Supplementary Materials for

Decoding human cytomegalovirus

Noam Stern-Ginossar, Ben Weisburd, Annette Michalski, Vu Thuy Khanh Le, Marco Y. Hein[2], Sheng-Xiong Huang, Ming Ma, Ben Shen, Shu-Bing Qian, Hartmut Hengel, Matthias Mann, Nicholas T. Ingolia, Jonathan S. Weissman.
correspondence to: A.M. (michalsk@biochem.mpg.de) J.S.W. (weissman@cmp.ucsf.edu)

This PDF file includes:

Other Supplementary Materials for this manuscript includes the following:

Materials and Methods

Cells and virus

Human fibroblasts (CRL-1634) and the HCMV Merlin strain (VR-1590) were obtained from American Type Culture Collection (ATCC). The virus was propagated twice on HFF cells before the preparations of samples for sequencing. We chose the Merlin strain because it has been reported to contain the complete set of HCMV ORFs, only two of which (RL13 and UL128) are mutated (*1-3*). Cells were grown on 15cm plates and were infected at a multiplicity of infection (MOI) of 5. To achieve maximum synchronization cells were infected for 1 hour, then the medium containing the virus was removed and fresh medium was added.

Ribosome profiling samples preparation

Cylcoheximide and harringtonine drug treatments were carried out as previously described(*4*). Cells were lysed in lysis buffer (20mM Tris 7.5, 250mM NaCl, 15mM $MgCl_2$, 1mM dithiothreitol , 8% glycerol) supplemented with 0.5% triton, 30 U/ml Turbo DNase (Ambion) and 100μg/ml cycloheximide, ribosome protected fragments were then generated as previously described(*4*). We also lysed cells in a lysis buffer containing lower concentrations of salt and lower concentrations of Magnesium (20mM Tris 7.5, 150mM KCl, 5mM MgCl2, 1mM dithiothreitol , 8% glycerol). The read densities we got agreed very well between the two lysis conditions ($R^2 = 0.996$) (Fig.S5). The lower-salt and lower-magnesium lysis buffer, however gave footprints that had stronger frame bias, and the frame analysis was done on this sample.
For the LTM (50 μM for 30min) and PateamineA (200nM for 5 min) treatments cells were lysed with the lower-salt and lower-magnesium lysis buffer.

mRNA-Seq samples preparation

Total RNA was isolated from infected cells using Tri-Reagent (Sigma). Polyadenylated RNA was purified from total RNA sample using Oligotex mRNA mini kit (Qiagen). The resulting mRNA was modestly fragmented by partial hydrolysis in bicarbonate buffer so that the average size of fragments would be ~ 100bp. The fragmented mRNA was separated by denaturating PAGE and fragments 50-80 nt were selected as previously described (*4*).
Typically the mRNAs samples were fragmented to give an average size of 100bp. Thus theoretically (if there are no other effects) our procedure could have generated upto a 100-fold enrichment of the mRNAs 5' termini (as breaks along the body will accrue randomly every 100 bp). The observed enrichment of reads is not as large as this theoretical value (~10 fold enrichment for most viral messages), perhaps because the 5' cap makes the mRNA 5' end less likely to be captured since it is known to cause untemplated addition of nucleotide during reverse transcription reactions. Additionally, there may be multiple closely spaced 5' transcription starts associated with each transcript. In all figures where mRNA profiles are presented only the 5' end of the sequencing read is presented in the figure. To assess significance of 5' end enrichment we also measured the ratio of mRNA reads compared to a downstream 10 bp window in every position (with more with more than 7 reads) along the HCMV genome. The observed mean of this distribution is 1.2 and standard deviation is 1.6 and since all the 5' end peaks mentioned in the paper (Fig.1A, Fig.2C and Fig.4) show >5 enrichment they

are statistically significant.

