Supporting Information

Fang et al. 10.1073/pnas.1211145109

SI Materials and Methods

GenBank Accession Numbers of Sequences Used in the Bioinformatic Analysis. *Porcine reproductive and respiratory syndrome virus (European genotype, type I).* M96262, AY366525, AY375474, AY588319, DQ489311, DQ864705, EU076704, FJ349261, GQ461593, GU047344, GU047345, GU067771, GU737264, JF276430, JF276431, JF276432, JF276433, JF276434, JF276435, JF802085. *Porcine reproductive and respiratory syndrome virus (North American genotype, type II).* NC_001961 = AF046869, AB288356, AF066183, AF159149, AF176348, AF184212, AF303354, AF303355, AF303356, AF303357, AF325691, AF331831, AF494042, AY032626, AY150312, AY150564, AY262352, AY424271, AY457635, AY545985, AY585241, AY612613, DQ056373, DQ176019, DQ176020, DQ176021, DQ217415, DQ459471, DQ473474, DQ779791, DQ988080, EF075945, EF112445, EF112446, EF112447, EF153486, EF484031, EF484033, EF488048, EF488739, EF517962, EF532801, EF532802, EF532803, EF532804, EF532805, EF532806, EF532807, EF532808, EF532809, EF532810, EF532811, EF532812, EF532813, EF532814, EF532815, EF532816, EF532817, EF532818, EF532819, EF535999, EF536000, EF536001, EF536002, EF536003, EF635006, EF641008, EU097706, EU097707, EU106888, EU109502, EU109503, EU144079, EU187484, EU200961, EU200962, EU236259, EU262603, EU360128, EU360129, EU360130, EU624117, EU678352, EU708726, EU807840, EU825723, EU825724, EU860248, EU860249, EU864231, EU864232, EU864233, EU880431, EU880432, EU880433, EU880434, EU880435, EU880436, EU880437, EU880438, EU880439, EU880440, EU880441, EU880442, EU880443, EU939312, FJ175687, FJ175688, FJ175689, FJ393456, FJ393457, FJ393458, FJ393459, FJ394029, FJ536165, FJ548851, FJ548852, FJ548853, FJ548854, FJ548855, FJ797690, FJ889129, FJ895329, FJ899592, GQ330474, GQ351601, GQ359108, GQ374441, GQ374442, GQ475526, GQ499193, GQ499194, GQ499195, GQ499196, GQ857656, GU143913, GU168567, GU168568, GU168569, GU169411, GU232735, GU232736, GU232737, GU232738, GU269541, GU454850, GU461292, HM011104, HM016158, HM016159, HM189676, HM214913, HM214914, HM214915, HM853673, HQ233604, HQ233605, HQ315835, HQ315836, HQ315837, HQ401282, HQ416720, HQ699067, HQ843178, HQ843179, HQ843180, HQ843181, JF268672, JF268673, JF268674, JF268675, JF268676, JF268677, JF268678, JF268679, JF268680, JF268681, JF268682, JF268683, JF268684, JF748717, JF748718, JF796180, JF800911, JN256115, JN387271, JN387272, JN387273, JN387274, JN626287, JN662424, U87392.
Lactate dehydrogenase-elevating virus. NC_001639 = U15146, L13298.
Simian hemorrhagic fever virus. NC_003092 = AF180391, HQ845737, HQ845738.
Equine arteritis virus. NC_002532 = X53459, AY349167, AY349168, DQ846750, EU252113, EU252114, EU586273, EU586274, EU586275, GQ903794, GQ903795, GQ903796, GQ903797, GQ903798, GQ903799, GQ903800, GQ903801, GQ903802, GQ903803, GQ903804, GQ903805, GQ903806, GQ903807, GQ903808, GQ903809, GQ903810, GQ903811, JN211316, JN211317, JN211318, JN211319, JN211320.

