for RiboSeq and two for RNASeq libraries. Reads were mapped to each of the following databases in order and only reads that failed to align were mapped to the next database: ribosomal RNA (rRNA), virus RNA (vRNA), mRNA, non-coding RNA (ncRNA), genomic DNA (gDNA). Viral genome sequences were verified by *de novo* genome assembly using Trinity (version 2.9.1). The rRNA database comprised the following GenBank accessions: NR_003287.2, NR_023379.1, NR_003285.2, NR_003286.2, AY603036.1, AF420058.1, AF420040.1, AY633510.1, AF352382.1, L35185.1, DQ983926.1, KJ193255.1, M30951.1, M30950.1, M30952.1, KJ193272.1, KJ193259.1, KJ193258.1, KJ193256.1, KJ193255.1, KJ193045.1, KJ193042.1, KJ193044.1, KJ193041.1, KJ193019.1, KJ193018.1, KJ193017.1 and AF420040.1. The mRNA database was compiled from the available *Chlorocebus sabaeus* RefSeq mRNAs after removing transcripts with annotated changes in frame. The ncRNA database was Chlorocebus_sabaeus.ChlSab1.1.ncrna.fa and the gDNA database was the ChlSab1.1 genome assembly, both from Ensembl. The position of vRNA within the database mapping order was altered to confirm that significant numbers of viral reads were not erroneously mapping to host databases or *vice versa*. All analyses were carried out using reads mapped by bowtie as described above, except for running PRICE, analyses using junction-spanning reads, or host differential gene expression.

Quality control plots and analyses were performed as described in Irigoyen *et al.*[49], modified for the timecourse libraries to account for the longer RNASeq reads, so that a 3′ UTR of at least 90 nt (as opposed to 60 nt) was required for inclusion of transcripts in the metagene analysis of 5′ end positions relative to start and stop codons (Supplementary Figure 1). For quality control analyses of read length and phasing, reads mapping to ORF1ab (excluding nsp2TF, all phases) were used for the virus versus host analyses (e.g. Figure 1C, Supplementary Figure 2), while for analyses of specific regions (e.g.

Supplementary Figures 3 and 4), overlapping regions of ORFs were permitted for the length distribution but not phasing analyses. For these phasing plots, phase 0 was designated independently for each region, relative to the first nucleotide of the ORF in that region. For the negative-sense read analysis, reads mapping to anywhere on the viral genome were used, and phase was determined using the 5′ end of the read (the 5′ end of the reverse complement reported by bowtie plus the read length). The coordinates of the regions of the viral genome used for all analyses are given in Supplementary Table 1.

For plots of read distributions on viral genomes, read densities were plotted at the inferred ribosomal P site position, obtained by applying a static +12 nt offset to the 5′ end coordinate of the read (applied to both RiboSeq and RNASeq for comparability although ribosomal P site is not relevant to RNASeq reads). For Supplementary Figure 23, species-specific tRNA adaptation index (stAI) values for *Macaca mulatta* (in the absence of *C. sabaeus* data) were obtained from STADIUM[156] on 4th Oct 2020, and a heatmap constructed using the minimum and maximum values for *M. mulatta* codons as the most extreme colours available in the gradient. For plots of the percentage of reads in each phase across ORF1a, positive-sense RiboSeq reads were separated according to phase and a 183-codon running mean filter applied to avoid any instances of zero across ORF1a (excluding the halfwindow at each end) after sliding window application. From this, the percentage of reads in each phase at each codon in-frame with ORF1a was calculated. All plots and analyses use RiboSeq read lengths identified as having minimal RNP contamination (Supplementary Figures 4 and 12F) unless otherwise specified.

Significance testing for proportion of host RPFs which are short

RiboSeq libraries were grouped into early (3 and 6 hpi) and late (9 and 12 hpi) timepoints, to provide enough replicates in each group to perform a two-tailed *t* test (WT and KO2 were treated as equivalent). Positive-sense RPFs mapping to host mRNA were used, and short reads were defined as 19–24 nt long, with the denominator formed by 19–34 nt long reads. The early timepoint group was used as a control, for which there was no significant difference in the percentage of short RPFs in infected cells compared to mock-infected cells ($p = 0.52$).

Junction-spanning read analysis for novel transcript discovery

Reads which did not map to any of the host or viral databases (rRNA, vRNA, mRNA, ncRNA or gDNA) in the core pipeline (described above) were used as input for mapping using STAR[60], version 2.7.3a. Mapping parameters were selected based on those suggested in Kim *et al.*[46] to switch off penalties for non-canonical splice junctions:

--runMode alignReads --outSAMtype BAM SortedByCoordinate --outFilterType BySJout --outFilterMultimapNmax 2 --alignSJoverhangMin 12 --outSJfilterOverhangMin 12 12 12 12 --outSJfilterCountUniqueMin 1 1 1 1 --outSJfilterCountTotalMin 1 1 1 1 --outSJfilterDistToOtherSJmin 0 0 0 0 --outFilterMismatchNmax 2 --scoreGapNoncan -4 --scoreGapATAC -4 --chimOutType Junctions --chimScoreJunctionNonGTAG 0 --alignEndsType EndToEnd --alignSJstitchMismatchNmax -1 -1 -1 -1 --alignIntronMin 20 --alignIntronMax 1000000 --outSAMattributes NH HI AS nM jM jI

First, junctions were processed within each library. To avoid junction clusters becoming inflated, late timepoint libraries (8 hpi onwards) were filtered to remove junctions with fewer than four supporting reads. Reads were split into two categories, TRS-spanning and non-TRS-spanning, according to whether the donor site of the junction overlapped the leader TRS (genomic coordinates in Supplementary Table 1). Junctions were clustered so that all junctions within a cluster had acceptor coordinates within seven (for TRS-spanning junctions) or two (for non-TRS-spanning junctions) nucleotides of at least one other junction in the cluster, with the same requirement applied to donor coordinates. This was to group highly similar junctions together and account for the fact that the precise location of a junction is ambiguous in cases where there is similarity between the donor and acceptor sites (such as between the 6-nt leader and body TRSs). The junctions within each cluster were merged, with donor and acceptor sites defined as the midpoints of the ranges of coordinates observed in the cluster, and the number of reads supporting the merged junction defined as the sum of the supporting read counts for all the input junctions in the cluster.

Then, junctions were filtered to keep only those present in multiple libraries and merged to generate one dataset per timepoint. Merged junctions from individual libraries were filtered so that only junctions which were present in more than one replicate (considering WT and KO2 as one group) passed the filter. Junctions were defined as matching if the ranges of the donor and acceptor coordinates for the junction in one library overlap with those of a junction in a second library. Matching junctions from all replicates were merged as described above to make the final merged junction. For the NA PRRSV M junction there is a stretch of six bases that is identical upstream of the leader TRS and the body TRS, leading to separation of the two alternative junction position assignments by a distance greater than the seven bases required to combine TRS-spanning junctions into clusters. The two junction clusters that are assigned either side of this stretch of identical bases were specifically selected and merged at this stage. To ensure this merging strategy did not lead to clusters spanning overly wide regions, widths of merged junction donor and acceptor sites were assessed, and the mean and median junction width for all analyses was found to be < 3 nt (maximum width 17 nt, for the NA PRRSV M junction). TRS-spanning junctions were designated as "known" junctions if they were the major junction responsible for one of the known canonical sgRNAs of PRRSV. Non-TRS-spanning junctions were filtered according to whether they represent local (≤ 2000 nt) or distant (> 2000 nt) deletions.

