Expanded subgenomic mRNA transcriptome and coding capacity of a nidovirus

Han Di[a], Joseph C. Madden Jr.[a], Esther K. Morantz[a], Hsin-Yao Tang[b], Rachel L. Graham[c], Ralph S. Baric[c,d], and Margo A. Brinton[a,1]

[a]Department of Biology, Georgia State University, Atlanta, GA 30303; [b]Proteomics and Metabolomics Facility, The Wistar Institute, Philadelphia, PA 19104; [c]Department of Epidemiology, University of North Carolina at Chapel Hill, Chapel Hill, NC 27599; and [d]Department of Microbiology and Immunology, University of North Carolina at Chapel Hill, Chapel Hill, NC 27599

Edited by Paul Ahlquist, University of Wisconsin–Madison, Madison, WI, and approved August 28, 2017 (received for review April 24, 2017)

Members of the order Nidovirales express their structural protein ORFs from a nested set of 3′ subgenomic mRNAs (sg mRNAs), and for most of these ORFs, a single genomic transcription regulatory sequence (TRS) was identified. Nine TRSs were previously reported for the arterivirus *Simian hemorrhagic fever virus* (SHFV). In the present study, which was facilitated by next-generation sequencing, 96 SHFV body TRSs were identified that were functional in both infected MA104 cells and macaque macrophages. The abundance of sg mRNAs produced from individual TRSs was consistent over time in the two different cell types. Most of the TRSs are located in the genomic 3′ region, but some are in the 5′ ORF1a/1b region and provide alternative sources of nonstructural proteins. Multiple functional TRSs were identified for the majority of the SHFV 3′ ORFs, and four previously identified TRSs were found not to be the predominant ones used. A third of the TRSs generated sg mRNAs with variant leader–body junction sequences. Sg mRNAs encoding E′, GP2, or ORF5a as their 5′ ORF as well as sg mRNAs encoding six previously unreported alternative frame ORFs or 14 previously unreported C-terminal ORFs of known proteins were also identified. Mutation of the start codon of two C-terminal ORFs in an infectious clone reduced virus yield. Mass spectrometry detected one previously unreported protein and suggested translation of some of the C-terminal ORFs. The results reveal the complexity of the transcriptional regulatory mechanism and expanded coding capacity for SHFV, which may also be characteristic of other nidoviruses.

nidovirus | *Simian hemorrhagic fever virus* | transcription regulatory sequences | subgenomic mRNAs | next-generation sequencing

The virus families *Coronaviridae*, *Arteriviridae*, *Mesoniviridae*, and *Roniviridae* constitute the order Nidovirales. Nidoviruses are single-stranded, positive-sense RNA viruses that share a similar genome organization and generate a 3′ coterminal nested set of subgenomic (sg) mRNAs to express structural and accessory proteins (1–3). The genomes of the arteriviruses are approximately half the size of those of Coronaviruses. Members of the family *Arteriviridae* include *Simian hemorrhagic fever virus* (SHFV), *Lactate dehydrogenase-elevating virus* (LDV), and the well-studied viruses *Equine arteritis virus* (EAV) and *Porcine reproductive and respiratory syndrome virus* (PRRSV). SHFV infections in species of African monkeys are typically persistent and asymptomatic (4–7). In contrast, SHFV infections in Asian macaque monkeys trigger an acute, fatal hemorrhagic fever disease with death occurring 7–14 d after infection (6, 8). Macrophages (MΦs) and dendritic cells are target cells for SHFV (9).

Arterivirus genomes have a 5′ cap and a 3′ poly(A) tail. The 5′ two-thirds of the genome encodes two polyproteins, ORF1a and ORF1ab, that are auto-cleaved into 13–15 nonstructural proteins required for virus replication and transcription (10, 11). The 3′ one-third of the arterivirus genome encodes either five or nine minor structural proteins and three major structural proteins that are required for infectious virus (iv) particle production. Among the major structural proteins, GP5 and M form heterodimers in the viral envelop and were shown to be essential for EAV particle assembly (12–14). The nucleocapsid (N) protein forms homodimers that interact with each other as well as with the genomic RNA to form a "cage-like" nucleocapsid (15–17). The minor structural glycoproteins GP2, GP3, and GP4 form heterotrimers and function in cell-receptor recognition (18, 19). The minor structural protein E forms oligomers and is thought to function as an ion channel in the virion membrane during cell entry (20). The function of the recently discovered ORF5a, a possible additional minor structural protein, is not known, but knocking out the expression of this protein in the EAV genome reduced the virus yield (21, 22). Among arteriviruses, SHFV has the largest genome at ∼15.7 kb and encodes an additional nsp1 protein (nsp1γ) and an extra set of minor structural proteins (GP2′, GP3′, GP4′, and E′) (23, 24). The functions of these additional minor structural proteins are not currently known, but each is required for the production of infectious extracellular virions (25).

All nidoviruses generate a 3′ coterminal nested set of sg mRNAs to express structural and also, in some cases, accessory proteins. The production of the minus-strand templates for these sg mRNAs is regulated by transcription regulatory sequences (TRSs) located in the 3′ region of the genome that are called "body TRSs." In addition to being 3′ coterminal, the sg mRNAs of coronaviruses, arteriviruses, and mesoniviruses are also 5′ coterminal due to a discontinuous RNA synthesis mechanism (1, 3, 26, 27). The arterivirus TRSs are usually 30 to 40 nt in length and consist of a 6- to 9-nt core sequence with 10- to 15-nt flanking sequences on each side. It has been proposed that the leader TRS folds into a stem–loop structure with the core sequence located in the loop and the flanking sequences forming the stem (28, 29). There is a single 5′ leader TRS but multiple 3′ body TRSs in arterivirus genomes. The core sequences of

Significance

All members of the order Nidovirales, including *Simian hemorrhagic fever virus* (SHFV), produce subgenomic mRNAs (sg mRNAs) for their 3′ genes regulated by genomic transcription regulatory sequences (TRSs). We used a next-generation sequencing–facilitated approach to comprehensively analyze a nidovirus sg mRNA transcriptome. The discovery of high sg mRNA redundancy for individual genes and multiple previously unreported sg mRNAs encoding nonstructural proteins, alternative reading frame proteins, or C-terminal peptides of known proteins represents a paradigm shift in our understanding of SHFV genome-coding capacity and the complexity of transcription regulation that is expected to also be characteristic of other nidoviruses. High sg mRNA redundancy would ensure continued protein synthesis if a TRS is inactivated by random mutation.

Author contributions: H.D. and M.A.B. designed research; H.D., J.C.M., E.K.M., H.-Y.T., and R.L.G. performed research; R.L.G. and R.S.B. contributed new reagents/analytic tools; H.D., J.C.M., H.-Y.T., and M.A.B. analyzed data; and H.D. and M.A.B. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

[1]To whom correspondence should be addressed. Email: mbrinton@gsu.edu.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1706696114/-/DCSupplemental.

different body TRSs vary in the extent of their sequence homology to the core sequence of the leader TRS (30). As the viral RNA polymerase copies a minus strand from the 3′ end of the genome, it sequentially encounters the 3′ body TRSs. At each body TRS, the polymerase either reads through or terminates prematurely within the TRS sequence. If termination occurs, the polymerase carrying a partially transcribed minus-strand RNA disassociates from the genome. Because the 3′ end of the nascent minus-strand RNA contains a portion of a sequence complementary to the body TRS, which is also partially complementary to the 5′ leader TRS, it can realign at the 5′ end of the genome (31). Transcription then continues, generating a minus-strand sg RNA with a unique leader–body junction sequence (32, 33). The minus-strand sg RNAs generated are then efficiently transcribed into sg mRNAs. Each sg mRNA typically expresses one structural protein from the first 5′ start codon (5′ proximal ORF), but in a few instances the expression of one or more additional ORFs from a single sg mRNA has been proposed (21, 22, 24, 34).

Similar to other arteriviruses, the SHFV genome was initially thought to encode six structural protein ORFs because six strong sg mRNA bands were detected by Northern blotting of infected cell extracts (35). However, subsequent sequencing of the 3′ region of the SHFV genome revealed the presence of ORFs for an extra set of minor structural proteins, GP2′, GP3′, and GP4′ (36). The body TRSs for GP2′ and GP4′ were then discovered, but a separate body TRS for SHFV GP3′ was not identified (24, 36). In a recent study, a sg mRNA3′ encoding GP3′ as the 5′ ORF was detected by Northern blotting, and the corresponding body TRS3′ was identified by RT-PCR amplification, cloning, and sequencing (25). This study also detected an additional sg mRNA band between the sg mRNA5 and sg mRNA6 bands on the Northern blots that was not characterized (25). Additional ORFs encoding the E and ORF5a proteins were discovered in other arterivirus genomes, and the SHFV genome was predicted to encode three additional proteins, E, E′, and ORF5a. It was proposed that the arterivirus E and ORF5a proteins are expressed from the second ORF of bicistronic sg mRNAs (21, 22, 34).