Statistical Significance of the Long ORF in Simian Hemorrhagic Fever Virus. The statistical significance of the conserved presence of the long TF ORF (Fig. 1*C*) in the three highly divergent simian hemorrhagic fever virus (SHFV) sequences was evaluated by randomly shuffling ORF1a-frame codon columns within the TF ORF region and calculating what fraction of shuffled alignments preserve an ORF in the +1 frame. This procedure controls for any bias for or against random long +1 frame ORFs resulting from ORF1a-frame amino acid use, codon use, or nucleotide biases and also controls for phylogenetic nonindependence. In fact, the proportion of randomizations that preserve a +1 frame ORF was too small to estimate directly (0 occurrences in 4,000 randomizations) and so was estimated instead from the mean number (per randomized alignment) of +1 frame alignment codon columns containing stop codons in one or more sequences (*viz*., 24.70), assuming Poisson statistics. Using this method, the *P* value for such a long +1 frame ORF occurring by chance in this region of the SHFV alignment is 1.9×10^{-11}. Neither this statistic for SHFV (1.9×10^{-11}) nor the conservation statistic quoted in the main text for porcine reproductive and respiratory syndrome virus (PRRSV) (1×10^{-64}) has been corrected for multiple tests. In principle one might consider testing the whole genome (~15,000 nt) for conserved regions and/or conserved ORFs of ~200 codons, and, in principle, one might apply such an analysis to the ~1,000 RNA virus species represented in GenBank, making a total of ~25,000 independent tests. Thus, the *P* values should be scaled by ~25,000 (giving 2.5×10^{-60} and 4.8×10^{-7}, respectively), although a correction for multiple testing is not, in fact, required for the SHFV statistic, because the location of the 5′ end of the TF ORF in SHFV is known a priori (it aligns to the 5′ end of the TF ORF in PRRSV).

DNA Constructs. Plasmid pL1a was a derivative of a similar equine arteritis virus (EAV) ORF1a expression vector, in which the foreign gene is under the control of a T7 RNA polymerase promoter and an encephalomyocarditis virus internal ribosomal entry site and is followed by a downstream T7 terminator sequence (1). The EAV ORF1a sequence was replaced by PRRSV SD01-08 ORF1a (nucleotides 1–7137 of the genome). To construct the nsp2TF knockout mutants (Fig. 2*B*), mutations were introduced into the nsp2 region of the expression vector pL1a or a PRRSV full-length cDNA infectious clone plasmid pSD01-08 (2). Except for the KO2 mutant, for which a synthesized oligonucleotide was used, all mutations were introduced by site-directed mutagenesis using the Quick-Change site-directed PCR mutagenesis kit (Stratagene). The pCAGGS-nsp2 and pCAGGS-nsp2TF plasmids were constructed by PCR amplification of the nsp2- and nsp2TF-encoding regions from plasmid DNA of pL1a and pL1a-IFC, respectively, and cloning of PCR products into a eukaryotic expression vector, pCAGGS (3).

Antibodies. For type I PRRSV SD01-08, anti-nsp2 mAbs 36-19 and 58-46 were produced by immunizing mice with amino acids 386–821 of pp1a. Anti-nsp4 mAb 54-19 was produced by immunizing mice with amino acids 1677–1879 of pp1a (4). For type II PRRSV SD23983, anti-nsp2 mAbs 140-68 and 148-43 were produced by immunizing mice with amino acids 435–514 of SD23983 pp1a. A polyclonal affinity-purified rabbit Ab, pAb-TF, was produced by GenScript using the synthetic peptide CPKGVVTSVGESV (C-terminal 13 amino acids of nsp2TF). Abs for detection of cellular marker proteins comprised a pAb against protein disulfide isomerase (PDI) (Enzo), anti-giantin mAb G1/93 (Alexis), and anti-ERGIC-p53 mAb G1/133 (Alexis).