The proportion of junction-spanning reads at donor and acceptor sites was calculated as junction-spanning / (junction-spanning + continuously aligned to reference genome). The number of non-junction-spanning reads at the junction site was defined as the number of bowtie-aligned reads (from the core pipeline) spanning at least the region 12 nt either side of the midpoint position of the donor or acceptor site (note that for sgRNA acceptor sites the denominator will include not only gRNA but also sgRNAs with body TRSs upstream). For all TRS-spanning junctions the donor midpoint was set according to the known leader TRS sequence (genomic coordinate 188 for NA PRRSV and 219 for EU PRRSV).

Detection of novel viral ORFs

Novel ORF discovery was performed using PRICE[64] (version 1.0.3b). A custom gtf file was made for each virus, with only the gRNA transcript and ORF1b CDS annotated. Other known viral ORFs were not annotated and served as positive controls. The custom viral gtf files were each individually concatenated with the host gtf file (ChlSab1.1.101, downloaded from Ensembl) to make the input of known ORFs (treating the viral genome as an additional host chromosome). The reference fasta files for the host and viruses were similarly concatenated to make the input reference sequences. RiboSeq reads were mapped to these combined references using STAR[60], version 2.7.3a, with parameters as described for novel transcript discovery but with the following changes: --outFilterType Normal --outFilterMultimapNmax 10 --outFilterMismatchNmax 1 --chimOutType WithinBAM --outSAMattributes MD NH HI AS nM jM jI

The STAR alignments were used as input for PRICE, and *p* values were corrected for multiple testing using the Benjamini-Hochberg method before filtering results to select significant viral ORFs. For noCHX NA PRRSV libraries, read lengths with minimal RNP contamination were used, while for EU PRRSV libraries all read lengths were used. The CHX NA PRRSV library was not used for PRICE, as CHX pre-treatment can artefactually increase uORF translation[157].

Analysis of sequence conservation

Sequences were selected for inclusion in the alignment based on a requirement for ≥ 50% amino acid identity (across all ORFs excluding overlapping regions), ≥ 70% nucleotide identity and ≥ 95% coverage compared with the prototype NA (NC_001961) or EU (NC_043487) PRRSV reference genomes, resulting in 661 and 120 sequences, respectively. For analyses of NA PRRSV genomes "representative of NA PRRSV diversity", the NA PRRSV sequences were clustered using CD-HIT[158] (version 4.8.1) based on the whole genome and with a nucleotide similarity threshold of 95% (all other parameters set to default), and one representative sequence from each cluster was selected to make a

sequence alignment of 137 sequences. Logo plots and mini-alignment plots were generated using CIAlign[68] and, for the uORF analyses, genome sequences which began partway through the ORF were excluded, as was KT257963 which has a likely sequencing artefact in the 5′ UTR. Synonymous site conservation was analysed, for the representative NA PRRSV sequences or for all EU sequences, using SYNPLOT2[67] and *p* values plotted after application of a 25-codon running mean filter.

Transcript abundance, total translation, and translation efficiency analyses

For the main analysis, RiboSeq RPKM values were calculated using the read counts and ORF "Location"s from the PRICE output (Supplementary Table 8), using the same library size normalisation factors as the core pipeline (where positive-sense virus- and host-mRNA-mapping reads from the bowtie output are the denominator). Each ORF was paired with the transcript(s) most likely to facilitate its expression (see schematic in Figure 7 and junction coordinates in Supplementary Tables 1 and 5–7). For some ORFs (NA PRRSV nsp10-iORF2, nsp11-iORF3, GP3-iORF and GP4-iORF), this included transcripts which are expected to produce slightly N-terminally truncated ORFs compared to the PRICE designation. ORFs overlapping ORF1ab for which there were no novel transcripts expected to facilitate expression were paired with gRNA. All ORF1b sgRNAs, defined as sgRNAs with body TRSs within ORF1b and ≥ 50 or ≥ 10 junction-spanning reads at 12 hpi (NA PRRSV) or 8 hpi (EU PRRSV), were included in the transcript abundance analysis regardless of whether they are expected to result in expression of a novel ORF.

To estimate transcript abundance, reads aligned to the viral genome by STAR (see junction-spanning read analysis pipeline) were normalised by library size using the same library size normalisation factors as the core pipeline. In cases where multiple body TRSs are expected to give rise to two different forms of a transcript that express the same ORF(s), these were treated as a single transcript for the purposes of this analysis, and read counts for all junctions were combined. Abundance of the gRNA transcript was defined as the number of reads which span 12 nt either side of the midpoint of the leader TRS (genomic coordinate 188 for NA PRRSV and 291 for EU PRRSV). This is analogous to the 12-nt

overhang required either side of a junction to qualify for mapping by STAR; however, these reads are not junction-spanning, and map specifically to gRNA (and a small proportion of non-canonical transcripts such as heteroclite sgRNAs). Leader abundance was defined as the total number of reads for all other transcripts in the analysis combined, as the leader is present on all sgRNA species and the gRNA. TE was calculated by dividing RiboSeq RPKM values by RNASeq junction-spanning read RPM values, excluding conditions where the denominator was zero.

For plots with logarithmic axes, data points with a value of zero were excluded from the plot, but not from mean calculations. WT and KO2 were treated as equivalent unless specified. For libraries with shorter read lengths (EU libraries and NA 9 hpi replicate 1 libraries) junction-spanning read counts are lower (and also subject to greater inaccuracies as a result of less dilution of possible read start- and end-point specific ligation biases) due to the requirement for a 12-nt overhang either side of the junction effectively representing a much larger proportion of the total read length. As such, these libraries are not directly comparable to the remaining NA PRRSV libraries and they were plotted separately and not included in NA PRRSV mean calculations.

For the estimation of translation of different regions of ORF1b, sections were designated as the regions between the downstream-most body TRS of one ORF1b sgRNA and the upstream-most body TRS of the next (all region coordinates given in Supplementary Table 1). Bowtie-aligned RiboSeq reads (from the core pipeline) which mapped in-phase with ORF1b in the designated regions were counted, using only read lengths with minimal RNP contamination (NA PRRSV) or good phasing (EU PRRSV). Total read counts were normalised by library size and region length to give RPKM. The same process was applied to the region of ORF1a between the major heteroclite junction and the nsp2 PRF site for comparison, counting reads mapping in-phase with ORF1a.

For investigation of gene expression in different regions of ORF1a, transcript abundance for the heteroclite sgRNAs was calculated by subtracting the gRNA RNASeq RPKM (measured in the region between the major [S-2] heteroclite sgRNA junction and the first ORF1b sgRNA body TRS) from the RPKM in the region between the leader TRS and the major heteroclite sgRNA junction (“heteroclite”).

This provides an averaged result for all heteroclite sgRNAs, although it does not take into account the reduced transcript length for the minor heteroclite sgRNAs compared to S-2. RiboSeq read density for the different regions of ORF1a (Supplementary Table 1) was calculated as described above for sections of ORF1b. For TE of the heteroclite region, the denominator was both gRNA and (decumulated) heteroclite sgRNAs combined.