Library generation and sequencing

The sequencing libraries were prepared and sequenced as in(*4*) with the following oligos used for rRNA substraction:
oNTI298: (biotin)-GGGGGGATGCGTGCATTTATCAGATCA
oNTI299 : (biotin)-TTGGTGACTCTAGATAACCTCGGGCCGATCGCACG
oNTI300: (biotin)-AGCCGCCTGGATACCGCAGCTAGGAATAATGGAAT
oNTI301: (biotin)-TCGTGGGGGGCCCAAGTCCTTCTGATCGAGGCCC
oNTI302h: (biotin)-GCACTCGCCGAATCCCGGGGCCGAGGGAGCGA
oNTI303h : (biotin)-GGGGCCGGGCCACCCCTCCCACGGCGCG
oNTI305: (biotin)-TCCGCCGAGGGCGCACCACCGGCCCGTCTCGCC
oNTI306: (biotin)-AGGGGCTCTCGCTTCTGGCGCCAAGCGT
oNTI307h: (biotin)-GAGCCTCGGTTGGCCTCGGATAGCCGGTCCCCCGC
oNTI308: (biotin)-TCGCTGCGATCTATTGAAAGTCAGCCCTCGACACA
oNTI309: (biotin)-TCCTCCCGGGGCTACGCCTGTCTGAGCGTCGCT
NG30h: (biotin)-CCCAGTGCGCCCCGGGCGGGTCGCGCCGTCGG
NG31h: (biotin)- CCGCCGGTGAAATACCACTACTCTGATCGTTTTTTC

Sequence alignments and normalization

Prior to alignment, linker and polyA sequences were removed from the 3' ends of reads. Bowtie v0.12.7 (*5*) (allowing up to 2 mismatches) was used to perform the alignments. First, reads that aligned to human rRNA sequences were discarded. All remaining reads were aligned to the concatenated viral (NC_006273.2) and human (hg19) genomes. Finally, still-unaligned reads were aligned to 200bp sequences that spanned splice junctions (see below). Reads with unique alignments (Table S1) were used to compute the total number of reads at each position in the viral genome. Footprint densities were calculated in units of reads per kilobase per million (rpkM) in order to normalize for gene length and total viral reads per sequencing run.
The PatA and parallel CHX treated samples were sequenced without rRNA substraction and absolute footprints read densities were calculated relative to the amounts of rRNA reads (Table S1). Note that there is some degree of variability (<50%) of rRNA between independent CHX treated samples but these differences are far less than the 8-fold increase in rRNA levels seen with PatA treated sample.

Splice junction mapping

De novo splice junction discovery was performed by running two radically different detection algorithims- TopHat v1.1.4 (which compute initial consensus of mapped regions and sequences flanking potential donor/acceptor splice sites within neighboring regions are then joined to form potential splice junctions which are used to try and align unalinged reads(*6*)) and HMMSplicer v0.9.5 (which uses half reads alignments to train a Hidden Markov Model (HMM) to detect the most probable splice position (*7*))– on mRNA reads from each time point. Both tools were executed with default options except the minimum and maximum allowed splice junction sizes which were set to 5 and 500K respectively. Splice junctions were kept only if they obtained (in at least one time point) both an HMMSplicer score of at least 1200, and a TopHat score of at least 7. These thresholds were chosen by looking at TopHat and HMMSplicer scores

distribution for the 17 known CMV splice junctions (at the time the analysis was performed), as well as at the overall distribution of HMMSplicer scores(*7*). The one known splice junction that didn't pass thresholds – a splice variant of UL37 - was manually added to the list as it was detected with lower scores (Table S3). The default output of these aligners is not restricted to canonical sites but no non canonical sites passed the required thresholds. Out of the 88 splice junction we identified all but 6 were recently independently identified by Gatherer et al. (Table S3). It should be noted that RNA processing into deep sequencing libraries involves a reverse transcription step during which template switching can occur and this could lead to false positives. Although in our protocol the reverse transcription step is done only after the messages are fragmented into small pieces (which significantly reduces the likelihood of template switching) we cannot exclude the possibility that some of the identified splice junctions are false positive identifications.