Immunofluorescence Microscopy. MARC-145 cells were infected with PRRSV [multiplicity of infection (MOI) 0.1] or were transfected with pCAGGS-nsp2 or pCAGGS-nsp2TF. At 18–24 h postinfection (p.i.) or 24–48 h posttransfection (p.t.) cells were

fixed with paraformaldehyde, and single or double immunolabeling and immunofluorescence microscopy were performed essentially as described previously (4, 5).

Immunoprecipitation and SDS/PAGE. Whole-cell lysates of PRRSV-infected MARC-145 cells were suspended in RIPA buffer [0.5% (wt/vol) sodium deoxycholate, 1% (wt/vol) SDS, 1% (vol/vol) Nonidet P-40, 100 mM NaCl, 1 mM EDTA (pH 8.0), 10 mM Tris·HCl (pH 7.5)]. To reduce nonspecific background, cell lysates were precleared with preimmune rabbit sera or non-specific mouse ascites. Protein A-Sepharose CL-4B beads (Pharmacia Biotech) and an nsp2-specific mAb were added to precleared cell lysates. After incubating overnight at 4 °C, immune complexes were washed three times with RIP buffer (10 mM Tris·HCl, 150 mM NaCl, 1% Nonidet P-40, 0.5% sodium deoxycholate) and three times with deionized H_2O. After boiling in 2× Laemmli sample buffer for 5 min, proteins were separated on a 6% SDS/PAGE gel.

Western Blot. Western blot was performed as described previously (4). The membrane was probed with primary nsp2- and/or nsp2TF-specific Abs. IRDye 680-conjugated goat anti-rabbit Ab and/or IRDye 800CW-conjugated goat anti-mouse Ab (LI-COR Biosciences) were used as secondary Abs. Imaging of the blot was performed using an Odyssey infrared imaging system (LI-COR Biosciences).

Radioactive Labeling and Radioimmunoprecipitation Analysis. To analyze PRRSV SD01-08 nsp2TF expression in infected MARC-145 cells (MOI 0.1), cells were starved for 30 min in Met- and Cys-free medium, and at 24 h p.i. proteins were pulse-labeled for 1 h with 500 μCi/mL of a [^{35}S]Met/Cys mixture (EXPRE^{35}S^{35}S Protein Labeling Mix; Perkin-Elmer). Following two PBS washes, cells were incubated for various chase periods up to 24 h in DMEM containing 2 mM each of unlabeled Met and Cys and 2% FCS. Transient PRRSV ORF1a expression in RK-13 cells, using plasmid pL1a and the recombinant vaccinia virus/T7 polymerase expression system, was performed as described previously (1). Pulse-chase experiments were performed as described above at 4 h p.t. using a 30-min pulse of 500 μCi of [^{35}S]Met/Cys per mL. Protocols for cell lysis, immunoprecipitation, SDS/PAGE, and imaging (Typhoon Variable Mode Imager; GE Healthcare) have been described previously (4).

Calculation of Frameshifting Efficiencies. Band intensities (nsp2, nsp2′, nsp2TF, nsp2TF′, and nsp2N) were quantified with ImageQuant TL (GE Healthcare) and normalized by the Met+Cys content of the respective products assuming that ^{35}S Met and ^{35}S Cys are incorporated with an efficiency ratio of 73:22 (the Met:Cys ratio in the mixture according to the manufacturer's documentation). (Note that calculated frameshifting efficiencies were only 1.06–1.07 times higher if equal incorporation efficiencies were assumed instead.) Radioactivity in the two bands at around 200 kDa was quantified and normalized assuming that these products represent nsp2–8; the product migrating at around 98 kDa was assumed to have Met and Cys content similar to that of nsp2N; and the two products migrating at around 90 kDa were assumed, based on migration position, to have half as many Met and Cys residues as nsp2. Using these values (Table S1), frameshifting efficiencies were calculated as (nsp2TF + nsp2TF′)/(nsp2 + nsp2′ + nsp2TF + nsp2TF′ + nsp2N) (upper bound) and (nsp2TF + nsp2TF′)/(nsp2 + nsp2′ + nsp2TF + nsp2TF′ + nsp2N + 200K products + 98K product + 90K products) (lower bound) for −2 frameshifting, and similarly, using nsp2N as the dividend, for −1 frameshifting.