Nsp2 site PRF efficiency calculations based on phasing

The proportion of reads in each phase in the upstream and transframe regions (coordinates in Supplementary Table 1) was calculated, where in both regions phase is taken relative to the ORF1a reading frame. It was assumed that all ribosomes in the upstream region were translating in the 0 frame, and the phase distribution in this region was used to estimate what proportion of 0-frame ribosomes generate reads attributed to the 0 phase ($upstream_0$) and the −2 phase ($upstream_{-2}$). This was extrapolated to determine what proportion of reads are expected to be in the −2 phase in the transframe region ($transframe_{-2}$) in the absence of frameshifting (which is expected to be the same as in the upstream region). A proportion (FS_proportion) of ribosomes undergoing −2 PRF is expected to mean that, between the upstream and downstream region, FS_proportion of phase 0 reads change from the 0 to the −2 phase and FS_proportion of −2 phase reads move to the −1 phase (leaving 1 – FS_proportion in the −2 phase). These concepts were combined to make the equation:

$$transframe_{-2} = (FS\_proportion \times upstream_0) + (1 - FS\_proportion) \times upstream_{-2}$$

This was rearranged to calculate percentage frameshift efficiency (which is FS_proportion expressed as a percentage):

$$FS\_efficiency = 100 \times (transframe_{-2} - upstream_{-2})/(upstream_0 - upstream_{-2})$$

This phasing-based method of calculating frameshift efficiency should theoretically be unaffected by RNP contamination, provided the RNP footprints are equally distributed between the three phases. Let R be the proportion of total reads that are RNPs, and let P_0 and P_{-2} be the proportion of total RNPs that

are attributed to the 0 and −2 phases, respectively. The phasing of reads originating from RNPs is not expected to change due to frameshifting. Therefore, the equation for calculating the fraction of reads that change from the 0 to −2 phase becomes:

$$FS\_proportion \times (upstream_0 - RP_0)$$

and the equation for calculating the fraction of reads that remain in the −2 phase becomes:

$$RP_{-2} + [(1 - FS\_proportion) \times (upstream_{-2} - RP_{-2})]$$

Combining these makes the equation:

$$transframe_{-2} = FS\_proportion \times [upstream_0 - RP_0] + RP_{-2} + [(1 - FS\_proportion) \times (upstream_{-2} - RP_{-2})]$$

This rearranges to:

$$FS\_proportion = (upstream_{-2} - transframe_{-2})/(RP_0 - upstream_0 - RP_{-2} + upstream_{-2})$$

If $P_0 = P_{-2}$ (for example if RNPs are equally distributed between all three phases) then this causes both terms involving R to cancel out of the equation, meaning RNPs would not affect the result. The same would hold for any other form of uniform non-phased contamination.

For NA PRRSV libraries, read lengths identified as having minimal RNP contamination (indicated in Supplementary Figure 4) were used and, for EU PRRSV, read lengths with good phasing (Supplementary Figure 6D) were used.

Nsp2 site PRF efficiency calculations based on KO2-normalised read density

The density normalisation-based method calculates frameshift efficiency as 100 × [1–(downstream/upstream)], where downstream and upstream represent the RPKM values for the respective regions after normalisation of WT density by the density in its KO2 counterpart. WT and KO2 libraries were paired first according to processing batches, and within each batch (if there were multiple replicates) the WT library with the higher ratio of virus:host RiboSeq reads was paired with

the KO2 library with the higher ratio. This resulted in libraries with the same replicate number being paired, except 9 hpi WT-3 was paired with KO2-4, and WT-4 with KO2-3. For RiboSeq libraries, read lengths identified as having minimal RNP contamination (indicated in Supplementary Figure 4) were used, whereas for RNASeq negative control libraries all read lengths were used. This method could not be applied to EU PRRSV libraries as no KO2 libraries were made.

ORF1ab site PRF efficiency calculations

Frameshift efficiency was calculated as 100 × (downstream/upstream), where downstream and upstream represent the RPKM values for the respective regions. Mutation of this frameshift site prevents viral replication so normalisation by a frameshift-defective mutant was not possible. KO2 and WT libraries were treated as equivalent for the calculations and statistical tests. For NA PRRSV RiboSeq libraries, read lengths identified as having minimal RNP contamination (indicated in Supplementary Figure 4) were used, and for EU PRRSV RiboSeq libraries read lengths with good phasing (Supplementary Figure 6D) were used. For all RNASeq negative control libraries, all read lengths were used. A two-tailed Mann-Whitney U test was employed to assess statistical significance of differences between groups of observed values.

Bootstrap resampling for phasing-based nsp2 −2 PRF frameshift efficiency calculations

100,000 randomised resamplings of codons in each respective region were performed. Each WT library was paired with its corresponding KO2 library (as described above), with matched codons selected for the two libraries in each resampling, and reads with 5′ ends mapping to these codons were used as input for the PRF efficiency calculations. Calculation of nsp2 −2 PRF efficiency for each resampling was performed using the phasing-based method as described above, with the results of individual libraries recorded, and then the mean of all libraries in each group calculated. Regions and bounding coordinates used were as described above, with an additional 147-codon downstream region (the same length as the region of nsp2TF used) added as a negative control (coordinates in Supplementary Table 1). For all

resamplings, *n* codons were sampled with replacement, where *n* is the total number of codons in the region undergoing resampling. Bootstrap resamplings were used to empirically determine *p* values. Confidence intervals for each bootstrap distribution were calculated using the bias-corrected accelerated (BCa) method, implemented through the R package coxed (version 0.3.3). This was performed for 95%, 99.5% and 999.5% confidence intervals, and pairs of groups were considered as significantly different with $p < 0.05$, 0.005 or 0.005, respectively, if the mean of the "group one" bootstrap distribution was not within the confidence intervals of "group two" and *vice versa*.

Host differential gene expression

After basic processing and removal of rRNA- and vRNA-mapping reads using bowtie as described in the core analysis pipeline, remaining reads were aligned to the host genome (fasta and gtf from genome assembly ChlSab1.1) using STAR[60] (version 2.7.3a) with the following parameters: --runMode alignReads --outSAMtype BAM SortedByCoordinate --outFilterMismatchNmax n_mismatches --outFilterIntronMotifs RemoveNoncanonicalUnannotated --outMultimapperOrder Random (where n_mismatches was one for RiboSeq libraries and two for RNASeq libraries). Reads were tabulated using htseq-count[159] (version 0.13.5), with parameters -a 0 -m union -s yes -t gene (covering the whole gene) for the differential transcription and -a 0 -m intersection-strict -s yes -t CDS (covering only the CDS) for the differential TE. Genes with fewer than ten reads between all libraries in the analysis combined were excluded, and quality control was performed according to the recommendations in the DESeq2[85] user guide, with all replicates deemed to be of sufficient quality. Read counts were normalised for differences in library size using DESeq2 (version 1.30.1), providing the input for differential transcription using DESeq2 (default parameters) or differential TE using xtail[84] (version 1.1.5; parameters: normalize = FALSE). Shrinkage was applied to the DESeq2 output using lfcShrink (parameters: type = “normal”). Where necessary (i.e. for the KO2 vs WT comparison), fdrtool was used to correct conservative *p* values for differential transcription (version 1.2.16; parameters: statistic = "normal"), in addition to the Benjamini-Hochberg correction for multiple testing. This correction could

not be applied to the xtail results as the test statistic is not included in the xtail output. Genes were considered significantly differentially expressed if they had FDR-corrected p value ≤ 0.05 and $\log_2$(fold change) of magnitude > 1 (for comparisons to mock) or > 0.7 (for KO2 vs WT comparison). GO terms associated with lists of significantly differentially expressed genes were retrieved and tested, using DAVID[160] (version 6.8, functional annotation chart report, default parameters), for enrichment against a background of GO terms associated with all genes that passed the threshold for inclusion in that differential expression analysis.