Translation initiation sites prediction

To identify translation initiation sites we applied machine learning approach as described before (*4*) with minor modifications. Initiation site scoring vectors were constructed using the normalized footprint profiles of the harringtonine and cycloheximide samples and were computed separately for 5hpi and 72hpi. Each vector was further normalized to have magnitude 1. In the training step, canonical ORF start codons that contained at least 6 normalized harringtonine read counts were used as positive examples, yielding 66 and 122 positive example vectors at 5hpi and 72hpi respectively. 10 negative examples were computed for each positive example as in (*4*). The SVMLight21 program was trained on 2/3 of the combined set of 188 positive and 1880 negative examples, using a radial basis kernel, $\gamma = 0.5$, $C = 100$, relative positive example weighing of 1.0, and without iterative removal and retraining. Cross-validation using the remaining 1/3 of examples showed 6 / 62 (13%) false negatives and 5/626 (1%) false positives. The trained classifier was then run on all plus and minus strand codons that had at least 6 normalized harringtonine read counts (5574 (1.2%) and 36931 (7.8%) of all possible codons at 5hpi and 72hpi respectively). Initiation sites that scored less than 0 were discarded. Contiguous blocks of initiation site nucleotide positions were then collected into single initiation sites as in (*4*). ORFs were then defined by extending each initiating codon to the next in-frame stop codon, and incorporating any intervening splice junctions detected within that time point. The 5hpi and 72hpi ORFs were then combined into one list of ORFs, and further processed to discard internal ORFs with non-AUG start codons and to include the manual ORFs (table S5 and table S6).
The manual curation was done on ORFs that included at list 10 harringtonine reads at the initiating codon and was focused mainly on the ORFs that have been previously annotated.
When running the trained SVM classifier on a biological replicate we obtained high agreement between replicates with 86% of the original ORFs being predicted by the biological replicate (Table S5, Table S6 and Fig.S4). Given our estimate of 1% false positive rate (from cross-validation), the primary reason for reduced agreement between replicates is the presence of false negatives in each replicate. In this light, our estimate of 13% false negatives agrees well with the 86% overlap we report between replicates.

Transcript 3' end detection

Prior to alignment, polyA sequences were removed from the 3' ends of mRNA reads. These sequences were defined as 5 or more consecutive A's at the 3' end of a read (after linker removal), allowing for 1 non-A base for every 5 A's (to account for sequencing errors). Then the reads were aligned to the viral genome, and 3' ends of polyA-containing reads were counted. To prevent long stretches of A's in the genome from causing false positives, the polyA sequences that were removed from the reads prior to alignment were compared to the underlying reference sequence after alignment. If the polyA sequence was found to have fewer than 3 mismatches with the reference, this read was not used as evidence for the 3' end of a transcript.

Analysis of LTM enrichment

The ribosome footprints density at each start codon relative to the median density across 20 bp downstream was calculated for every ORF. As a control, the occupancy of a codon five positions downstream of the start codon was calculated in the same way. ORFs were binned according to the level of enrichment and new ORFs that fell into bins that were likely to have more than one false positive (estimated by the distribution of the control positions) were removed (13 ORFs).

Clustering

Clustering was performed using Cluster 3.0 (*8*). Ribosome footprints were normalized to the total number of viral reads at each time point (RPMK). Genes were then centered so that the median value of each row is 0, and normalized such that the square root of the sum of the squares (ie. the magnitude) of the row was 1. Clustering data were then visualized using TreeView (*9*).

ORFs analysis and nomenclature

The analysis and figures in the manuscript were done using the GenBank accession NC_006273.2 (24 February 2011 version) that contained 165 annotated ORFs in the Merlin genome. During the preparation of this manuscript Gatherer et al(*3*) have added 6 new ORFs to these annotations (GenBank accession AY446894), 5 of which were also identified here and their names were revised according to Gatherer et al. nomenclature(*3*).

We named the ORFs we have identified in a way that does not overlap with the established nomenclature but still provides information about their location in the genome.

ORFs longer than 20aa were named:

ORF (L/S) (number) (W/C)

L/S indicates the location in the genome (long or short region)

W/C indicates the strand + (Watson) or - (Crick).

For the previously annotated ORFs we added the official names in brackets.

ORFs that are shorter than 20 amino acids were named: rORF (*number*) (table S6).

Conservation analysis**:**

Each ORF was used as a query sequence in a tBlastN search against a database of sequenced HCMV strains (GQ221973.1, GQ466044.1 ,AC146904.1, GQ221974.1, GU179290.1, GU179291.1, GQ221975.1, X17403.1, GQ396662.1, HQ380895.1,

GU937742.1, FJ616285.1, GU179289.1, AC146907.1, EF999921.1, AF480884.1). An alignment with e-value > 10^{-4} was scored as an orthologue. The analysis included only ORFs were longer than 20 amino acids and internal in-frame ORFs were discarded. For spliced ORFs only the sequence of the 1st exon was used.