In the present study, the presence of an additional sg mRNA band between sg mRNA5 and sg mRNA6 was confirmed, and this band was shown to contain seven different sg mRNAs of similar size, each regulated by a unique body TRS. All seven of these sg mRNAs encode the same in-frame C-terminal region of GP5. To identify additional functional TRSs, the SHFV sg mRNA transcriptome was analyzed first by amplification, cloning, and sequencing of sg mRNA leader–body junctions and then by next-generation sequencing (NGS) of mRNAs extracted from SHFV-infected MA104 cells and macaque MΦs. A total of 96 functional body TRSs were identified in the SHFV genome; the majority were located in the 3′ structural protein region of the genome, but some were also located in the ORF1a/1b region and produced sg mRNAs encoding nonstructural proteins. Thirty-four of the identified TRSs produced sg mRNAs with two or three variant leader–body junction sequences. The relative abundance of sg mRNAs produced at individual TRSs remained consistent at early and late times post infection in both cell types analyzed. The majority of the newly identified TRSs produced alternative sg mRNAs for known structural proteins. Separate TRSs and sg mRNAs were identified for E′, GP2, and ORF5a. In addition, sg mRNAs encoding in-frame C-terminal ORFs of a number of structural proteins as well as previously unreported ORFs in alternative reading frames were detected. Mass spectrometry analysis of the viral proteome in SHFV-infected MA104 cells detected all the previously identified and predicted SHFV proteins except for ORF5a and nsp6. This analysis also detected one of the alternative reading frame proteins and suggested the production of some of the C-terminal ORFs. The expanded sg mRNA transcriptome and coding capacity and the complex but consistently regulated sg mRNA production for individual ORFs discovered for SHFV are likely characteristic of other nidoviruses.

Results

An Additional sg mRNA Band Was Consistently Detected in SHFV-Infected MA104 Cell Lysates by Northern Blotting. The SHFV structural protein ORFs are expressed from sg mRNAs. In an initial study, six strong sg mRNA bands, which were thought to encode three minor and three major structural proteins as in other arteriviruses, were detected by Northern blotting of RNA extracted from SHFV-infected MA104 cell lysates (35, 37). However, subsequent sequencing of the 3′ end of the SHFV genome predicted the presence of three additional minor structural protein ORFs in this region and suggested that nine sg mRNAs should have been detected (36). A recent Northern blot analysis performed with a digoxigenin (DIG)-labeled 5′ leader probe on RNA extracted from SHFV-infected MA104 cells [multiplicity of infection (MOI) of 1] at 8, 16, and 24 h post infection (hpi) detected nine sg mRNA bands with sizes corresponding to those predicted for the nine ORFs (25). However, an additional sg mRNA band was consistently detected between sg mRNA5 and sg mRNA6 (25). To confirm the production of the additional sg mRNA band, MA104 cells were infected with SHFV infectious clone (SHFVic) virus at an MOI of 1, and total intracellular RNA was extracted at different times after infection and subjected to Northern blotting analysis using a DIG-labeled sg mRNA7 probe (Table S1). A genomic RNA band and 10 sg mRNA bands were detected by 8 hpi (Fig. 1*A*). The sizes of nine of these sg mRNA bands were 5.0, 4.4, 4.0, 3.5, 2.8, 2.6, 1.9, 1.2, and 0.6 kb and corresponded to the sizes of the previously identified sg mRNA2′, sg mRNA3′, sg mRNA4′, sg mRNA2, sg mRNA3, sg mRNA4, sg mRNA5, sg mRNA6, and sg mRNA7, respectively (25). The additional ∼1.7-kb band was detected between sg mRNA5 (1.9 kb) and sg mRNA6 (1.2 kb). To further characterize this sg mRNA, MA104 cells were infected with SHFVic at an MOI of 1, and at 12 hpi total intracellular RNA was extracted from three biological-repeat experiments and was subjected to Northern blotting using either a sg mRNA5 probe (targeting the region between TRS5 and TRS6) or a sg mRNA6 probe (targeting the region between TRS6 and TRS7) (Table S1). The sg mRNA5 probe detected the bands corresponding to all the previously identified sg mRNAs except for sg mRNA6 and sg mRNA7 and also detected the additional ∼1.7-kb band (Fig. 1*B*). The sg mRNA6 probe detected all the sg mRNAs except for sg mRNA7 and also strongly detected the ∼1.7-kb band. Based on the Northern blotting data, the ∼1.7-kb sg mRNA band was predicted to be generated from a body TRS located between TRS5 and TRS6.

Identification of the Body TRS for the ∼1.7-kb sg mRNA. To identify the body TRS for the ∼1.7-kb sg mRNA, a set of primers was designed to be complementary to the 5′ leader region and a region upstream of TRS6 and used for RT-PCR amplification of RNA extracted from SHFVic-infected MA104 cells (MOI of 1) at 24 hpi (Fig. S1*A*). Multiple strong bands close to the predicted size (852 bp) of the sg mRNA5 leader–body junction as well as a cluster of fainter, faster-migrating bands (∼500 bp) were detected (Fig. S1*B*). No bands were detected in the mock-infected samples. The region of the gel containing the faster-migrating bands was excised, and the DNA was extracted and cloned into a TA vector. Forty colonies were randomly selected, and the plasmid DNAs were extracted. The leader–body junction inserts were cut out by restriction digestion and separated by gel electrophoresis (Fig. S1*C*). Clones containing leader–body junctions of different sizes were sequenced. Alignment of the resulting sequences with both the 5′ leader and the 3′ SHFV genome sequences revealed seven additional functional body TRSs located between TRS5 and TRS6 (Fig. S2). The sg mRNAs generated from all seven of these TRSs encoded the same in-frame, C-terminal peptide of GP5 that was designated "ORF5-C-68aa" (Fig. S1*D*).

Identification of Additional Functional Body TRSs in the SHFV Genome. The discovery of multiple previously unreported functional body TRSs between TRS5 and TRS6 suggested that the additional band detected on the Northern blots (Fig. 1) contained a group of sg mRNAs generated from nearby body TRSs. Multiple PCR bands were also detected in the region of the gel containing the predicted

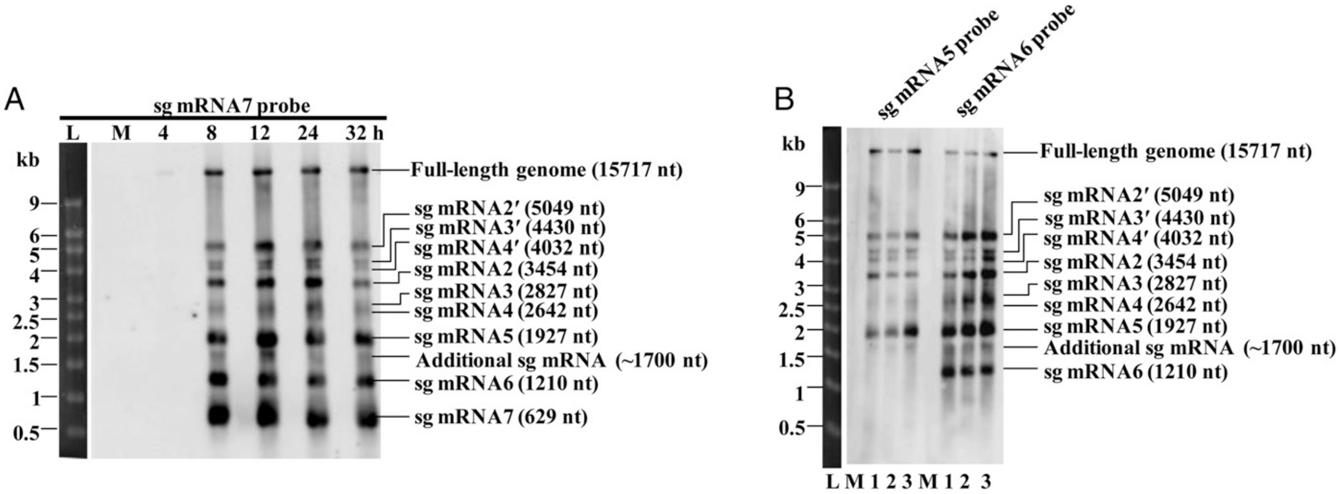

Fig. 1. Northern blot analysis of sg mRNAs produced in SHFVic-infected MA104 cells. (*A*) MA104 cells were mock infected or infected with SHFVic at an MOI of 1. At different times post infection, total intracellular RNA was extracted, and 1 μg of RNA was separated on a 1% denaturing agarose gel followed by transfer to an Hybond-N$^+$ membrane. After UV cross-linking, the membrane was hybridized to a DIG-labeled RNA probe specific for sg mRNA7. (*B*) Total intracellular RNA collected at 12 hpi from three biological repeats was hybridized separately to DIG-labeled RNA probes specific for sg mRNA5 or sg mRNA6. The RNA bands detected were labeled based on the estimated sizes of the predicted structural protein sg mRNAs and on the probe used. The additional sg mRNA band detected is indicated. The RNA ladder was cut from the membrane, stained with methylene blue, and imaged. L, DNA ladder; M, mock-infected; 1, 2, and 3 indicate the different biological repeat samples tested.

sg mRNA5 leader–body junction (852 bp) (Fig. S1*B*), strongly suggesting the existence of additional functional TRSs near TRS5 and possibly also adjacent to the previously identified body TRSs for the other 3′ SHFV ORFs. To identify additional functional TRSs within the 3′ end of the SHFV genome, nine reverse primers were designed, each targeting a region downstream of the identified body TRS of one of the 3′ ORFs. Each reverse primer together with the same 5′ leader forward primer was used for RT-PCR amplification of the leader–body junctions of sg mRNAs generated from the corresponding 3′ region of the genome (Table S2). After RT-PCR amplification and TA cloning, 20 clones were randomly selected for each of the nine regions and sequenced. All nine of the previously published body TRSs were detected, and a total of 36 additional functional body TRSs were discovered (Table S3). Twenty-one of these 36 body TRSs functioned as redundant body TRSs generating minus-strand templates for sg mRNAs encoding known structural proteins (Table S3). The remaining 15 of these additional body TRSs generated minus-strand templates for sg mRNAs encoding an in-frame, C-terminal peptide of a known structural protein. These peptides were designated "ORF2a′-C-25aa," "ORF4′-C-84aa," "ORF2a-C-161aa," ORF5-C-68aa, and "ORF6-C-79aa" (Fig. 2 and Table S3).