Acknowledgements

This work was supported by the Wellcome Trust (U.K.) through PhD studentships to G.M.C. (203864/Z/16/Z) and L.S. (102163/Z/13/Z), Senior Research Fellowships to A.E.F. (106207/Z/14/Z, 220814/Z/20/Z) and an Investigator Award (202797/Z/16/Z) to I.B. Additional support was from a European Research Council grant (646891; to A.E.F.) and an Agriculture and Food Research Initiative Competitive Grant (2015-67015-22969; to Y.F.) from the USDA National Institute of Food and Agriculture. We would like to thank Dr David Brown for supporting discussions.

Author contributions

G.M.C. and L.S. performed the ribosome profiling and biochemical experiments. G.M.C., P.S., Y.L., L.S., A.M., Y.F. and I.B. carried out virus infections. Bioinformatic analysis was carried out by G.M.C., K.B., L.S., A.M.D., C.T. and A.E.F. G.M.C, A.E.F. and I.B. wrote the manuscript with contributions from all authors.

Competing interests

The authors declare no competing interests.

References

1. Meulenberg, J. J. *et al.* Lelystad virus belongs to a new virus family, comprising lactate dehydrogenase-elevating virus, equine arteritis virus, and simian hemorrhagic fever virus. *Arch. Virol. Suppl.* **9**, 441–448 (1994).

2. Cavanagh, D. Nidovirales: a new order comprising Coronaviridae and Arteriviridae. *Arch. Virol.* **142**, 629–633 (1997).

3. Nan, Y. *et al.* Improved vaccine against PRRSV: Current Progress and future perspective. *Frontiers in Microbiology* **8**, 1635 (2017).

4. Holtkamp, D. J. *et al.* Assessment of the economic impact of porcine reproductive and respiratory syndrome virus on United States pork producers. *Journal of Swine Health and Production* **21**, 72–84 (2013).

5. Kappes, M. A. & Faaberg, K. S. PRRSV structure, replication and recombination: Origin of phenotype and genotype diversity. *Virology* **479–480**, 475–486 (2015).

6. Collins, J. E. *et al.* Isolation of swine infertility and respiratory syndrome virus (isolate ATCC VR-2332) in North America and experimental reproduction of the disease in gnotobiotic pigs. *J. Vet. Diagnostic Investig.* **4**, 117–126 (1992).

7. Wensvoort, G. *et al.* Mystery swine disease in The Netherlands: the isolation of Lelystad virus. *Vet. Q.* **13**, 121–130 (1991).

8. Guo, Z., Chen, X. X., Li, R., Qiao, S. & Zhang, G. The prevalent status and genetic diversity of porcine reproductive and respiratory syndrome virus in China: A molecular epidemiological

perspective. *Virology Journal* **15**, (2018).

9. Snijder, E. J., Kikkert, M. & Fang, Y. Arterivirus molecular biology and pathogenesis. *Journal of General Virology* **94**, 2141–2163 (2013).

10. Posthuma, C. C., te Velthuis, A. J. W. & Snijder, E. J. Nidovirus RNA polymerases: Complex enzymes handling exceptional RNA genomes. *Virus Research* **234**, 58–73 (2017).

11. Meulenberg, J. J. M. *et al.* Lelystad virus, the causative agent of porcine epidemic abortion and respiratory syndrome (PEARS), is related to LDV and EAV. *Virology* **192**, 62–72 (1993).

12. Nelsen, C. J., Murtaugh, M. P. & Faaberg, K. S. Porcine Reproductive and Respiratory Syndrome Virus Comparison: Divergent Evolution on Two Continents. *J. Virol.* **73**, 270–280 (1999).

13. Snijder, E. J., Wassenaar, A. L. & Spaan, W. J. Proteolytic processing of the replicase ORF1a protein of equine arteritis virus. *J. Virol.* **68**, 5755–64 (1994).

14. Gorbalenya, A. E., Enjuanes, L., Ziebuhr, J. & Snijder, E. J. Nidovirales: Evolving the largest RNA virus genome. *Virus Res.* **117**, 17–37 (2006).

15. Rodnina, M. V *et al.* Survey and summary: Translational recoding: Canonical translation mechanisms reinterpreted. *Nucleic Acids Res.* **48**, 1056–1067 (2020).

16. Atkins, J. F., Loughran, G., Bhatt, P. R., Firth, A. E. & Baranov, P. V. Ribosomal frameshifting and transcriptional slippage: From genetic steganography and cryptography to adventitious use. *Nucleic Acids Res.* **44**, 7007–78 (2016).

17. Firth, A. E. & Brierley, I. Non-canonical translation in RNA viruses. *Journal of General Virology* **93**, 1385–1409 (2012).

18. Plant, E. P. & Dinman, J. D. Torsional restraint: A new twist on frameshifting pseudoknots. *Nucleic Acids Res.* **33**, 1825–1833 (2005).

19. Namy, O., Moran, S. J., Stuart, D. I., Gilbert, R. J. C. & Brierley, I. A mechanical explanation

of RNA pseudoknot function in programmed ribosomal frameshifting. *Nature* **441**, 244–247 (2006).

20. Caliskan, N., Katunin, V. I., Belardinelli, R., Peske, F. & Rodnina, M. V. Programmed –1 Frameshifting by Kinetic Partitioning during Impeded Translocation. *Cell* **157**, 1619–1631 (2014).

21. den Boon, J. A. *et al.* Equine arteritis virus is not a togavirus but belongs to the coronaviruslike superfamily. *J. Virol.* **65**, 2910–20 (1991).

22. Bekaert, M. & Rousset, J. P. An extended signal involved in eukaryotic -1 Frameshifting operates through modification of the E site tRNA. *Mol. Cell* **17**, 61–68 (2005).

23. Fang, Y. *et al.* Efficient -2 frameshifting by mammalian ribosomes to synthesize an additional arterivirus protein. *Proc. Natl. Acad. Sci.* **109**, E2920–E2928 (2012).

24. Li, Y. *et al.* Transactivation of programmed ribosomal frameshifting by a viral protein. *Proc. Natl. Acad. Sci. U. S. A.* **111**, E2172-81 (2014).

25. Napthine, S. *et al.* A novel role for poly(C) binding proteins in programmed ribosomal frameshifting. *Nucleic Acids Res.* **44**, 5491–5503 (2016).

26. Li, Y. *et al.* Programmed –2/–1 Ribosomal Frameshifting in Simarteriviruses: an Evolutionarily Conserved Mechanism. *J. Virol.* **93**, (2019).

27. Han, J., Rutherford, M. S. & Faaberg, K. S. The Porcine Reproductive and Respiratory Syndrome Virus nsp2 Cysteine Protease Domain Possesses both trans- and cis-Cleavage Activities. *J. Virol.* **83**, 9449–9463 (2009).

28. van Kasteren, P. B. *et al.* Arterivirus and Nairovirus Ovarian Tumor Domain-Containing Deubiquitinases Target Activated RIG-I To Control Innate Immune Signaling. *J. Virol.* **86**, 773–785 (2012).