Construction of recombinant HCMV viruses

For generation of the HCMV mutants expressing HA-tagged forms of ORFS370W_(US33A) and ORFL49W_(UL16) PCR fragments were amplified using the plasmid pSLFRTKn (*10*) as template DNA. The PCR fragments containing a kanamycin resistance gene were inserted into the parental TB40 BAC (*11*) by homologous recombination in E. coli. The kanamycin cassette was excised from both BACs by FLP-mediated recombination (*12*). Correct mutagenesis was confirmed by southernblot and PCR analysis (data not shown). Recombinant parental and mutant HCMVs were reconstituted from HCMV BAC DNA by Superfect (Qiagen) transfection into MRC-5 cells.

Primer sequences (TB40 tagged viruses):

KL-ORF370W_(US33A) -Kana1

TTCGTCCCAAGCGGGTTCGTCATGGGCCATGTTGCAATTATCAGCGTCACCAC
TGACGTGCCAGTGAATTCGAGCTCGGTAC

KL- ORF370W_(US33A)-Kana2

CAACTCATGATGTACCCGCTGGTGTTGCCGTTCACTTTTCCGTTTTACGTGCCG
CGGTCCGGTGGTGGTTACCCATACGATGTTCCAGATTACGCTTAGGACCATGA
TTACGCCAAGCTC

KL-UL16-Kana1

GGGGGACATGGCCTTCTTATAGCAGCGTGAACGTTGCACGTGGCCTTTGCGG
TTATCCGTCCAGTGAATTCGAGCTCGGTAC

KL-UL16-Kana2

CTGTGTCAGCGGCTGCGCATTCGCCTGCCGCATCGATACCAGCGGTTACGCA
CCGAGGACTACCCATACGATGTTCCAGATTACGCTTAAGACCATGATTACGC
CAAGCTCC

Primer sequences (pIRES-EGFP expression plasmids):

KL-UL16-49aa-1

CGgaattcCCTCCGTAAGCGCCGACGAGG

KL-UL16-49aa-2

CGgaattcTTAAGCGTAATCTGGAACATCGTATGGGTAGTCCTCGGTGCGTAACC
GCTG

KL- ORF370W_(US33A) -1

CGgaattcGGAGTCTCTGGCGCAAAAGCAC

KL- ORF370W_(US33A) -2

CGgaattcTTAAGCGTAATCTGGAACATCGTATGGGTAACCACCACCGGACCGCG
GCACGTAAAACG

UL138 Immunoprecipitation

MRC-5 cells were infected for 48hr were labeled with 35S methionine/cysteine for 2hr before lysis. Cells were lysed (0.1 mM EDTA, 150 mM NaCl, 10 mM KCl, 10

mM $MgCl_2$, 10% [vol/vol] glycerol, 20 mM HEPES [pH 7.4], 0.5% [vol/vol] IGEPAL, 0.1 mM phenylmethylsulfonyl fluoride, 1 mM dithiothreitol, 0.4 mM pepstatin A, 0.1 mM Na-vanadate, Complete protease inhibitor [Roche]). Lysates were spun for 30 min at 4°C and 16,000 × *g*. Immunoprecipitation was done using an anti-UL138 rabbit serum that was generated by peptide immunization (sequence: VRRESDRRYRFSER) by Coring Systems Diagnostix. The varL and varS AD169 variants used in this experiment are described in (*13*).

Northern blotting

Total RNA was isolated from infected cells using Tri-Reagent (Sigma) and Northern blotting was performed using the NorthernMax system (Ambion). The transcripts were detected using strand-specific RNA probes to the ~200 base pairs at the 3' end of each ORF. Probes were synthesized using the T7 Maxiscript kit (Ambion) and P32 labeled UTP. Methylene blue staining was used as a loading control.