1. Snijder EJ, Wassenaar AL, Spaan WJ (1994) Proteolytic processing of the replicase ORF1a protein of equine arteritis virus. *J Virol* 68:5755–5764.
2. Fang Y, et al. (2006) A full-length cDNA infectious clone of North American type 1 porcine reproductive and respiratory syndrome virus: Expression of green fluorescent protein in the Nsp2 region. *J Virol* 80:11447–11455.
3. Niwa H, Yamamura K, Miyazaki J (1991) Efficient selection for high-expression transfectants with a novel eukaryotic vector. *Gene* 108:193–199.
4. Li Y, Tas A, Snijder EJ, Fang Y (2012) Identification of porcine reproductive and respiratory syndrome virus ORF1a-encoded non-structural proteins in virus-infected cells. *J Gen Virol* 93:829–839.
5. Knoops K, et al. (2012) Ultrastructural characterization of arterivirus replication structures: Reshaping the endoplasmic reticulum to accommodate viral RNA synthesis. *J Virol* 86:2474–2487.

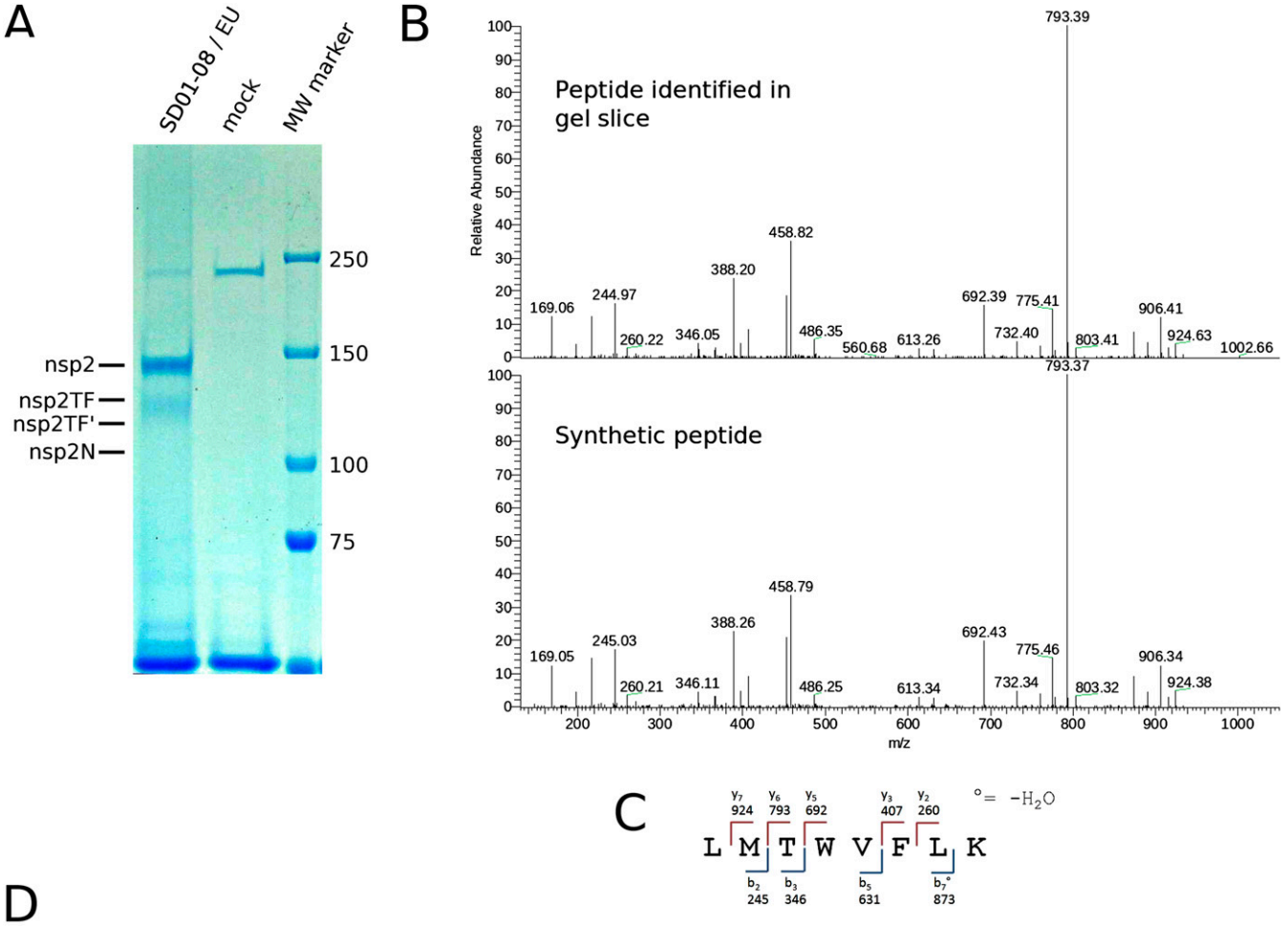

Fig. S1. Mass spectrometric analysis of nsp2TF and nsp2N purified from cells infected with type I PRRSV isolate SD01-08. (*A*) PRRSV-infected or mock-infected MARC-145 cell lysates were immunoprecipitated with nsp2-specific mAb36-19. Immunoprecipitated proteins were separated by SDS/PAGE and stained with Coomassie Blue. Positions of molecular mass markers and putative PRRSV proteins are indicated. (*B*) Fragmentation spectra of the shift-site peptide LMTW**VF**LK (shift site-encoded amino acids in bold) identified from the gel slice containing the band labeled nsp2TF (*Upper*) and the synthetic peptide (*Lower*). (*C*) Peptide sequence of the nsp2TF shift-site peptide. The fragment ions that were identified in the LC-MS/MS analysis of the gel slice are indicated. (*D*) Complete amino acid sequence of the predicted −1 frameshift product, nsp2N (see main text for the corresponding figure, Fig. 4*B*, for nsp2TF). Peptides identified by mass spectrometry of the gel slice containing the band labeled nsp2N are in red. The peptide compatible with −1 frameshifting and termination at the −1 frame stop codon is underlined in green.