29. Frias-Staheli, N. *et al.* Ovarian Tumor Domain-Containing Viral Proteases Evade Ubiquitin- and

ISG15-Dependent Innate Immune Responses. *Cell Host Microbe* **2**, 404–416 (2007).

30. Sun, Z., Chen, Z., Lawson, S. R. & Fang, Y. The Cysteine Protease Domain of Porcine Reproductive and Respiratory Syndrome Virus Nonstructural Protein 2 Possesses Deubiquitinating and Interferon Antagonism Functions. *J. Virol.* **84**, 7832–7846 (2010).

31. Sun, Z., Li, Y., Ransburgh, R., Snijder, E. J. & Fang, Y. Nonstructural Protein 2 of Porcine Reproductive and Respiratory Syndrome Virus Inhibits the Antiviral Function of Interferon-Stimulated Gene 15. *J. Virol.* **86**, 3839–3850 (2012).

32. Li, Y. *et al.* Nonstructural proteins nsp2TF and nsp2N of porcine reproductive and respiratory syndrome virus (PRRSV) play important roles in suppressing host innate immune responses. *Virology* **517**, 164–176 (2018).

33. Kappes, M. A., Miller, C. L. & Faaberg, K. S. Porcine reproductive and respiratory syndrome virus nonstructural protein 2 (nsp2) topology and selective isoform integration in artificial membranes. *Virology* **481**, 51–62 (2015).

34. Snijder, E. J., Van Tol, H., Roos, N. & Pedersen, K. W. Non-structural proteins 2 and 3 interact to modify host cell membranes during the formation of the arterivirus replication complex. *J. Gen. Virol.* **82**, 985–994 (2001).

35. Knoops, K. *et al.* Ultrastructural Characterization of Arterivirus Replication Structures: Reshaping the Endoplasmic Reticulum To Accommodate Viral RNA Synthesis. *J. Virol.* **86**, 2474–2487 (2012).

36. Guo, R. *et al.* A swine arterivirus deubiquitinase stabilizes two major envelope proteins and promotes production of viral progeny. *PLOS Pathog.* **17**, e1009403 (2021).

37. Patel, A. *et al.* Molecular characterization of the RNA-protein complex directing -2/-1 programmed ribosomal frameshifting during arterivirus replicase expression. *J. Biol. Chem.* **295**, 17904–17921 (2020).

1467 38. Loughran, G., Firth, A. E. & Atkins, J. F. Ribosomal frameshifting into an overlapping gene in
1468 the 2B-encoding region of the cardiovirus genome. *Proc. Natl. Acad. Sci. U. S. A.* **108**, E1111
1469 (2011).

1470 39. Finch, L. K. *et al.* Characterization of Ribosomal Frameshifting in Theiler's Murine
1471 Encephalomyelitis Virus. *J. Virol.* **89**, 8580–8589 (2015).

1472 40. Napthine, S. *et al.* Protein-directed ribosomal frameshifting temporally regulates gene
1473 expression. *Nat. Commun.* **8**, (2017).

1474 41. Napthine, S., Bell, S., Hill, C. H., Brierley, I. & Firth, A. E. Characterization of the stimulators
1475 of protein-directed ribosomal frameshifting in Theiler's murine encephalomyelitis virus. *Nucleic*
1476 *Acids Res.* **47**, 8207–8223 (2019).

1477 42. Hill, C. H. *et al.* Structural and molecular basis for Cardiovirus 2A protein as a viral gene
1478 expression switch. *bioRxiv* 2020.08.11.245035 (2021). doi:10.1101/2020.08.11.245035

1479 43. Hill, C. H. *et al.* Investigating molecular mechanisms of 2A-stimulated ribosomal pausing and
1480 frameshifting in Theilovirus. *Nucleic Acids Res.* **49**, 11938–11958 (2021).

1481 44. Yuan, S., Murtaugh, M. P. & Faaberg, K. S. Heteroclite subgenomic RNAs are produced in
1482 porcine reproductive and respiratory syndrome virus infection. *Virology* **275**, 158–169 (2000).

1483 45. Di, H. *et al.* Expanded subgenomic mRNA transcriptome and coding capacity of a nidovirus.
1484 *Proc. Natl. Acad. Sci. U. S. A.* **114**, E8895–E8904 (2017).

1485 46. Kim, D. *et al.* The Architecture of SARS-CoV-2 Transcriptome. *Cell* **181**, 914-921.e10 (2020).

1486 47. Wang, D. *et al.* The SARS-CoV-2 subgenome landscape and its novel regulatory features. *Mol.*
1487 *Cell* **81**, 2135-2147.e5 (2021).

1488 48. Stewart, H. *et al.* Transcriptional and Translational Landscape of Equine Torovirus. *J. Virol.* **92**,
1489 589–607 (2018).

1490 49. Irigoyen, N. *et al.* High-Resolution Analysis of Coronavirus Gene Expression by RNA

1491 Sequencing and Ribosome Profiling. *PLoS Pathog.* **12**, e1005473 (2016).

1492 50. Dinan, A. M. *et al.* Comparative Analysis of Gene Expression in Virulent and Attenuated Strains
1493 of Infectious Bronchitis Virus at Subcodon Resolution. *J. Virol.* **93**, (2019).

1494 51. Finkel, Y. *et al.* The coding capacity of SARS-CoV-2. *Nature* **589**, 125–130 (2021).

1495 52. Ingolia, N. T., Ghaemmaghami, S., Newman, J. R. S. & Weissman, J. S. Genome-Wide Analysis
1496 in Vivo of Translation with Nucleotide Resolution Using Ribosome Profiling. *Science* **324**, 218–
1497 223 (2009).

1498 53. Steitz, J. A. Polypeptide Chain Initiation : Nucleotide Sequences of the Three Ribosomal
1499 Binding Sites in Bacteriophage R17 RNA. *Nature* **224**, 957–964 (1969).

1500 54. Wolin, S. L. & Walter, P. Ribosome pausing and stacking during translation of a eukaryotic
1501 mRNA. *EMBO J.* **7**, 3559–3569 (1988).

1502 55. Wu, C. C.-C., Zinshteyn, B., Wehner, K. A. & Green, R. High-Resolution Ribosome Profiling
1503 Defines Discrete Ribosome Elongation States and Translational Regulation during Cellular
1504 Stress. *Mol. Cell* **73**, 959-970.e5 (2019).

1505 56. Lareau, L. F., Hite, D. H., Hogan, G. J. & Brown, P. O. Distinct stages of the translation
1506 elongation cycle revealed by sequencing ribosome-protected mRNA fragments. *Elife* **2014**, 1–
1507 16 (2014).

1508 57. Li, Y., Tas, A., Snijder, E. J. & Fang, Y. Identification of porcine reproductive and respiratory
1509 syndrome virus ORF1a-encoded non-structural proteins in virus-infected cells. *J. Gen. Virol.*
1510 **93**, 829–839 (2012).

1511 58. Kreutz, L. C. & Ackermann, M. R. Porcine reproductive and respiratory syndrome virus enters
1512 cells through a low pH-dependent endocytic pathway. *Virus Res.* **42**, 137–147 (1996).

1513 59. Viehweger, A. *et al.* Direct RNA nanopore sequencing of full-length coronavirus genomes
1514 provides novel insights into structural variants and enables modification analysis. *Genome Res.*

29, 1545–1554 (2019).