The SHFV sg mRNA2′, sg mRNA2, and sg mRNA5 were previously predicted to bicistronically express GP2′/E′, E/GP2, and GP5/ORF5a proteins, respectively. In sg mRNA2′, the start codon for E′ is located 91 nt downstream of the GP2′ start codon. Two additional functional body TRSs were identified in this 91-nt region that produced sg mRNAs encoding E′ as the 5′ proximal ORF (Table S3). In sg mRNA2, the GP2 start codon is located only 37 nt downstream of the E start codon. One additional functional body TRS was identified in this 37-nt region that produced a sg mRNA encoding GP2 as the 5′ proximal ORF (Table S3). The data indicate that E′ and GP2 can be expressed monocistronically from separate sg mRNAs. However, the possibility that bicistronic expression of these proteins may also occur was not ruled out. A body TRS that generates a separate sg mRNA encoding ORF5a as its 5′ proximal ORF was not identified after TA cloning of the amplified leader–body junctions.

Mutation of the Start Codon of ORF5-C-68aa or ORF6-C-79aa Reduced SHFV Yield from MA104 Cells. Among the 36 body TRSs identified in this work, 15 produced sg mRNAs encoding the in-frame C-terminal peptide of a known structural protein as the 5′ proximal ORF. As an initial means of analyzing the functional relevance of these C-terminal peptides, the start codon for each peptide was separately mutated in the SHFVic LVR strain to generate SHFVic-ΔORF2a′-C-25aa, SHFVic-ΔORF4′-C-84aa, SHFVic-ΔORF2a-C-161aa, SHFVic-ΔORF5-C-68aa, and SHFVic-ΔORF6-C-79aa. In each case, the nucleotide substitution did not change the amino acid in the overlapping ORF (Table S4). Passage 1 (P1) virus stocks were generated as described in *Materials and Methods*. MA104 cells were infected with either wild-type infectious clone P1 virus or with one of the mutant P1 viruses at an MOI of 0.5, and viral infectivity in culture fluids harvested at different times after infection was assessed by plaque assay in MA104 cells. Compared with wild-type SHFV, the SHFVic-ΔORF5-C-68aa and SHFVic-ΔORF6-C-79aa mutant viruses produced decreased virus yields starting at 24 hpi (Fig. 3*A*), whereas the SHFVic-ΔORF2a′-C-25aa, SHFVic-ΔORF4′-C-84aa, and SHFVic-ΔORF2a-C-161aa mutant viruses produced yields similar to those of the wild-type virus (Fig. 3*B*). The SHFVic-ΔORF5-C-68aa and SHFVic-ΔORF6-C-79aa mutant viruses also produced small plaques (Fig. 3*C*). These data suggested that ORF5-C-68aa and ORF6-C-79aa could be functionally important during the viral replication cycle.

NGS Identification of Functional SHFV Body TRSs and sg mRNAs with Alternative Leader–Body Junctions Generated from the Same Body TRS. Although the strategy based on RT-PCR amplification of leader–body junctions from SHFV sg mRNAs facilitated the discovery of a number of previously unreported functional body TRSs, the depth of this analysis was limited both by the total number of clones screened in each region and by the relative abundance of each sg mRNA. To increase the depth of the analysis and to identify sg mRNAs with low abundance, MA104 cells were infected with SHFVic at an MOI of 1, total RNA was extracted at 8 and 18 hpi, and mRNAs were isolated and subjected to Illumina HiSeq analysis. Three biological repeats for each time point were sequenced and analyzed with CLC Genomics Workbench software. A workflow, which is described in *Material and Methods*, was designed and used to identify sg mRNA reads containing a leader–body junction sequence and then to sequentially map them to each of the 45 identified leader–body junction sequences (15 nt in length). Multiple reads in both the 8- and 18-h infected samples mapped to each of the 45 junction sequences, confirming the

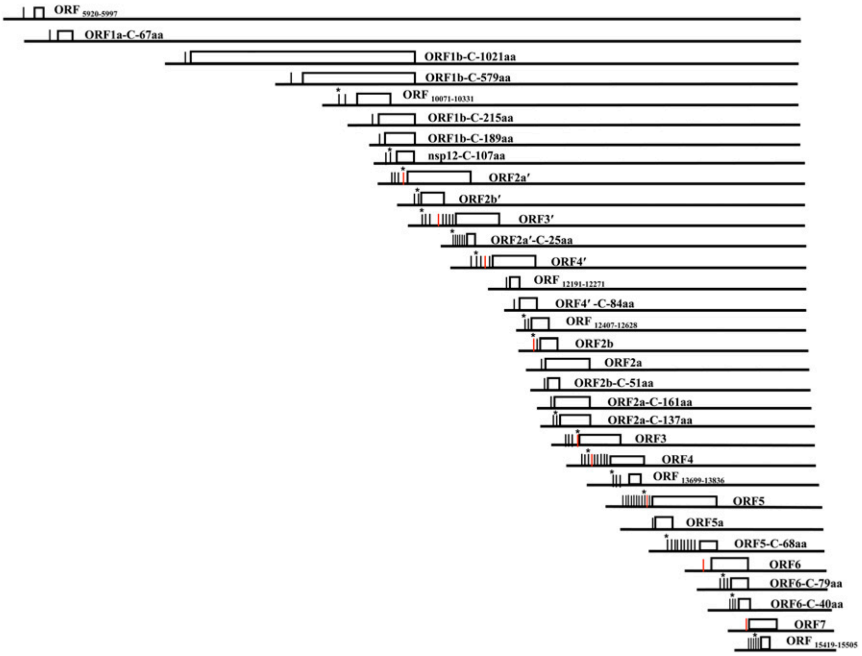

Fig. 2. Diagram of the genome locations of the identified SHFV body TRSs and ORFs. Previously published body TRSs are in red. When more than one body TRS is used for an ORF, an asterisk indicates the TRS with the highest sg mRNA abundance.

production of sg mRNAs from all the identified SHFV body TRSs at both an early and a later time post infection. Surprisingly, ~64% of the sg mRNA reads in the samples collected at both times remained unmapped, indicating the junction sequences they contained did not have 100% homology with any of the identified leader–body junctions and suggesting the existence of many additional sg mRNAs and body TRSs. To locate the positions of the additional functional body TRSs in the SHFV genome, all the remaining sg mRNA reads from the 18-h MA104 sample were mapped to an SHFV genome sequence without the 5′ end (nucleotides 250–15,717) (Fig. 4). The majority of the remaining sg mRNA reads mapped to the 3′ region of the genome, with the highest number mapping close to ORF7 as indicated by the large peak in the read coverage that contained about half the remaining reads. Inspection of the sequences of the reads mapping to this region revealed an sg mRNA with an alternative leader–body junction sequence (5′-TCCTTAACCTGAGGA-3′) generated from the published TRS7 during discontinuous synthesis of minus-strand RNA (Table S5). Unexpectedly, some of the remaining sg mRNA reads mapped to various locations in the ORF1a/1b region (Fig. 4). Although the abundance of these reads was lower than that of those mapping to the 3′ end of the genome, their detection indicated that discontinuous RNA synthesis can also occur within the ORF1a/1b region, with a higher number of reads from the ORF1b region than from the ORF1a region.

To identify all the functional body TRSs in the SHFV genome and all the alternative leader–body junctions generated from the same body TRS revealed by the NGS data, all the mapped remaining sg mRNA reads were analyzed for leader–body junction sequences. Fifty-one additional previously unreported body TRSs were identified, giving a total of 96 functional body TRSs in the SHFV genome (Fig. 2). Thirty-four of these 96 body TRSs generate two or more sg RNAs with alternative leader–body junction sequences, resulting in a total of 137 sg mRNAs, each containing a unique leader–body junction sequence (Table S5). A previously unreported body TRS (5′-TACTTATGT-3′) that generates the template for a separate sg mRNA with ORF5a as its 5′ proximal ORF was identified. Two previously unreported functional body TRSs were discovered in the ORF1a region, and eight were discovered in the ORF1b region that generated minus-strand templates for long sg mRNAs with 30 or more reads mapped at 18 hpi. Seven of these long sg mRNAs encode different lengths of in-frame C-terminal ORF1a or ORF1b peptides. Three of these long sg mRNAs encode two previously unreported ORFs in alternative reading frames: $ORF_{10,071–10,331}$ encoding an 86-aa protein and $ORF_{5,920–5,997}$ encoding a 25-aa protein (Fig. 2), Eleven functional TRSs were also identified in the 3′ structural protein region that generated sg mRNAs encoding four additional previously unreported ORFs in alternative reading frames: $ORF_{15,419–15,505}$ encoding a 28-aa protein, $ORF_{12,407–12,628}$ encoding a 73-aa protein, $ORF_{12,191–12,271}$ encoding a 26-aa protein, and $ORF_{13,699–13,836}$ encoding a 45-aa protein (Fig. 2).