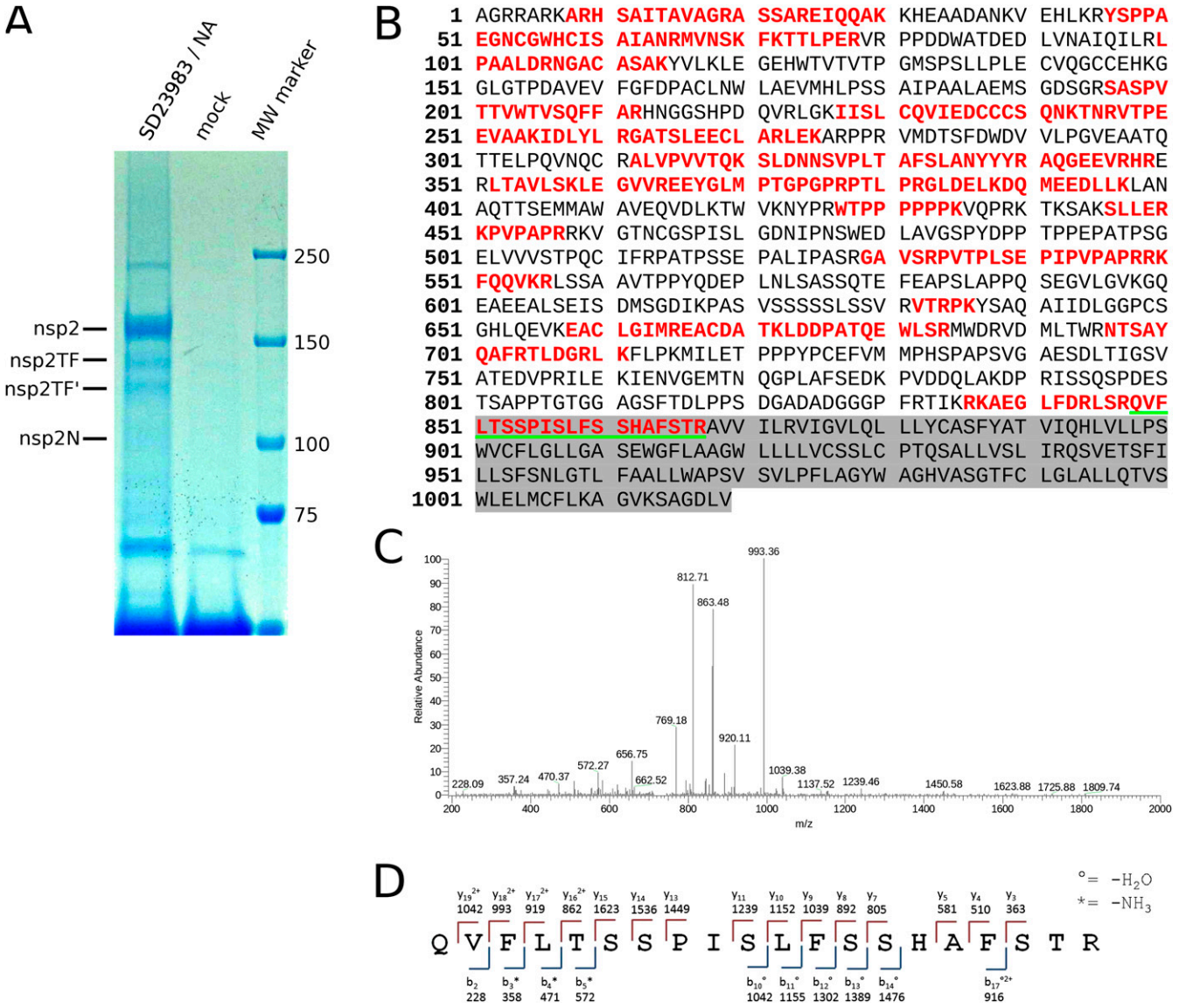

Fig. S2. Mass spectrometric analysis of nsp2TF purified from cells infected with type II PRRSV isolate SD23983. (*A*) PRRSV-infected or mock-infected MARC-145 cell lysates were immunoprecipitated with nsp2-specific mAb140-68. Immunoprecipitated proteins were separated by SDS/PAGE and stained with Coomassie Blue. Positions of molecular mass markers and putative PRRSV proteins are indicated. (*B*) Complete amino acid sequence of nsp2TF. Peptides identified by mass spectrometry are in red. The C-terminal 169 amino acids encoded by the +1 reading frame are highlighted in gray. The N-terminal 850 amino acids are shared with nsp2. The peptide spanning the frameshift site is underlined in green. (*C*) Fragmentation spectrum of the shift-site peptide Q**VF**LTSSPISLFSSHAFSTR (shift site-encoded amino acids in bold). (*D*) Peptide sequence of the nsp2TF shift-site peptide. The fragment ions that were identified in the LC-MS/MS analysis of the gel slice are indicated.