60. Dobin, A. *et al.* STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15–21 (2013).

61. Lin, Y. C., Chang, R. Y. & Chueh, L. L. Leader-body junction sequence of the viral subgenomic mRNAs of porcine reproductive and respiratory syndrome virus isolated in Taiwan. *J. Vet. Med. Sci.* **64**, 961–965 (2002).

62. Yuan, S., Murtaugh, M. P., Schumann, F. A., Mickelson, D. & Faaberg, K. S. Characterization of heteroclite subgenomic RNAs associated with PRRSV infection. *Virus Res.* **105**, 75–87 (2004).

63. Xiao, C. T. *et al.* Identification of new defective interfering RNA species associated with porcine reproductive and respiratory syndrome virus infection. *Virus Res.* **158**, 33–36 (2011).

64. Erhard, F. *et al.* Improved Ribo-seq enables identification of cryptic translation events. *Nat. Methods* **15**, 363–366 (2018).

65. Ingolia, N. T., Brar, G. A., Rouskin, S., McGeachy, A. M. & Weissman, J. S. The ribosome profiling strategy for monitoring translation in vivo by deep sequencing of ribosome-protected mRNA fragments. *Nat. Protoc.* **7**, 1534–1550 (2012).

66. Kearse, M. G. & Wilusz, J. E. Non-AUG translation: A new start for protein synthesis in eukaryotes. *Genes Dev.* **31**, 1717–1731 (2017).

67. Firth, A. E. Mapping overlapping functional elements embedded within the protein-coding regions of RNA viruses. *Nucleic Acids Res.* **42**, 12425–12439 (2014).

68. Tumescheit, C., Firth, A. E. & Brown, K. CIAlign-A highly customisable command line tool to clean, interpret and 1 visualise multiple sequence alignments. 2. *bioRxiv* 2020.09.14.291484 (2020). doi:10.1101/2020.09.14.291484

69. Ingolia, N. T., Lareau, L. F. & Weissman, J. S. Ribosome profiling of mouse embryonic stem cells reveals the complexity and dynamics of mammalian proteomes. *Cell* **147**, 789–802 (2011).

70. Artieri, C. G. & Fraser, H. B. Accounting for biases in riboprofiling data indicates a major role for proline in stalling translation. *Genome Res.* **24**, 2011–2021 (2014).

71. Pavlov, M. Y. *et al.* Slow peptide bond formation by proline and other N-alkylamino acids in translation. *Proc. Natl. Acad. Sci. U. S. A.* **106**, 50–54 (2009).

72. Brierley, I., Jenner, A. J. & Inglis, S. C. Mutational analysis of the "slippery-sequence" component of a coronavirus ribosomal frameshifting signal. *J. Mol. Biol.* **227**, 463–479 (1992).

73. Choi, J., O'Loughlin, S., Atkins, J. F. & Puglisi, J. D. The energy landscape of −1 ribosomal frameshifting. *Sci. Adv.* **6**, (2020).

74. Caliskan, N. *et al.* Conditional Switch between Frameshifting Regimes upon Translation of dnaX mRNA. *Mol. Cell* **66**, 558-567.e4 (2017).

75. Qin, P., Yu, D., Zuo, X. & Cornish, P. V. Structured mRNA induces the ribosome into a hyper-rotated state. *EMBO Rep.* **15**, 185–190 (2014).

76. Chen, J. *et al.* Dynamic pathways of -1 translational frameshifting. *Nature* **512**, 328–32 (2014).

77. Dong, Q. *et al.* Gene expression in tonsils in swine following infection with porcine reproductive and respiratory syndrome virus. *BMC Vet. Res.* **17**, 1–21 (2021).

78. Lim, B. *et al.* Integrated time-serial transcriptome networks reveal common innate and tissue-specific adaptive immune responses to PRRSV infection. *Vet. Res.* **51**, 1–18 (2020).

79. Pröll, M. J. *et al.* Transcriptome profile of lung dendritic cells after in vitro porcine reproductive and respiratory syndrome virus (PRRSV) infection. *PLoS One* **12**, (2017).

80. Wilkinson, J. M. *et al.* Genome-wide analysis of the transcriptional response to porcine reproductive and respiratory syndrome virus infection at the maternal/fetal interface and in the fetus. *BMC Genomics* **17**, 1–17 (2016).

81. Zeng, N. *et al.* Transcriptome Analysis Reveals Dynamic Gene Expression Profiles in Porcine Alveolar Macrophages in Response to the Chinese Highly Pathogenic Porcine Reproductive and

Respiratory Syndrome Virus. *Biomed Res. Int.* **2018**, (2018).

82. Zhang, J. *et al.* Genome-wide analysis of long noncoding RNA profiling in PRRSV-infected PAM cells by RNA sequencing. *Sci. Rep.* **7**, 4952 (2017).

83. Zhang, K. *et al.* Global miRNA, lncRNA, and mRNA transcriptome profiling of endometrial epithelial cells reveals genes related to porcine reproductive failure caused by porcine reproductive and respiratory syndrome virus. *Front. Immunol.* **10**, 1221 (2019).

84. Xiao, Z., Zou, Q., Liu, Y. & Yang, X. Genome-wide assessment of differential translations with ribosome profiling data. *Nat. Commun. 2016 71* **7**, 1–11 (2016).

85. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15**, 550 (2014).

86. Badaoui, B. *et al.* Pig immune response to general stimulus and to porcine reproductive and respiratory syndrome virus infection: a meta-analysis approach. *BMC Genomics* **14**, 220 (2013).

87. Crisci, E. *et al.* Distinctive Cellular and Metabolic Reprogramming in Porcine Lung Mononuclear Phagocytes Infected With Type 1 PRRSV Strains. *Front. Immunol.* **11**, (2020).

88. Kusnadi, E. P., Timpone, C., Topisirovic, I., Larsson, O. & Furic, L. Regulation of gene expression via translational buffering. *Biochim. Biophys. Acta - Mol. Cell Res.* 119140 (2021). doi:10.1016/j.bbamcr.2021.119140

89. Kim, S. Y., Suh, H.-W., Chung, J. W., Yoon, S.-R. & Choi, I. *Diverse Functions of VDUP1 in Cell Proliferation, Differentiation, and Diseases*. **4**, (2007).

90. Alhawiti, N. M., Mahri, S. Al, Aziz, M. A., Malik, S. S. & Mohammad, S. TXNIP in Metabolic Regulation: Physiological Role and Therapeutic Outlook. *Curr. Drug Targets* **18**, 1095 (2017).

91. Xiang, G. *et al.* Catalytic degradation of vitamin D up-regulated protein 1 mRNA enhances cardiomyocyte survival and prevents left ventricular remodeling after myocardial ischemia. *J. Biol. Chem.* **280**, 39394–39402 (2005).

1587 92. Zhou, R., Tardivel, A., Thorens, B., Choi, I. & Tschopp, J. Thioredoxin-interacting protein links
1588 oxidative stress to inflammasome activation. *Nat. Immunol.* **11**, 136–140 (2010).

1589 93. Shalev, A. Minireview: Thioredoxin-interacting protein: Regulation and function in the
1590 pancreatic β-cell. *Molecular Endocrinology* **28**, 1211–1220 (2014).