DNA constructs

For N terminal fusion to luciferase we used the firefly luciferase from the pSP-luc-N1 (promega) and cloned it into the pGL3 control vector (Promega) using the HindIII and XbaI sites. The 5′UTR and part of the coding region of UL119 and US9 were amplified by PCR and cloned to the vector using HindIII and BsteII sites. The inserts and their proper orientation were confirmed by sequencing. The primers used to amplify the various UTR and coding regions were: For UL119-long forward; AAGCTTTAGGCGTCACAAGAGGTGAC, UL119-short forward; AAGCTTATAGCTCGTCCACACGCCG and GGTGACCCGCGCCAGCATGTAGGTCA as a reverse primer.
US9-long forward; 5'- AAGCTTACGTCGGGGCTGACGCG -3', US9-short forward; 5'- AAGCTTCGGGCGGCTTCCTGCGG -3' and reverse; GGTGA*C*CTCAGTTTCGTAACGTAGAG
For making the mutated UL119 plasmid we used the quikchange site-directed mutagenesis kit (Agilent) with the following primers; GACCTCCTGC CACAGATACC TCGTCCACAC GCC, GGCGTGTGGACGAGGTATCTGTGGCAGGAGGTC
For making the mutated US9 plasmid we used the multi site-directed quikchange directed mutagenesis kit (Agilent) with the following primers; GCGGATCGGACATCCTACAGTCGTGAGGCGCTCCG,ACGGGGGTTTCCCCAGA CACGTAATACTCG, ATAAAGAGGCACGGAGTTTCGGCTCCCGC
For GFP fusion the ORFs and their 5' UTRs were amplified by PCR and cloned into pEGFP-N1 vector (Clontech) using XhoI and EcoRI sites using the following primers were used (forward; reverse): For ORFL88C; GCGCCTCGAGCGCGCCGGGGAGAAGAAT, GCGCGAATTCGACCGCCGTGACACATCAGG. For ORFL88C.2; GCGCCTCGAGCGCGCCGGGGAGAAGAAT, GCGCGAATTCGCTGTCCGAGAGGGTGCGAT. For ORFS362W; GCGCCTCGAG*TGTTCACGTCGTTCGTGGG*, GCGCGAATTC GAAGCCACGT GCTATTCTGAC; For ORFS370W; GCGCCTCGAGAGGAGTCTCTGGCGCAAAAG, GCGCGAATTCGGGACCGCGGCACGTAAAAC For ORFL167W;

GCGCCTCGAGCGCACCTCGCCGCCGTG,
GCGCGAATTCGCAGACGCAAAAGTGCTCGGA
For making the mutated plasmids a nucleotide was inserted down stream of the predicted start site using quikchange site-directed mutagenesis kit (Agilent) with the following primers: For ORFL88C and ORFL88C.2;
GTCCGCTTCATCAGAGCTGCGTACGCTTTGG
CCAAAGCGTACGCAGCTCTGATGAAGCGGAC. For ORFS362W;
AGACTCACGTAATTTGGCAACATGTGCGACCGCAA,
TTGCGGTCGCACATGTTGCCAAATTACGTGAGTCT. For ORFS370W;
TTCCCCGAGCAGGGCGGGTTACG, CGTAACCCGCCCTGCTCGGGGAACCTG.
For ORFL167W; GTAGAGCGGCCGACGTCTCGGCCTC,
GAGGCCGAGACGTCGGCCGCTCTAC.
For GFP fusion for microscopy experiments the ORFs were amplified from start to stop codon (excluding the actual stop codon) and cloned into pEGFP-N1 vector (Clontech) using XhoI and EcoRI sites.

Luciferase assay, real time PCR and western

HeLa were grown to 70% confluence in 24-well plates. Cells were transfected using FuGene6 (Roche) with 600 ng of the *firefly* luciferase reporter vector and 10ng of the control vector containing *Renilla* luciferase, pRL-CMV (Promega), in a final volume of 0.5 ml. *Firefly* and *Renilla* luciferase activities were measured using the Dual-luciferase assays (Promega), 36 hr after transfection.
For real time PCR analysis cells were transfected in 6-well plates using 1.8µg of the firefly luciferase reporter vector and 30ng *Renilla* luciferase vector. Total RNA was then extracted using the QIAGEN RNeasy Mini Kit and RNA was treated with DNase I. 1 µg of purified total RNA was combined with 2 µl of 50 µM random hexamer solution (ABI), 1 µl of 20 U/µl Superase· In (Ambion), 1 µl of 10 mM dNTPs (NEB), 2 µl of 10x M-MuLV buffer (NEB) and 1 µl of 200 U/µl M-MuLV reverse transcriptase (NEB). This mixture was incubated at room temperature for 10 min, next at 42°C for 1 hr and finally at 90°C for 10 min. Real time PCR was performed using the DyNAmo HS SYBR Green qPCR Kit (Finnzymes) on a DNA Opticon Real Time Cycler (MJ Research) with the following primers;
Firefly-FW; TCGCCAGTCAAGTAACAAC
Firefly-REV; 5′-ACTTCGTCCACAAACACAA
Renilla-FW; AACGCGGCCTCTTCTTATTT
Renilla-REV; GTCTGGTATAATACACCGCG
For western, HeLa cells were grown to 70% confluence in 6-well plates. Cells were transfected with 2.5ug of the different plasmids. 48hr after transfection cells were lysed using 250µl of RIPA buffer. Lysates were placed on ice for 30 min and then centrifuged at 20,000 × *g* for 15 min at 4°C. Samples were then separated by 4-12% Bis-tris gel (Invitrogen), blotted onto nitrocellulose membranes and immunoblot with anti-GFP antibody (Roche) or anti-Flag (Sigma).