The Relative Abundance of Individual SHFV sg mRNAs Was Consistent at Early and Late Times Post Infection in MA104 Cells and Macaque MΦs. The number of reads mapping to each of the 137 unique leader–body junction sequences was used to estimate the transcription level of sg mRNA produced from each body TRS. Separate calculations were made from the data generated for each of the three biological repeats. Because the values obtained for the three repeats were very similar, read data from one representative repeat are shown in Table S6. In cases where sg mRNAs with variant leader–body junction sequences were produced from a single body TRS (indicated by a dagger in Table S6), the reads for each of these sg mRNAs were added together to obtain the total transcription level from that body TRS (Table S6, "per TRS" column). When multiple body TRSs produced sg mRNAs encoding the same ORF, the sum of the reads for all these TRSs was used to indicate the total transcription level for that ORF (Table S6, "per ORF" column). The total number of reads for each body TRS was higher at 18 hpi than at 8 hpi, indicating increased transcription from each body TRS with time after infection.

The relative abundance of sg mRNAs generated from each SHFV body TRS was calculated by dividing the transcription level from each body TRS by the total transcription level from all the body TRSs (Table S7, per TRS column). The relative abundance of sg mRNAs generated from the different body TRSs varied. The sg mRNAs generated from the published TRS7 had the highest abundance (232,720 reads at 8 hpi and 310,224 at 18 hpi, ~49.53% and 50.55% abundance, respectively). The sg mRNA with the

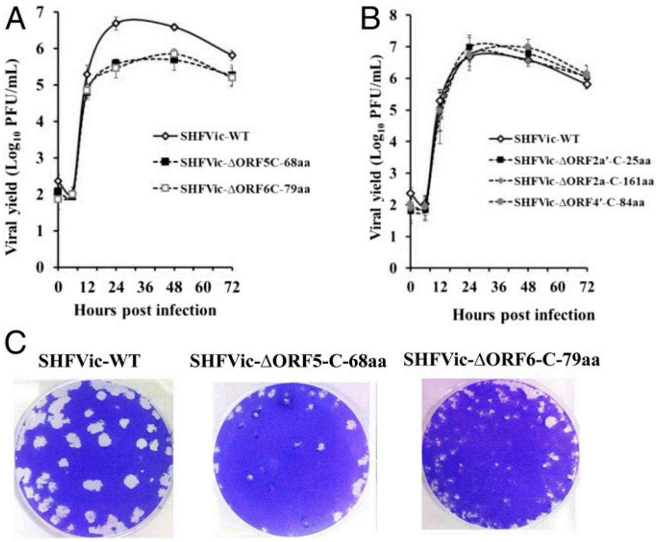

Fig. 3. Effect of mutagenesis of the start codons of the C-terminal ORFs encoded by identified sg mRNAs on virus production in MA104 cells. (*A* and *B*) MA104 cells were infected (MOI of 0.5) with P1 of wild-type or the indicated start codon-mutant SHFVic virus. At different times after infection, the culture fluid was collected, and virus infectivity was determined by plaque assay on MA104 cells. (*C*) Comparison of the diameters of the plaques produced by wild-type and mutant SHFVic viruses harvested at 24 hpi.

lowest abundance was one generated from a previously unreported body TRS (5′-ttcttcgcc-3′) located near the published TRS5 (four reads at 18 hpi and three at 8 hpi, 0.0009% and ∼0.0004% abundance, respectively) (Tables S6 and S7). Interestingly, the sg mRNA generated from an alternative TRS5 (5′-atcataacc-3′), rather than the published TRS5, was the third most abundant, indicating that the published TRS5 is not the primary TRS used. The previously identified TRSs for GP3′, GP4′, and GP4 were also not the major ones used. To compare the relative abundance of the sg mRNAs generated from each TRS at 8 and 18 hpi, the fold change over time was calculated for each TRS. Interestingly, the relative abundance of sg mRNAs for each body TRS showed little change with time after infection (Table S7). These data indicate that the production of SHFV sg mRNAs from each body TRS is consistent throughout the infection cycle.

To analyze the effect of the location of a TRS in the genome on the abundance of sg mRNA generated, the location of each body TRS in the SHFV genome was plotted against the corresponding relative sg mRNA abundance from that TRS. The highest density of body TRSs is in the 3′ region where the SHFV structural proteins are encoded (nucleotides 10,953–15,630) (Fig. 5). Although the TRS for the 3′ terminal ORF7 produced the highest abundance of sg mRNAs, a sequential decrease in sg mRNA abundance with increasing distance of a TRS from the 3′ end of the genome was not observed. Instead, multiple regions containing body TRSs with high transcription activity were flanked by regions containing body TRSs with low transcription activity. To analyze the impact of duplex stability between leader and body TRSs on sg mRNA abundance, the lowest free energy (kcal/mol) of the duplex (12 bp) formed between the leader TRS and each of the seven closely located body TRSs for ORF5-C-68aa was calculated (Fig. S2). A linear correlation between duplex stability and sg mRNA abundance was not observed, suggesting that other factors, such as RNA secondary structure, play a major role in determining sg mRNA abundance.

MΦs and dendritic cells are targeted by SHFV during infections in monkeys. To determine whether the TRS use and relative abundance of individual sg mRNAs observed in MA104 cells are accurate representations of what occurs during an in vivo infection, primary rhesus macaque MΦs were infected with SHFVic at an MOI of 1, and mRNAs isolated from total RNA extracted at 7 or 16 hpi was subjected to NGS using Illumina MiSeq. The sequencing data were analyzed using CLC Genomics Workbench software and the same workflow described above. Although fewer total reads were obtained due to the lower capacity of Illumina MiSeq compared with HiSeq, each of the 96 body TRSs identified in the SHFV genome was also functional at both early and late times post infection in MΦs, with the exception of the body TRS (5′-ttcttcgcc-3′). This body TRS had the lowest number of mapped reads in MA104 cells (Table S6). There were no reads mapped to this TRS in the 7-h infected MΦ sample, but two reads mapped to it in the 16-h sample (Table S8). The lower capacity of Illumina MiSeq likely negatively impacted the detection of the rare sg mRNAs produced from this TRS. The relative sg mRNA abundance produced at the majority of the SHFV body TRSs was similar in the infected MΦ and MA104 samples and remained consistent in both types of cells with time after infection (Table S7). The consistent sg mRNA abundance for individual TRSs at different times after infection in both a cell line and primary MΦs indicates that transcription is regulated by viral, not cellular, factors.

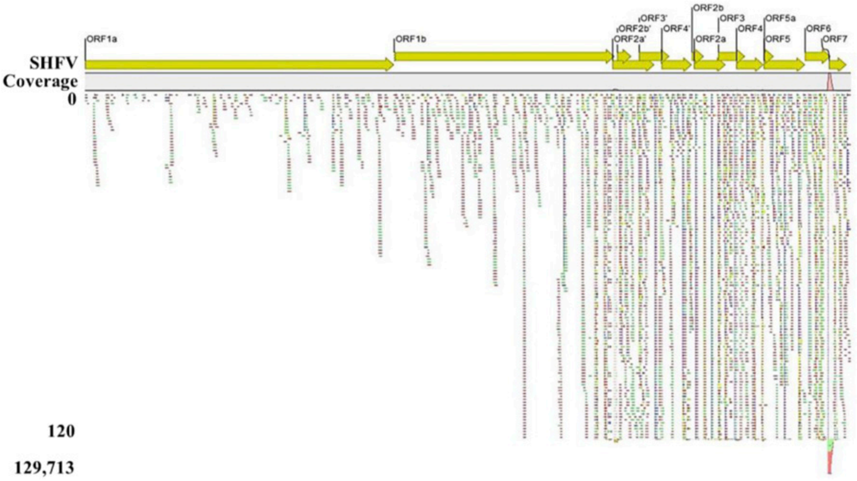

Fig. 4. Genome alignment of the NGS sg mRNA reads that did not contain one of the 45 identified leader–body junction sequences. MA104 cells were infected with SHFVic at an MOI of 1. At 18 hpi, total intracellular RNA was extracted, and mRNA was isolated and subjected to library preparation and RNA sequencing (RNA-seq) using Illumina HiSeq. The resulting reads were trimmed, and SHFV sg mRNA reads were extracted as described in *Material and Methods*. All the sg mRNA reads were mapped to each of the 36 newly identified as well as to the nine previously published 3′ region leader–body junctions. The remaining unmapped sg mRNA reads were then mapped to an SHFV genome sequence without the 5′ end (nucleotides 250–15,717). The mapping results are displayed with the read depth of 0–120 shown in detail and the remainder of the reads shown in compact view. The previously known SHFV genome ORFs are indicated by horizontal arrows at the top with the start site of each ORF indicated by a vertical line. The pink coverage peak indicates a region with a very high number of mapped reads. Each green dot represents a single forward read, each red dot represents a single reverse read, and each blue dot represents a paired read. The numbers on the left side indicate the read depth.