1591 94. Nishiyama, A. *et al.* Identification of Thioredoxin-binding Protein-2/Vitamin D3 Up-regulated
1592 Protein 1 as a Negative Regulator of Thioredoxin Function and Expression *. *J. Biol. Chem.*
1593 **274**, 21645–21650 (1999).

1594 95. Lu, J. & Holmgren, A. The thioredoxin antioxidant system. *Free Radic. Biol. Med.* **66**, 75–87
1595 (2014).

1596 96. Matsuzawa, A. Thioredoxin and redox signaling: Roles of the thioredoxin system in control of
1597 cell fate. *Arch. Biochem. Biophys.* **617**, 101–105 (2017).

1598 97. Song, H. *et al.* Vitamin D3 up-regulating protein 1 (VDUP1) antisense DNA regulates
1599 tumorigenicity and melanogenesis of murine melanoma cells via regulating the expression of
1600 fas ligand and reactive oxygen species. *Immunol. Lett.* **86**, 235–247 (2003).

1601 98. Lee, K. N. *et al.* VDUP1 is required for the development of natural killer cells. *Immunity* **22**,
1602 195–208 (2005).

1603 99. Guo, R., Davis, D. & Fang, Y. Intercellular transfer of mitochondria rescues virus-induced cell
1604 death but facilitates cell-to-cell spreading of porcine reproductive and respiratory syndrome
1605 virus. *Virology* **517**, 122–134 (2018).

1606 100. Lee, S. M. & Kleiboeker, S. B. Porcine reproductive and respiratory syndrome virus induces
1607 apoptosis through a mitochondria-mediated pathway. *Virology* **365**, 419–434 (2007).

1608 101. Redza-Dutordoir, M. & Averill-Bates, D. A. Activation of apoptosis signalling pathways by
1609 reactive oxygen species. *Biochim. Biophys. Acta - Mol. Cell Res.* **1863**, 2977–2992 (2016).

1610 102. Herb, M. & Schramm, M. Functions of ros in macrophages and antimicrobial immunity.

Antioxidants **10**, 1–39 (2021).

103. HJ, F. & M, T. Redox signaling in macrophages. *Mol. Aspects Med.* **22**, 189–216 (2001).

104. Morgan, S. B. *et al.* Pathology and Virus Distribution in the Lung and Lymphoid Tissues of Pigs Experimentally Inoculated with Three Distinct Type 1 PRRS Virus Isolates of Varying Pathogenicity. *Transbound. Emerg. Dis.* **63**, 285–295 (2016).

105. Rendra, E. *et al.* Reactive oxygen species (ROS) in macrophage activation and function in diabetes. *Immunobiology* **224**, 242–253 (2019).

106. Tan, H. *et al.* The Reactive Oxygen Species in Macrophage Polarization : Human Diseases. *Oxid. Med. Cell. Longev.* **2016**, 1–16 (2016).

107. Hadri, K. El *et al.* Thioredoxin-1 Promotes Anti-Inflammatory Macrophages of the M2 Phenotype and Antagonizes Atherosclerosis. *Arterioscler. Thromb. Vasc. Biol.* **32**, 1445–1452 (2012).

108. Wang, L. *et al.* Porcine alveolar macrophage polarization is involved in inhibition of porcine reproductive and respiratory syndrome virus (PRRSV) replication. *J. Vet. Med. Sci.* **79**, 1906–1915 (2017).

109. Finkel, Y. *et al.* SARS-CoV-2 uses a multipronged strategy to impede host protein synthesis. *Nature* 1–9 (2021). doi:10.1038/s41586-021-03610-3

110. Alexander, M. R. *et al.* Ribosome-profiling reveals restricted post transcriptional expression of antiviral cytokines and transcription factors during SARS-CoV-2 infection. *Int. J. Mol. Sci.* **22**, (2021).

111. Gomez, G. N., Abrar, F., Dodhia, M. P., Gonzalez, F. G. & Nag, A. SARS coronavirus protein nsp1 disrupts localization of nup93 from the nuclear pore complex. *Biochem. Cell Biol.* **97**, 758–766 (2019).

112. Zhang, K. *et al.* Nsp1 protein of SARS-CoV-2 disrupts the mRNA export machinery to inhibit

host gene expression. *Sci. Adv.* **7**, (2021).

113. Han, M., Ke, H., Zhang, Q. & Yoo, D. Nuclear imprisonment of host cellular mRNA by nsp1β protein of porcine reproductive and respiratory syndrome virus. *Virology* **505**, 42–55 (2017).

114. Ke, H., Han, M., Kim, J., Gustin, K. E. & Yoo, D. Porcine Reproductive and Respiratory Syndrome Virus Nonstructural Protein 1 Beta Interacts with Nucleoporin 62 To Promote Viral Replication and Immune Evasion. *J. Virol.* **93**, 469–488 (2019).

115. Ke, H. *et al.* Type I interferon suppression-negative and host mRNA nuclear retention-negative mutation in nsp1β confers attenuation of porcine reproductive and respiratory syndrome virus in pigs. *Virology* **517**, 177–187 (2018).

116. Zhang, H., Wang, Y. & Lu, J. Function and Evolution of Upstream ORFs in Eukaryotes. *Trends Biochem. Sci.* **44**, 782–794 (2019).

117. Kim, D. *et al.* A high-resolution temporal atlas of the SARS-CoV-2 translatome and transcriptome. *Nat. Commun.* **12**, 1–16 (2021).

118. Kheyar, A., St-Laurent, G. & Archambault, D. Sequence determination of the extreme 5' end of equine arteritis virus leader region. *Virus Genes* **12**, 291–295 (1996).

119. Archambault, D., Kheyar, A., De Vries, A. A. F. & Rottier, P. J. M. The intraleader AUG nucleotide sequence context is important for equine arteritis virus replication. *Virus Genes* **33**, 59–68 (2006).

120. Molenkamp, R. *et al.* The arterivirus replicase is the only viral protein required for genome replication and subgenomic mRNA transcription. *J. Gen. Virol.* **81**, 2491–2496 (2000).

121. Snijder, E. J. & Meulenberg, J. J. M. The molecular biology of arteriviruses. *Journal of General Virology* **79**, 961–979 (1998).

122. Andreev, D. E. *et al.* Translation of 5' leaders is pervasive in genes resistant to eIF2 repression. *Elife* **2015**, 1–21 (2015).

123. Wissink, E. H. J. *et al.* Envelope Protein Requirements for the Assembly of Infectious Virions of Porcine Reproductive and Respiratory Syndrome Virus. *J. Virol.* **79**, 12495–12506 (2005).

124. Cao, Q. M., Subramaniam, S., Ni, Y.-Y., Cao, D. & Meng, X.-J. The non-structural protein Nsp2TF of porcine reproductive and respiratory syndrome virus down-regulates the expression of Swine Leukocyte Antigen class I. *Virology* **491**, 115–24 (2016).

125. Huo, Y. *et al.* Involvement of unfolded protein response, p53 and Akt in modulation of porcine reproductive and respiratory syndrome virus-mediated JNK activation. *Virology* **444**, 233–240 (2013).

126. Lopinski, J. D., Dinman, J. D. & Bruenn, J. A. Kinetics of Ribosomal Pausing during Programmed -1 Translational Frameshifting. *Mol. Cell. Biol.* **20**, 1095–1103 (2000).