Fluorescence microscopy and subcellular localization

HeLa cells were transfected with 2µg of the various GFP tagged plasmids. For ER co-localization cells were cotransfected with 1 µg mCherry-KDEL (mCherry with KDEL-ER retention sequence). For mitochondia staining cells were incubated in 250nM MitoTracker Red CM-H2XRos (invitrogen) for 30 min at 37°C, for nucleus staining cells were incubated in 2 µg/mL Hoechst 33342 stain for 30 min at 37°C. Cells were imaged 48 hours post transfection using a spinning disk confocal microscope.

Affinity-Purification, Mass Spectrometry and western

HeLa expressing the different GFP tagged proteins were cultured in standard cell culture conditions, expanded to 3 × 15 cm cell culture dishes in triplicate experiments, harvested with Trypsin, and flash frozen in liquid nitrogen. Cells were lysed with 1 ml per plate of 1.5% Digitonin containing IP buffer (50 mM HEPES-KOH, 150 mM KOAc, 2 mM MgOAc, 1 mM CaCl2, 15% glycerol, protease inhibitor). After a 45 min incubation step the cell lysate was clarified by a ultracentrifugation spin at 20 min, 58,000 g, 4°C. The supernatant was incubated with a 10 µl bed volume of anti-GFP-antibody coupled agarose beads (trapA beads, Chromotek) for 3 hr at 4°C. The beads were then washed five times with 10 ml of wash buffer I (0.1% Digitonin, 50 mM HEPES-KOH, 150 mM KOAc, 2 mM MgOAc, 1 mM CaCl2, 15% glycerol), washed once with 10 ml of wash buffer II (50 mM HEPES-KOH, 150 mM KOAc, 2 mM MgOAc, 1 mM CaCl2, 15% glycerol) and once with 1 ml of wash buffer II for transfer of the beads to 1 ml snap lid tubes. Proteins recovered by the immunoaffinity purification were trypsin digested on the agarose beads and peptides eluted and treated as previously described (*14*). Resulting peptides were subjected to reverse phase liquid chromatography on a Proxeon easy-nLC system (Thermo Fisher) and directly sprayed into an LTQ-Orbitrap mass spectrometer (Thermo Fisher). Raw files were processed with MaxQuant, version 1.2.6.20 (*15*), using the label free protein quantification algorithm and matching of peptide identifications between runs.

For western analysis the same affinity purification protocol was used. Proteins were eluted from the beads by boiling with SDS sample buffer. Samples were then separated by 4-12% Bis-tris gel (Invitrogen), blotted onto nitrocellulose membranes and immunoblot with anti-GFP antibody (Roche), anti-TIMM8A (Sigma), anti-TIMM22 (Abcam) and anti TAP1 (Millipore)

Mass spectrometry sample preparation and analysis

HFF cells were infected at MOI of 5 for 72 hr. 5hr prior to lysis, cells were either treated with MG132 (Sigma) at a final concentration of 10 µM or left untreated. Cells were lysed with 4% (w/v) SDS, 100mM DTT in 100mM Tris pH 7.6 as previously described(*16*). Samples were then separated on 4-12% Bis-Tris gels and were then subjected to in-gel digestion (*17*). The peptide mixtures were desalted on reversed phase C18 StageTips (*18*) directly before analysis, peptides were eluted with 50 µL buffer B (80% ACN in 0.5% acetic acid). Organic solvent was removed in a SpeedVac concentrator and the final sample volume was adjusted with buffer A* (2% ACN in 0.1 % TFA) to 6 µL. To capture a larger fraction of the small proteins we also tested other approaches such as 10 kDa size exclusion filters to filter the lysate before digestion, but the LC-MS data contained few peptides and high amounts of contaminants. We found

that gel-based separation provided the best results and that the main limitation for detection was length (Fig. S22).