Mass Spectrometry Analysis of SHFV Proteins in Infected MA104 Cells. The relative abundance of an SHFV ORF product can be estimated by the relative abundance of the sg mRNAs expressing it. When an ORF was encoded as the 5′ ORF in multiple sg mRNAs generated from different TRSs, then the combined sg mRNA abundance from all these TRSs was used to estimate the relative abundance of that ORF (Table S7, per ORF column). Among the three major structural proteins, ORF7 (N, 50.55%) had the highest abundance, followed by ORF6 (M, 10.15%) and ORF5 (GP5, 9.50%). The abundances for the minor structural proteins were lower: 5.69% for ORF2a′ (GP2′), 4.73% for ORF4 (GP4), 4.14% for ORF2b (E), 3.65% for ORF2b′ (E′), 2.66% for ORF4′ (GP4′), 1.48% for ORF3′ (GP3), and 1.15% for ORF3 (GP3). ORF2a (GP2, 0.09%) and ORF5a (0.002%) had the lowest abundance. Among all the in-frame C-terminal ORFs identified, ORF5-C-68aa had the highest abundance (1.43%), which is consistent with the detection of its sg mRNAs on the Northern blots (Fig. 1). Each of the previously unreported ORFs encoded in an alternative reading frame had low abundance (Table S7, per ORF column).

To directly analyze the expression of the SHFV proteins, MA104 cells were infected with SHFVic virus at an MOI of 1. At 18 hpi, cell lysate was collected in radioimmunoprecipitation assay (RIPA) buffer and briefly electrophoresed 1 cm into a 12% NuPAGE Bis-Tris gel. The region of the gel containing proteins smaller than 75 kDa was excised and subjected to LC-MS/MS analysis. The MS/MS spectra generated were searched against a SHFV protein database that contains all the predicted ORFs (≥20 aa) encoded in all three reading frames on both the plus- and minus-strand SHFV RNAs. Each SHFV protein detected was quantified using the intensity-based absolute quantification (iBAQ) method (Table 1). The mass spectrometry analysis detected all the previously identified or predicted SHFV structural proteins except for ORF5a, which had the lowest sg mRNA abundance (0.002%) (Table S7). The relative amount of each SHFV structural protein correlated well with its estimated sg mRNA abundance, with the exception of ORF2a (GP2), ORF2a′ (GP2′), and ORF3′ (GP3′). ORF2a′ (GP2′) had the fourth most abundant sg mRNA, whereas its protein level ranked eighth. Conversely, the ORF3′ (GP3′) protein level ranked sixth, while the sg mRNA level ranked ninth. ORF2a (GP2) had the most surprising discrepancy with a very low sg mRNA abundance (0.09%, ranked 11th) but a very high amount of protein (iBAQ intensity of 9.88 E+08, ranking seventh). This finding suggested that although a separate body TRS was identified for ORF2a (GP2), the high abundance of this protein strongly suggests that it is also bicistronically expressed from the second ORF in the sg mRNAs encoding ORF2b (E), which are highly abundant.

Because the mass spectrometry analysis could not differentiate trypsin cleavage peptides generated from in-frame C-terminal proteins from those generated from the overlapping full-length proteins, the amounts of the C-terminal proteins were not estimated. However, the peptide-mapping diagrams for ORF4′, ORF2b, ORF2a, ORF5, and ORF6 showed a greater abundance of peptides mapping to the C-terminal region of these proteins, suggesting the possibility that these in-frame C-terminal proteins as well as the full-length proteins are produced in infected cells. Among the ORF1a/1b C-terminal ORFs, the in-frame C-terminal ORF of nsp12 (nsp12-C-107aa) had the highest sg mRNA abundance, followed by ORF1b-C-579aa and ORF1b-C-215aa (Table S7). Accordingly, even though the ORF1b polyprotein is translated at a much lower efficiency (15–20%) than ORF1a (38), the mass spectrometry analysis detected a higher amount of nsp11 (iBAQ intensity of 1.03E+08) and nsp12 (iBAQ intensity of 2.58E+08), which are the 3′ terminal proteins of ORF1b, than of nsp8 (iBAQ intensity of 8.29E+07), which is the 3′ terminal protein of ORF1a (Table 1). The data suggest that the ORF1b C-terminal ORFs are translated and provide an additional source of nsp11 and nsp12.

Among the six previously unreported ORFs located in alternative reading frames encoded by newly identified sg mRNAs, only the product of $\mathrm{ORF}_{10,071-10,331}$ was detected by mass spectrometry, confirming its expression in infected cells (Table 1). $\mathrm{ORF}_{10,071-10,331}$ encodes an 86-aa protein from an alternative reading frame in the ORF1b region and is the largest of the proteins expressed from an alternative reading frame. The mass spectrometry analysis also detected all the previously known and predicted SHFV nonstructural proteins except for nsp6, which is only 13 aa in length and contains a trypsin cleavage site (Table 1). Evidence for both −1 and −2 frameshift activities at a site in the nsp2 region of the genome was previously reported for PRRSV (39). The nsp2 frameshift sequence is conserved in the SHFV genome. The mass spectrometry analysis detected the SHFV nsp2 −1 and −2 frameshift products and showed that the SHFV nsp2 −2 frameshift product (nsp2_−2_TF, iBAQ intensity of 3.08E+07) is more abundant than the −1 frameshift product (nsp2_−1_TF, iBAQ intensity of 3.76E+06) (Table 1). No peptides from any of the predicted ORFs in the SHFV minus-strand RNA were detected by the mass spectrometry analysis.

Discussion

The defining characteristic of members of the order Nidovirales is that the protein ORFs located at the 3′ end of the genome are expressed from a 3′ coterminal nested set of sg mRNAs, and the synthesis of the minus-strand templates for these sg mRNAs is regulated by TRSs. Eight body TRSs were initially identified in the SHFV genome, with each considered to be the TRS used for generating the sg RNA of one of the known structural proteins (24). We subsequently identified a body TRS and sg mRNA for a ninth structural protein GP3′ by Northern blotting and RT-PCR (25). In addition to the nine SHFV sg mRNA bands detected by Northern blotting, whose sizes corresponded to the sizes predicted for the sg mRNAs of the known structural proteins—GP2′, GP3′, GP4′, GP2, GP3, GP4, GP5, M, and N—an additional uncharacterized band was detected between sg mRNA5 and sg mRNA6 with a 5′ leader probe (25). Although not characterized, a sg mRNA band at the same relative gel position was evident on multiple previously published Northern blots of EAV sg mRNAs, suggesting that this additional sg mRNA may also be produced in EAV-infected cells (31, 33, 40). In the present study, the additional band was also detected with probes specific for sg mRNA7, sg mRNA6, or sg mRNA5, confirming its location between sg mRNA5 and sg mRNA6. Leader–body junctions amplified from sg mRNAs produced in SHFV-infected MA104 cells using a 5′ leader primer and a 3′ body primer located upstream of TRS6 identified seven adjacent body TRSs between TRS5 and TRS6 and indicated that the broad band detected by Northern blotting was composed of multiple sg mRNAs of slightly different sizes.

Subsequent NGS analyses identified two more body TRSs between TRS5 and TRS6 for a total of nine functional TRSs in this region. The sg mRNAs produced from each of these nine TRSs were predicted to encode the same in-frame, C-terminal peptide of GP5 (ORF5-C-68aa). The start codon for this ORF is located downstream of the last GP5 transmembrane domain, so this protein would be expected to be located in the cytoplasm of infected cells. Although the expression of an ORF5-C peptide by an arterivirus has not previously been predicted or reported, two previous studies of the coronaviruses *Severe acute respiratory syndrome virus* and *Infectious bronchitis virus* identified a novel sg mRNA encoding an in-frame, C-terminal peptide of the virion spike protein (41, 42), and an in-frame truncated spike protein was also reported for *Porcine respiratory coronavirus* (43, 44). Compared with the other identified SHFV in-frame C-terminal peptides, ORF5-C-68aa and ORF6-C-79aa have much higher sg mRNA abundances, and these were the two that showed a decreased virus yield after mutation of their start codon, suggesting that ORF5-C-68aa and ORF6-C-79aa are functionally important during the virus replication cycle. However, the possibility that the conserved Met-to-Leu substitution in the full-length protein was responsible for the effect observed was not ruled out. Although no evidence of function in cell culture was obtained for the ORF4′-C-84aa, ORF2a-C-161aa, and ORF2a′-C-25aa proteins, the

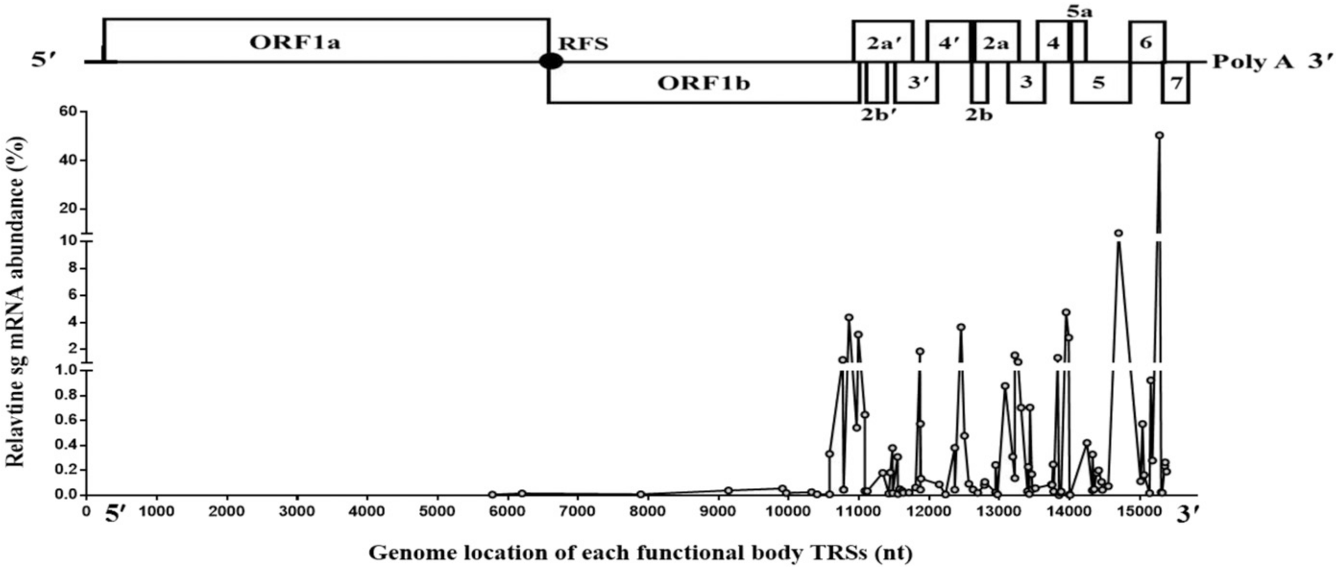

Fig. 5. Relative abundance of the sg mRNAs produced from each identified functional SHFV body TRS. The percent abundance of the sg mRNAs produced from each identified functional body TRS was graphed according to its genome location. A schematic diagram of the SHFV genome is shown at the top. RFS, ribosomal frameshift.

possibility that these peptides may be functionally relevant in infected animals has not been investigated.