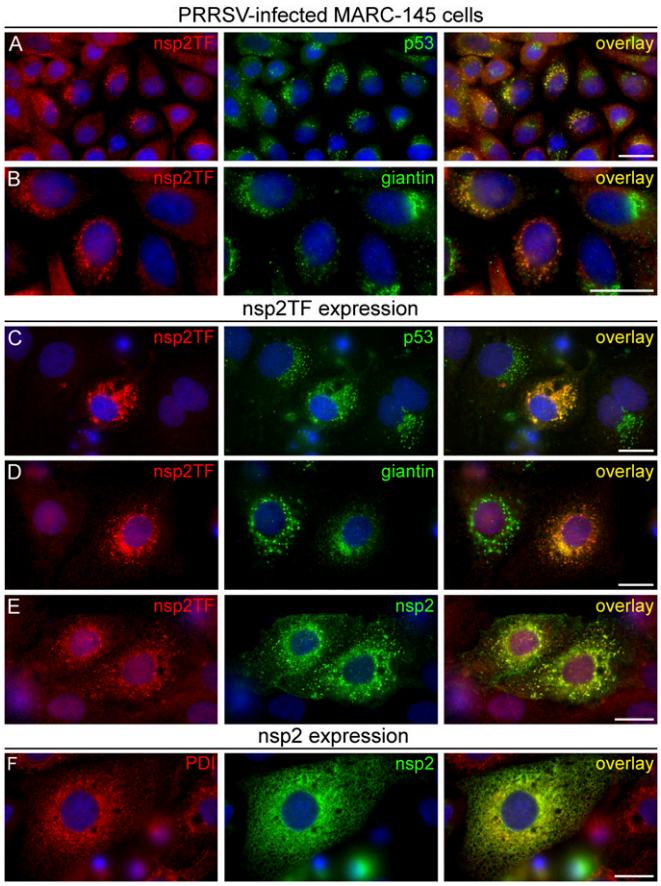

Fig. S3. IF microscopy analysis of the subcellular localization of nsp2TF in PRRSV-infected MARC-145 cells at 24 h p.i. (*A* and *B*) and a CMV promoter-driven expression system for nsp2TF (*C*–*E*) and nsp2 (*F*). (*A*) Infected cells were double labeled with pAb-TF and an mAb recognizing the ERGIC marker p53, revealing a substantial overlap between the two labeling patterns. (*B*) Double labeling with pAb-TF and an mAb directed against the Golgi marker giantin. Partial colocalization was observed, and the two labelings seemed to influence each other, because cells with a strong nsp2TF signal showed a weaker giantin signal (e.g., the cell in the middle). (*C*) Nsp2TF-expressing MARC-145 cells double labeled with pAb-TF and the anti-p53 mAb. As in infected cells, a considerable overlap was observed. (*D*) Double labeling with pAb-TF and the anti-giantin mAb, again showing partial colocalization. (*E*) Double labeling with pAb-TF and anti-nsp2 mAb 58-46 showing identical labeling patterns, indicating that mAb 58-46 does recognize nsp2TF in nsp2TF-expressing MARC-145 cells, in contrast to observations in PRRSV-infected cells (Fig. 7). (*F*) Nsp2-expressing MARC-145 cells double labeled for the endoplasmic reticulum marker PDI (pAb) and nsp2 (mAb58-46), showing partial colocalization. Cell nuclei were stained with Hoechst 33258. (Scale bars, 20 nm.)

Table S1. Radioactive incorporation into nsp2-related products in virus-infected cells

Band	% radioactive label	Adjusted for Met/Cys* (%s)	Adjusted for Met/Cys† (%s)
200K (upper)	2.4	1.1	1.0
200K (lower)	0.5	0.3	0.2
nsp2	59.1	50.7	51.8
nsp2′	5.2	4.5	4.6
nsp2TF	13.3	13.8	13.0
nsp2TF′	2.8	2.9	2.8
nsp2N	4.4	6.0	5.6
98K	1.2	1.6	1.5
90K (upper)	1.9	3.3	3.3
90K (lower)	9.2	15.8	16.1

*Assuming equal incorporation efficiencies for ^{35}S Met and ^{35}S Cys. Met and Cys content for the unidentified nsp2-related products migrating at around 200K, 98K, and 90K are estimated as described in *SI Materials and Methods*.
†Assuming ^{35}S Met and ^{35}S Cys are incorporated with an efficiency ratio of 73:22.