LC MS/MS analysis

Each peptide fraction was separated by nanoflow HPLC (Easy nLC 1000, Thermo Scientific) on a C_{18}-reversed phase column (20 cm long, 75 µm inner diameter; packed in-house with ReproSil-Pur C_{18}-AQ 1.8 µm resin (Dr. Maisch GmbH) with a linear gradient of 5–30% buffer B (80% ACN and 0.5% formic acid) at a flow rate of 250 nL/min and at 40 °C over 120 min. The HPLC was on-line coupled to a Q Exactive mass spectrometer (Thermo Scientific) via a nanoelectrospray ion source (Proxeon). The MS data was acquired with a data-dependent top10 method that automatically selects the most abundant precursor ions for HCD fragmentation. Full scans were acquired with target value 3e6 and resolution 70,000 at m/z 200 and MS/MS scans with target value 1e6 and resolution 17,500 at m/z 200 in a 1.6 Th isolation window. The injection time for MS spectra was restricted to 20 ms and for MS/MS spectra to 60 ms.

Proteomics data analysis

The RAW files were analyzed with the MaxQuant software (version 1.2.2.6) (*19*). Scoring of the peptides for identification was carried out with an initial allowed mass deviation of the precursor ion of up to 7 ppm for the search for peptides with a minimum length of 7 amino acids. The allowed fragment mass deviation was 20 ppm. The false discovery rate (FDR) was set to 0.01 for proteins and peptides. The Andromeda search engine with the 2nd peptide function enabled (*15*) was used to simultaneously search the MS/MS spectra against the IPI human data base (version 3.68, 87,061 entries) and an HCMV database, containing the officially annotated proteins as well as the predicted novel proteins combined with 262 common contaminants. Enzyme specificity was set as C-terminal to Arg and Lys and a maximum of two missed cleavages. Carbamidomethylation of cysteine was set as fixed modification and N-terminal protein acetylation and methionine oxidation were specified as variable modifications. The mass accuracy of the precursor ions was improved by the MaxQuant time-dependent recalibration algorithm. Furthermore, we manually verified the identity each of these high mass accuracy spectra (fragment mass deviation < 20 ppm), which also revealed that the majority of them contained complete sequence information (supplementary file 2)

MochiView database file

File S1 is a MochiView database file (*20*) that contains our annotations together with all the footprints (from cylcoheximide , harringtonine, LTM treated samples) and mRNA data.
This file could be viewed by downloading MochiView from:
(http://johnsonlab.ucsf.edu/sj/mochiview-start/).
After opening MochiView go to the menu
Database ➔ Manage Databases ➔ import ; Choose the file and give a name to the database when prompted. Next choose the database you loaded ➔ Activate and allow the program to restart. After the restart, 3 plots will be displayed. Each one contains; the merlin annotations (NC_006273.2), our footprints-based annotations, ribosome occupancy profiles and mRNA profile for each of the time points (5 hpi, 24 hpi and 72

hpi). The ribosome occupancy profiles at 5 hpi and 72 hpi are composed of four tracks (CHX, Harr, LTM and no drug) whereas the 24 hpi ribosome occupancy profile contains only the CHX track. Note that this genomic browser does not illustrate the splice junctions within ORFs.

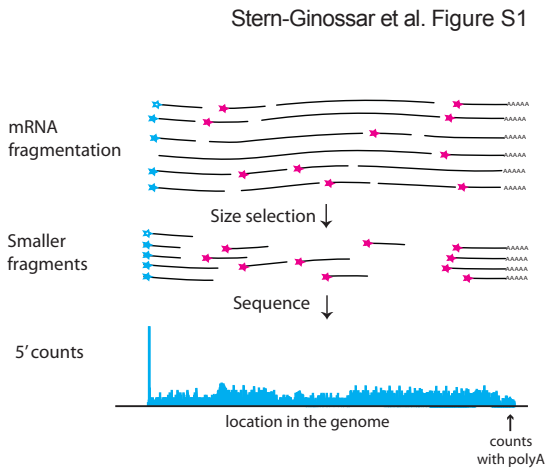


Fig. S1.
Schematic representation of RNA-Seq protocol that leads to mRNA 5’ end enrichment. Isolated mRNA is subjected to partial fragmentation followed by size selection of small fragments (50 to 80bp, marked with a star). This procedure generates enrichment in fragments that start at the 5’ end of transcripts that is ilustrated by plotting the 5’end of each sequencing read. We identified the 3’ end of messages by searching for the final nucleotide preceding the start of polyadenylation tracks from the sequenced mRNA fragments.