The initial discovery of multiple functional TRSs for ORF5-C-68aa suggested the possibility that additional TRSs were used for the expression of sg mRNAs for some or all of the other SHFV 3′ ORFs. A total of 36 functional TRSs were subsequently identified by amplification, cloning, and sequencing, and another 51 were identified by NGS analysis of SHFV sg mRNAs from infected MA104 cells. The majority of these TRSs produced alternative sg mRNAs for known SHFV structural proteins. For five of the structural protein ORFs, the relative abundance of the sg mRNA produced from the previously published TRS was higher than that of the sg mRNAs produced from any of the newly identified alternative TRSs, confirming that the published TRS is the major TRS used. However, for GP3′, GP4′, GP4, and GP5, one of the newly identified alternative TRSs produced sg mRNAs at a higher

Table 1. Mass spectrometry analysis of SHFV proteins produced in MA104 cells

SHFV proteins	kDa	% coverage	Razor + unique peptides	iBAQ Intensity*
SHFV ORF7 (N)	12.284	62.2	8	1.54E+10
SHFV ORF6 (M)	17.851	61.7	10	2.71E+09
SHFV ORF5 (GP5)	31.294	19.8	5	1.39E+09
SHFV ORF2b (E)	8.7474	20	2	1.25E+09
SHFV ORF4 (GP4)	19.793	24.2	3	1.23E+09
SHFV ORF3′ (GP3′)	22.732	26	4	9.92E+08
SHFV ORF2a (GP2)	24.13	28	5	9.88E+08
SHFV ORF2a′ (GP2′)	31.792	48	11	5.48E+08
SHFV ORF2b′ (E′)	10.456	56.4	5	5.01E+08
SHFV ORF4′ (GP4′)	23.061	29.3	3	1.33E+08
SHFV ORF3 (GP3)	19.655	26.8	2	9.20E+07
ORF$_{10,071–10,331}$	9.4705	19.8	2	8.54E+06
SHFV nsp1α	17.898	93.3	13	4.11E+09
SHFV nsp3	24.016	42.3	7	2.38E+09
SHFV nsp1β	20.663	78	10	1.98E+09
SHFV nsp4	20.858	47.8	8	1.80E+09
SHFV nsp1γ	15.313	77.6	6	1.66E+09
SHFV nsp2	81.172	51.9	31	1.63E+09
SHFV nsp7	22.932	73.1	11	8.21E+08
SHFV nsp5	18.089	4.8	1	5.94E+08
SHFV nsp12	19.828	45.2	6	2.58E+08
SHFV nsp9	70.347	42.6	24	1.08E+08
SHFV nsp11	24.767	43.5	6	1.03E+08
SHFV nsp8	5.249	68	2	8.29E+07
SHFV nsp10	49.348	38	10	7.31E+07
SHFV nsp2-2 TF	68.896	37.1	2	3.08E+07
SHFV nsp2-1 TF	52.161	45.5	1	3.76E+06

*The structural proteins and then the nonstructural proteins are listed in the order of their relative abundance.

abundance than the previously published TRS. The major TRSs for those proteins were likely not identified in the previous study due to the limited number of leader–body junction clones analyzed (24). Evidence from previous studies suggested that the genomes of other nidoviruses also contain functional alternative body TRSs for their structural proteins. One or two alternative body TRSs producing sg mRNAs encoding the structural proteins GP3, GP4, or GP5 were identified by amplification and cloning of a few sg mRNA leader–body junctions for PRRSV and EAV (40, 45). Although the relative abundance of the EAV sg mRNAs produced from these alternative body TRSs was lower than that from the major TRS, the combined amount produced from them was sufficient to generate infectious progeny virus when the major TRS was inactivated (although the virus yield was reduced) (40). Coronavirus genomes have also been shown to be able to use functional alternative body TRSs when the major body TRS of a 3′ ORF is mutated (46–48). However, the finding that, for the majority of the SHFV proteins detected by mass spectrometry, the relative protein level correlated well with the total sg mRNA abundance produced from multiple TRSs strongly suggests that the alternative body TRSs are not just reserve back-ups that function only to ensure virus survival when the primary TRS becomes mutated. Instead, the use of all the multiple alternative body TRSs for each ORF during each replication cycle appears to be the mechanism used to fine tune the production of the optimal total amount of each viral protein, especially for those proteins that lack a major TRS with high transcription activity.

After the discovery of the E ORF in arterivirus genomes and the prediction of the additional E′ ORF in the SHFV genome, it was proposed that SHFV sg mRNA2′ and sg mRNA2 bicistronically express GP2′/E′ and E/GP2, respectively (10, 34). In the present study, sg mRNAs were identified that encode E′ as the 5′ ORF. The total sg mRNA abundance estimated for GP2′ was ~1.5 times higher than that for E′, which correlated with the relative protein amounts estimated by mass spectrometry indicating that GP2′ was ~1.1 times more abundant than E′. These data suggest that the majority of E′ is expressed from the 5′ ORF of multiple E′ sg mRNAs. Multiple sg mRNAs were also identified with GP2 as the 5′ ORF. However, although the total abundance estimated for the GP2 sg mRNAs was ~46 times lower than that for E, the protein abundance estimated for GP2 was equivalent to that for E. This finding suggests that, although sg mRNAs expressing GP2 as the 5′ ORF are produced at low levels, the majority of GP2 produced is translated as a second ORF from the E sg mRNAs. Monocistronic sg mRNAs have also subsequently been identified for some coronavirus proteins, such as the *Mouse hepatitis virus* (MHV) E protein, that were initially thought to be expressed bicistronically (41, 49, 50). However, an in vitro translation study showed that the MHV E protein can also be expressed from the second ORF of a larger sg mRNA (50). The SHFV sg mRNA5 was previously predicted to express both GP5 and ORF5a. In the present study, separate ORF5a sg mRNAs were identified at the very low abundance of 0.002%. The lack of detection of ORF5a peptides by the mass spectrometry analysis was consistent with the low sg mRNA abundance and suggested that it is unlikely that ORF5a is expressed bicistronically from the highly abundant GP5 sg mRNAs.

About a third of the body TRSs identified in the SHFV genome produced detectable levels of sg mRNAs with two or three variant leader–body junction sequences. The junction sequence of a sg mRNA is determined by the position at which the polymerase disassociates from the body TRS core sequence and by the location where the 3′ end of the nascent minus strand anneals to the leader TRS core sequence. Heterogeneity of the leader–body junction sequences generated from the same body TRS has also been observed for some coronavirus sg mRNAs (51, 52). Despite the complexity of SHFV sg mRNA production, the total abundance of sg mRNA generated at each body TRS was consistent at early and late times post infection in two different cell types. Although the sg mRNAs generated at the 3′ terminal TRS7 were by far the most abundant, a serial decrease in sg mRNA abundance with increasing TRS distance from the 3′ end was not observed. Instead, multiple regions of high transcriptional activity flanked by regions of low transcriptional activity were detected. Previous studies showed that increasing the duplex stability between leader–body TRSs by site-directed mutagenesis enhanced sg mRNA synthesis but that duplex stability was not the only determinant (33, 47, 53, 54). In the present study, a linear correlation between the sg mRNA abundance and the stability of the duplex formed between the leader TRS and each of the seven nearby body TRSs for ORF5-C-68aa was not observed. These results are consistent with previous observations for both coronavirus and arterivirus genomes indicating that the local RNA secondary structural context of the different body TRSs as well as long-distance RNA interactions in the viral genome play important roles in determining the efficiency of sg mRNA production (40, 54–57).