127. Smith, A. M., Costello, M. S., Kettring, A. H., Wingo, R. J. & Moore, S. D. Ribosome collisions alter frameshifting at translational reprogramming motifs in bacterial mRNAs. *Proc. Natl. Acad. Sci. U. S. A.* **116**, 21769–21779 (2019).

128. Puray-Chavez, M. *et al.* The translational landscape of SARS-CoV-2 and infected cells. *bioRxiv* (2021). doi:10.1101/2020.11.03.367516

129. Chen, J. *et al.* Structural Basis for Helicase-Polymerase Coupling in the SARS-CoV-2 Replication-Transcription Complex. *Cell* **182**, 1560-1573.e13 (2020).

130. Van Dinten, L. C., Den Boon, J. A., Wassenaar, A. L. M., Spaan, W. J. M. & Snijder, E. J. An infectious arterivirus cDNA clone: Identification of a replicase point mutation that abolishes discontinuous mRNA transcription. *Proc. Natl. Acad. Sci. U. S. A.* **94**, 991–996 (1997).

131. Tang, C. *et al.* Helicase of type 2 porcine reproductive and respiratory syndrome virus strain HV reveals a unique structure. *Viruses* **12**, (2020).

132. Song, J. *et al.* Mapping the Nonstructural Protein Interaction Network of Porcine Reproductive and Respiratory Syndrome Virus. *J. Virol.* **92**, 1112–1130 (2018).

133. Nan, H. *et al.* The Network of Interactions Among Porcine Reproductive and Respiratory Syndrome Virus Non-structural Proteins. *Front. Microbiol.* **9**, 970 (2018).

134. Wang, T. Y. *et al.* The Nsp12-coding region of type 2 PRRSV is required for viral subgenomic mRNA synthesis. *Emerg. Microbes Infect.* **8**, 1501–1510 (2019).

135. Nedialkova, D. D. *et al.* Biochemical Characterization of Arterivirus Nonstructural Protein 11 Reveals the Nidovirus-Wide Conservation of a Replicative Endoribonuclease. *J. Virol.* **83**, 5671–5682 (2009).

136. Wang, D. *et al.* The nonstructural protein 11 of porcine reproductive and respiratory syndrome virus inhibits NF-κB signaling by means of its deubiquitinating activity. *Mol. Immunol.* **68**, 357–366 (2015).

137. Sun, Y. *et al.* Nonstructural protein 11 of porcine reproductive and respiratory syndrome virus suppresses both MAVS and RIG-I expression as one of the mechanisms to antagonize Type I interferon production. *PLoS One* **11**, e0168314 (2016).

138. Su, Y. *et al.* The Superimposed Deubiquitination Effect of OTULIN and Porcine Reproductive and Respiratory Syndrome Virus (PRRSV) Nsp11 Promotes Multiplication of PRRSV. *J. Virol.* **92**, (2018).

139. Wang, D. *et al.* Porcine Reproductive and Respiratory Syndrome Virus nsp11 Antagonizes Type I Interferon Signaling by Targeting IRF9. *J. Virol.* **93**, (2019).

140. Yang, L. *et al.* Nonstructural Protein 11 of Porcine Reproductive and Respiratory Syndrome Virus Induces STAT2 Degradation To Inhibit Interferon Signaling. *J. Virol.* **93**, (2019).

141. Shi, X. *et al.* Nonstructural protein 11 (nsp11) of porcine reproductive and respiratory syndrome virus (PRRSV) promotes PRRSV infection in MARC-145 cells. *BMC Vet. Res.* **12**, 90 (2016).

142. Chen, Z. *et al.* Sequences of 3' end of genome and of 5' end of open reading frame 1a of lactate dehydrogenase-elevating virus and common junction motifs between 5' leader and bodies of

seven subgenomic mRNAs. *J. Gen. Virol.* **74**, 643–659 (1993).

143. Taiaroa, G. *et al.* Direct RNA sequencing and early evolution of SARS-CoV-2. **2**, (2020).

144. Davidson, A. D. *et al.* Characterisation of the transcriptome and proteome of SARS-CoV-2 reveals a cell passage induced in-frame deletion of the furin-like cleavage site from the spike glycoprotein. *Genome Med.* **12**, 68 (2020).

145. St-Germain, J. R. *et al.* A SARS-CoV-2 BioID-based virus-host membrane protein interactome and virus peptide compendium: new proteomics resources for COVID-19 research. *bioRxiv* 2020.08.28.269175 (2020). doi:10.1101/2020.08.28.269175

146. Grenga, L. *et al.* Shotgun proteomics analysis of SARS-CoV-2-infected cells and how it can optimize whole viral particle antigen production for vaccines. *Emerg. Microbes Infect.* **9**, 1712–1721 (2020).

147. Han, M., Du, Y., Song, C. & Yoo, D. Degradation of CREB-binding protein and modulation of type I interferon induction by the zinc finger motif of the porcine reproductive and respiratory syndrome virus nsp1α subunit. *Virus Res.* **172**, 54–65 (2013).

148. Han, M. *et al.* Biogenesis of non-structural protein 1 (nsp1) and nsp1-mediated type I interferon modulation in arteriviruses. *Virology* **458–459**, 136–150 (2014).

149. Kim, O., Sun, Y., Lai, F. W., Song, C. & Yoo, D. Modulation of type I interferon induction by porcine reproductive and respiratory syndrome virus and degradation of CREB-binding protein by non-structural protein 1 in MARC-145 and HeLa cells. *Virology* **402**, 315–326 (2010).

150. Chen, Z. *et al.* Nuclear export signal of PRRSV NSP1α is necessary for type I IFN inhibition. *Virology* **499**, 278–287 (2016).

151. Song, C., Krell, P. & Yoo, D. Nonstructural protein 1α subunit-based inhibition of NF-κB activation and suppression of interferon-β production by porcine reproductive and respiratory syndrome virus. *Virology* **407**, 268–280 (2010).

152. Guo, H., Ingolia, N. T., Weissman, J. S. & Bartel, D. P. Mammalian microRNAs predominantly act to decrease target mRNA levels. *Nature* **466**, 835–840 (2010).

153. Chung, B. Y. *et al.* The use of duplex-specific nuclease in ribosome profiling and a user-friendly software package for Ribo-seq data analysis. *RNA* **21**, 1731–45 (2015).

154. Renaud, G., Stenzel, U. & Kelso, J. LeeHom: Adaptor trimming and merging for Illumina sequencing reads. *Nucleic Acids Res.* **42**, e141 (2014).

155. Langmead, B., Trapnell, C., Pop, M. & Salzberg, S. L. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol.* **10**, R25 (2009).

156. Yoon, J., Chung, Y.-J. & Lee, M. STADIUM: Species-Specific tRNA Adaptive Index Compendium. *Genomics Inform.* **16**, e28 (2018).

157. Gerashchenko, M. V & Gladyshev, V. N. Translation inhibitors cause abnormalities in ribosome profiling experiments. *Nucleic Acids Res.* **42**, e134 (2014).

158. Li, W. & Godzik, A. Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. *Bioinformatics* **22**, 1658–1659 (2006).

159. Anders, S., Pyl, P. T. & Huber, W. HTSeq—a Python framework to work with high-throughput sequencing data. *Bioinformatics* **31**, 166–169 (2015).

160. Huang, D. W., Sherman, B. T. & Lempicki, R. A. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat. Protoc. 2009 41* **4**, 44–57 (2008).