Surprisingly, multiple functional body TRSs were also identified in the ORF1a/1b region of the SHFV genome, indicating that sg RNA transcription is not restricted to the 3′ structural protein region. A previous study of the coronavirus MHV also identified three functional body TRSs in the ORF1a region (58). Most of the TRSs in the SHFV ORF1a/1b region produced long sg mRNAs encoding different lengths of in-frame, C-terminal portions of ORF1b (truncated ORF1b). A previous study on the arterivirus LDV also identified a body TRS in ORF1b that produced a long sg mRNA encoding a truncated ORF1b (200 aa) (59). ORF1b is translated as part of an ORF1a/1b polyprotein after a −1 frame shift occurs near the end of ORF1a and the 1b region is cleaved to produce nsp9 (RdRp), nsp10 (helicase), nsp11 (NendoU), and nsp12 (function unknown) (10). Except for nsp12-C, all the truncated SHFV ORF1b ORFs encode full-length nsp12. The longest truncated ORF1b ORF, ORF1b-C-1,021aa, encodes full-length nsp12, nsp11, and nsp10 as well as the C terminus of nsp9. The translation efficiency of EAV ORF1b was previously reported to be only 15–20% of that of ORF1a (38). The production of sg mRNAs encoding ORF1b proteins provides an alternative means for producing nsp10, nsp11, and especially nsp12. Consistent with this hypothesis, the mass spectrometry analysis detected three times more nsp12 (iBAQ intensity of 2.58E+08) than nsp8 (iBAQ intensity of 8.29E+07), which is cleaved from the 3′ end of the ORF1a polyprotein.

Studies on different coronaviruses identified additional functional body TRSs that produce sg mRNAs encoding accessory proteins or previously unidentified protein ORFs, some of which are virus-strain specific (41, 42, 51, 58, 60, 61). In the present study, additional functional body TRSs that produce sg mRNAs encoding six previously unidentified ORFs in an alternative reading frame were discovered. However, only the protein product expressed from $\mathrm{ORF}_{10,071–10,331}$, which is located within the ORF1b region, was detected by mass spectrometry analysis. This 86-aa product was the largest among those of the additional alternative reading frame ORFs and was predicted to be composed primarily of an α-helix, to have two transmembrane domains, and to contain a protein kinase C phosphorylation site (PredictProtein, https://www.predictprotein.org/). Five of the previously unreported body TRSs produced sg mRNAs encoding $\mathrm{ORF}_{15,419–15,505}$ (28 aa) from the +2 reading frame inside ORF7 (N), which is very similar to the recently identified small ORF7a of both type I and type II PRRSV that is also translated from the +2 reading frame inside ORF7 (N) (62). An alternative ORF inside the *N* gene encoding an accessary protein has also been identified in the genomes of many beta coronaviruses (63–65). The lack of detection of peptides from $\mathrm{ORF}_{15,419–15,505}$ and the other SHFV alternative frame ORFs by mass spectrometry could be due to small peptide size, low abundance, the location of trypsin cleavage sites, and/or posttranslational modifications preventing trypsin cleavage.

An additional functional arterivirus ORF1a frameshift site that primarily produces the −2 frameshift product nsp2TF was previously identified in the nsp2 region of the PRRSV genome. This site was predicted to be conserved in the SHFV LVR nsp2 region, and a −2 frameshift would produce a fusion protein consisting of the N-terminal two-thirds of nsp2 fused to an alternative 225-aa extension (39). Some −1 frameshift activity at the site in the PRRSV genome was also detected and produced a fusion protein (nsp2N) with a very short extension due to an immediate stop codon. Unlike PRRSV, the −1 frameshift in the SHFV LVR nsp2 produces a fusion protein with an alternative 77-aa extension. Peptides mapping to the unique extensions of the −1 or −2 nsp2 frameshift products were detected in SHFV-infected MA104 cells by mass spectrometry. Similar to the data obtained for PRRSV, the SHFV nsp2 −2 frameshift was more efficient than the −1 frameshift. The intracellular locations and biological functions of the SHFV nsp2 frameshift products are not known.

The detection of functional body TRSs in the SHFV 5′ ORF1a/1b region suggests that nidovirus sg RNA transcription can occur within the 5′ region as well as within the 3′ region of the genome. The SHFV ORF1b sg mRNAs identified provide an alternative means of increasing the abundance of the ORF1b nonstructural proteins, especially nsp11 and nsp12. The identification of multiple functional body TRSs for all but four of the SHFV 3′ ORFs and the consistency of sg mRNA abundance produced from individual TRSs at different times during the infection cycle in two different cell types indicate that nidovirus transcription is accurately regulated and more complex than previously appreciated. The alternating regions of high and low transcription activity detected across the 3′ region of the SHFV genome strongly suggest a role for local secondary and genomic higher-order RNA structures in regulating nidovirus transcription. The discovery of many previously unreported in-frame C-terminal ORFs and alternative frame ORFs encoded by newly identified sg mRNAs indicates that the coding capacity of the SHFV genome and very likely that of other nidovirus genomes was previously underestimated. In fact, a recent gene-expression study of the coronavirus MHV that used ribosome profiling and RNA sequencing identified 15 previously unreported sg mRNAs with unique junction sequences at 5 hpi and detected heterogeneous leader–body junction sequences generated from the same body TRS (66). Translation initiation was also detected at internal AUG codons in the MHV ORF5 by Ribo-Seq. In addition, translation of several small, previously unidentified ORFs either upstream of or embedded within known viral protein ORFs was detected that was often initiated from a noncanonical start codon, such as CUG. These data predict an even bigger coding capacity for nidoviruses. The small MHV ORFs were proposed to regulate the translation of the downstream ORF in the same sg mRNA. Some of the SHFV small C-terminal ORFs and alternative frame ORFs identified in the present study may also function as regulators of a downstream ORF. Alternatively, increasing evidence from different organisms suggests that small proteins, as short as 11 aa, translated from in-frame or alternative frame ORFs are functional and are involved in regulating a variety of cellular processes (67, 68).

Materials and Methods

Methods and sources of reagents can be found in *SI Materials and Methods*.

ACKNOWLEDGMENTS. We thank B. Stockman and F. Ede for technical assistance and M. Basu for help with the final figure files. The research was supported by Public Health Service Research Grants AI073824 (to M.A.B.), National Institute of Allergy and Infectious Diseases, NIH Grant U19 AI107810 (to R.S.B.), and by a Georgia State University Molecular Basis of Disease Seed Grant (to H.D.). H.D. was supported by a Georgia State University Molecular Basis of Disease Fellowship.

1. Pasternak AO, Spaan WJ, Snijder EJ (2006) Nidovirus transcription: How to make sense...? *J Gen Virol* 87:1403–1421.
2. Gorbalenya AE, Enjuanes L, Ziebuhr J, Snijder EJ (2006) Nidovirales: Evolving the largest RNA virus genome. *Virus Res* 117:17–37.
3. Zirkel F, et al. (2013) Identification and characterization of genetically divergent members of the newly established family Mesoniviridae. *J Virol* 87:6346–6358.
4. London WT (1977) Epizootiology, transmission and approach to prevention of fatal simian haemorrhagic fever in rhesus monkeys. *Nature* 268:344–345.
5. Lauck M, et al. (2013) Exceptional simian hemorrhagic fever virus diversity in a wild African primate community. *J Virol* 87:688–691.
6. Vatter HA, et al. (2015) A simian hemorrhagic fever virus isolate from persistently infected baboons efficiently induces hemorrhagic fever disease in Japanese macaques. *Virology* 474:186–198.
7. Bailey AL, et al. (2014) Two novel simian arteriviruses in captive and wild baboons (Papio spp.). *J Virol* 88:13231–13239.
8. Allen AM, Palmer AE, Tauraso NM, Shelokov A (1968) Simian hemorrhagic fever. II. Studies in pathology. *Am J Trop Med Hyg* 17:413–421.
9. Brinton MA, Di H, Vatter HA (2015) Simian hemorrhagic fever virus: Recent advances. *Virus Res* 202:112–119.
10. Snijder EJ, Kikkert M, Fang Y (2013) Arterivirus molecular biology and pathogenesis. *J Gen Virol* 94:2141–2163.
11. Snijder EJ, Meulenberg JJM (1998) The molecular biology of arteriviruses. *J Gen Virol* 79:961–979.
12. de Vries AA, Post SM, Raamsman MJ, Horzinek MC, Rottier PJ (1995) The two major envelope proteins of equine arteritis virus associate into disulfide-linked heterodimers. *J Virol* 69:4668–4674.
13. Wieringa R, et al. (2004) Structural protein requirements in equine arteritis virus assembly. *J Virol* 78:13019–13027.
14. Snijder EJ, Dobbe JC, Spaan WJ (2003) Heterodimerization of the two major envelope proteins is essential for arterivirus infectivity. *J Virol* 77:97–104.
15. Dea S, Gagnon CA, Mardassi H, Pirzadeh B, Rogan D (2000) Current knowledge on the structural proteins of porcine reproductive and respiratory syndrome (PRRS) virus: Comparison of the North American and European isolates. *Arch Virol* 145:659–688.
16. Doan DN, Dokland T (2003) Structure of the nucleocapsid protein of porcine reproductive and respiratory syndrome virus. *Structure* 11:1445–1451.
17. Dokland T (2010) The structural biology of PRRSV. *Virus Res* 154:86–97.
18. Tian D, et al. (2012) Arterivirus minor envelope proteins are a major determinant of viral tropism in cell culture. *J Virol* 86:3701–3712.
19. Das PB, et al. (2010) The minor envelope glycoproteins GP2a and GP4 of porcine reproductive and respiratory syndrome virus interact with the receptor CD163. *J Virol* 84:1731–1740.
20. Lee C, Yoo D (2006) The small envelope protein of porcine reproductive and respiratory syndrome virus possesses ion channel protein-like properties. *Virology* 355:30–43.
21. Firth AE, et al. (2011) Discovery of a small arterivirus gene that overlaps the GP5 coding sequence and is important for virus production. *J Gen Virol* 92:1097–1106.
22. Johnson CR, Griggs TF, Gnanandarajah J, Murtaugh MP (2011) Novel structural protein in porcine reproductive and respiratory syndrome virus encoded by an alternative ORF5 present in all arteriviruses. *J Gen Virol* 92:1107–1116.
23. Vatter HA, et al. (2014) Functional analyses of the three simian hemorrhagic fever virus nonstructural protein 1 papain-like proteases. *J Virol* 88:9129–9140.
24. Godeny EK, de Vries AA, Wang XC, Smith SL, de Groot RJ (1998) Identification of the leader-body junctions for the viral subgenomic mRNAs and organization of the simian hemorrhagic fever virus genome: Evidence for gene duplication during arterivirus evolution. *J Virol* 72:862–867.
25. Vatter HA, Di H, Donaldson EF, Baric RS, Brinton MA (2014) Each of the eight simian hemorrhagic fever virus minor structural proteins is functionally important. *Virology* 462–463:351–362, and erratum (2014) 464–465:461.
26. Smits SL, et al. (2005) Torovirus non-discontinuous transcription: Mutational analysis of a subgenomic mRNA promoter. *J Virol* 79:8275–8281.
27. Sawicki SG, Sawicki DL, Siddell SG (2007) A contemporary view of coronavirus transcription. *J Virol* 81:20–29.
28. Van Den Born E, Gultyaev AP, Snijder EJ (2004) Secondary structure and function of the 5′-proximal region of the equine arteritis virus RNA genome. *RNA* 10:424–437.
29. van den Born E, Posthuma CC, Gultyaev AP, Snijder EJ (2005) Discontinuous subgenomic RNA synthesis in arteriviruses is guided by an RNA hairpin structure located in the genomic leader region. *J Virol* 79:6312–6324.
30. de Vries AA, et al. (1990) All subgenomic mRNAs of equine arteritis virus contain a common leader sequence. *Nucleic Acids Res* 18:3241–3247.
31. van Marle G, et al. (1999) Arterivirus discontinuous mRNA transcription is guided by base pairing between sense and antisense transcription-regulating sequences. *Proc Natl Acad Sci USA* 96:12056–12061.
32. den Boon JA, Kleijnen MF, Spaan WJ, Snijder EJ (1996) Equine arteritis virus subgenomic mRNA synthesis: Analysis of leader-body junctions and replicative-form RNAs. *J Virol* 70:4291–4298.
33. Pasternak AO, van den Born E, Spaan WJ, Snijder EJ (2001) Sequence requirements for RNA strand transfer during nidovirus discontinuous subgenomic RNA synthesis. *EMBO J* 20:7220–7228.
34. Snijder EJ, van Tol H, Pedersen KW, Raamsman MJ, de Vries AA (1999) Identification of a novel structural protein of arteriviruses. *J Virol* 73:6335–6345.
35. Zeng L, Godeny EK, Methven SL, Brinton MA (1995) Analysis of simian hemorrhagic fever virus (SHFV) subgenomic RNAs, junction sequences, and 5′ leader. *Virology* 207:543–548.
36. Smith SL, Wang X, Godeny EK (1997) Sequence of the 3′ end of the simian hemorrhagic fever virus genome. *Gene* 191:205–210.
37. Godeny EK, Zeng L, Smith SL, Brinton MA (1995) Molecular characterization of the 3′ terminus of the simian hemorrhagic fever virus genome. *J Virol* 69:2679–2683.
38. den Boon JA, et al. (1991) Equine arteritis virus is not a togavirus but belongs to the coronaviruslike superfamily. *J Virol* 65:2910–2920.
39. Fang Y, et al. (2012) Efficient -2 frameshifting by mammalian ribosomes to synthesize an additional arterivirus protein. *Proc Natl Acad Sci USA* 109:E2920–E2928.

40. Pasternak AO, Gultyaev AP, Spaan WJ, Snijder EJ (2000) Genetic manipulation of arterivirus alternative mRNA leader-body junction sites reveals tight regulation of structural protein expression. *J Virol* 74:11642–11653.
41. Hussain S, et al. (2005) Identification of novel subgenomic RNAs and noncanonical transcription initiation signals of severe acute respiratory syndrome coronavirus. *J Virol* 79:5288–5295.
42. Bentley K, Keep SM, Armesto M, Britton P (2013) Identification of a noncanonically transcribed subgenomic mRNA of infectious bronchitis virus and other gammacoronaviruses. *J Virol* 87:2128–2136.
43. Vaughn EM, Halbur PG, Paul PS (1995) Sequence comparison of porcine respiratory coronavirus isolates reveals heterogeneity in the S, 3, and 3-1 genes. *J Virol* 69:3176–3184.
44. Callebaut P, Correa I, Pensaert M, Jiménez G, Enjuanes L (1988) Antigenic differentiation between transmissible gastroenteritis virus of swine and a related porcine respiratory coronavirus. *J Gen Virol* 69:1725–1730.
45. Lin YC, Chang RY, Chueh LL (2002) 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.
46. Ozdarendeli A, et al. (2001) Downstream sequences influence the choice between a naturally occurring noncanonical and closely positioned upstream canonical heptameric fusion motif during bovine coronavirus subgenomic mRNA synthesis. *J Virol* 75: 7362–7374.
47. Zúñiga S, Sola I, Alonso S, Enjuanes L (2004) Sequence motifs involved in the regulation of discontinuous coronavirus subgenomic RNA synthesis. *J Virol* 78:980–994.
48. Schelle B, Karl N, Ludewig B, Siddell SG, Thiel V (2005) Selective replication of coronavirus genomes that express nucleocapsid protein. *J Virol* 79:6620–6630.
49. O'Connor JB, Brian DA (2000) Downstream ribosomal entry for translation of coronavirus TGEV gene 3b. *Virology* 269:172–182.
50. Zhang X, Liu R (2000) Identification of a noncanonical signal for transcription of a novel subgenomic mRNA of mouse hepatitis virus: Implication for the mechanism of coronavirus RNA transcription. *Virology* 278:75–85.
51. Schaad MC, Baric RS (1993) Evidence for new transcriptional units encoded at the 3′ end of the mouse hepatitis virus genome. *Virology* 196:190–198.
52. Makino S, Soe LH, Shieh CK, Lai MM (1988) Discontinuous transcription generates heterogeneity at the leader fusion sites of coronavirus mRNAs. *J Virol* 62:3870–3873.
53. Pasternak AO, van den Born E, Spaan WJ, Snijder EJ (2003) The stability of the duplex between sense and antisense transcription-regulating sequences is a crucial factor in arterivirus subgenomic mRNA synthesis. *J Virol* 77:1175–1183.
54. Sola I, Moreno JL, Zúñiga S, Alonso S, Enjuanes L (2005) Role of nucleotides immediately flanking the transcription-regulating sequence core in coronavirus subgenomic mRNA synthesis. *J Virol* 79:2506–2516.
55. Pasternak AO, Spaan WJ, Snijder EJ (2004) Regulation of relative abundance of arterivirus subgenomic mRNAs. *J Virol* 78:8102–8113.
56. Moreno JL, Zúñiga S, Enjuanes L, Sola I (2008) Identification of a coronavirus transcription enhancer. *J Virol* 82:3882–3893.
57. Mateos-Gomez PA, Morales L, Zuñiga S, Enjuanes L, Sola I (2013) Long-distance RNA-RNA interactions in the coronavirus genome form high-order structures promoting discontinuous RNA synthesis during transcription. *J Virol* 87:177–186.
58. La Monica N, Yokomori K, Lai MM (1992) Coronavirus mRNA synthesis: Identification of novel transcription initiation signals which are differentially regulated by different leader sequences. *Virology* 188:402–407.
59. Chen Z, et al. (1993) 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.
60. Shieh CK, et al. (1989) Identification of a new transcriptional initiation site and the corresponding functional gene 2b in the murine coronavirus RNA genome. *J Virol* 63: 3729–3736.
61. Yokomori K, Banner LR, Lai MM (1991) Heterogeneity of gene expression of the hemagglutinin-esterase (HE) protein of murine coronaviruses. *Virology* 183:647–657.
62. Olasz F, et al. (2016) Immunological and biochemical characterisation of 7ap, a short protein translated from an alternative frame of ORF7 of PRRSV. *Acta Vet Hung* 64: 273–287.
63. Fischer F, Peng D, Hingley ST, Weiss SR, Masters PS (1997) The internal open reading frame within the nucleocapsid gene of mouse hepatitis virus encodes a structural protein that is not essential for viral replication. *J Virol* 71:996–1003.
64. Meier C, et al. (2006) The crystal structure of ORF-9b, a lipid binding protein from the SARS coronavirus. *Structure* 14:1157–1165.
65. Liu DX, Fung TS, Chong KK, Shukla A, Hilgenfeld R (2014) Accessory proteins of SARS-CoV and other coronaviruses. *Antiviral Res* 109:97–109.
66. Irigoyen N, et al. (2016) High-resolution analysis of coronavirus gene expression by RNA sequencing and ribosome profiling. *PLoS Pathog* 12:e1005473.
67. Pueyo JI, Magny EG, Couso JP (2016) New peptides under the s(ORF)ace of the genome. *Trends Biochem Sci* 41:665–678.
68. Landry CR, Zhong X, Nielly-Thibault L, Roucou X (2015) Found in translation: Functions and evolution of a recently discovered alternative proteome. *Curr Opin Struct Biol* 32:74–80.
69. Cox J, Mann M (2008) MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. *Nat Biotechnol* 26:1367–1372.
70. Schwanhäusser B, et al. (2011) Global quantification of mammalian gene expression control. *Nature* 473:337–342.