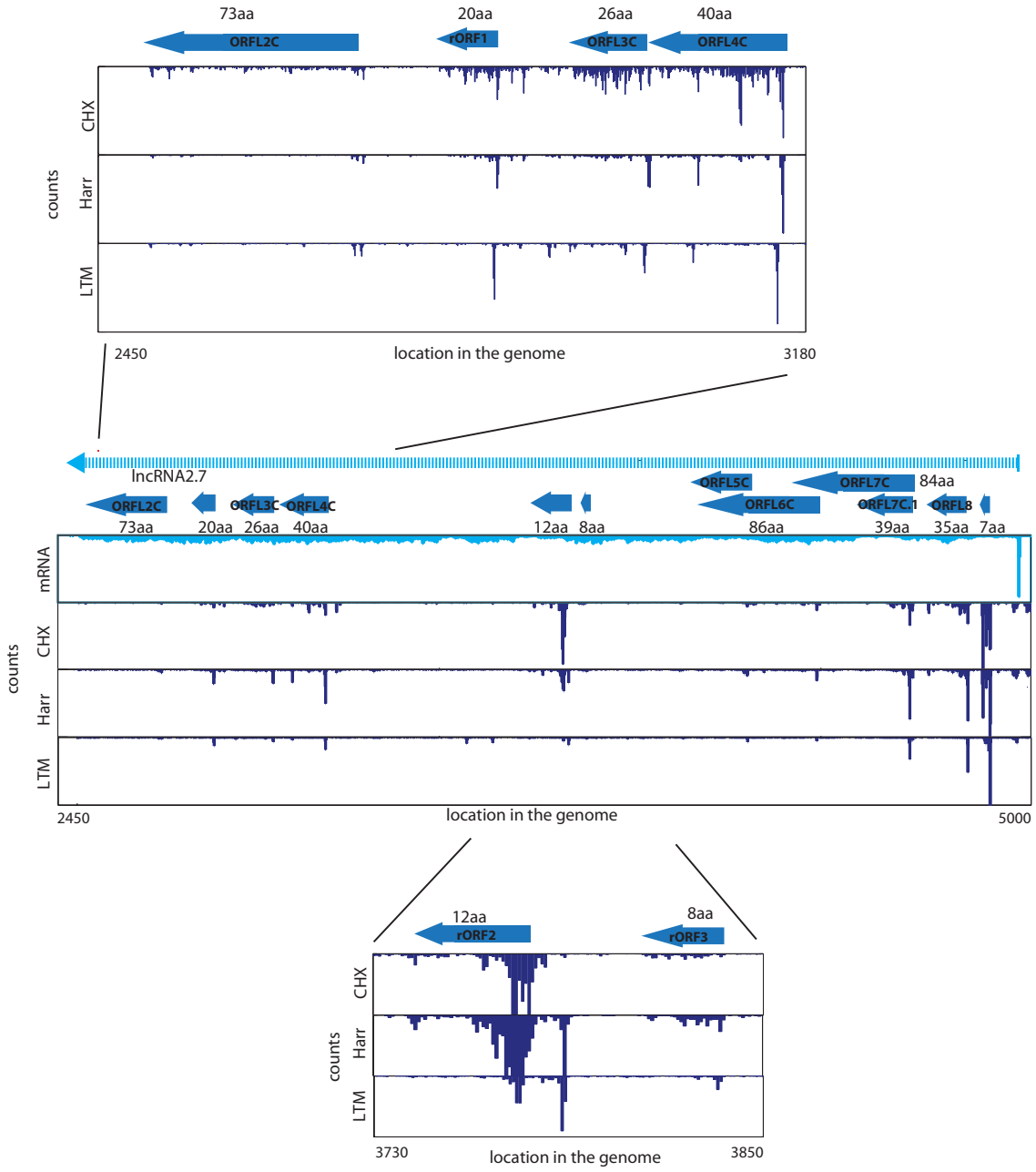


Fig. S2

The ribosome occupancy and mRNA profiles are shown around RNA β2.7. The lower and upper panels show expanded views for some of the ORFs.

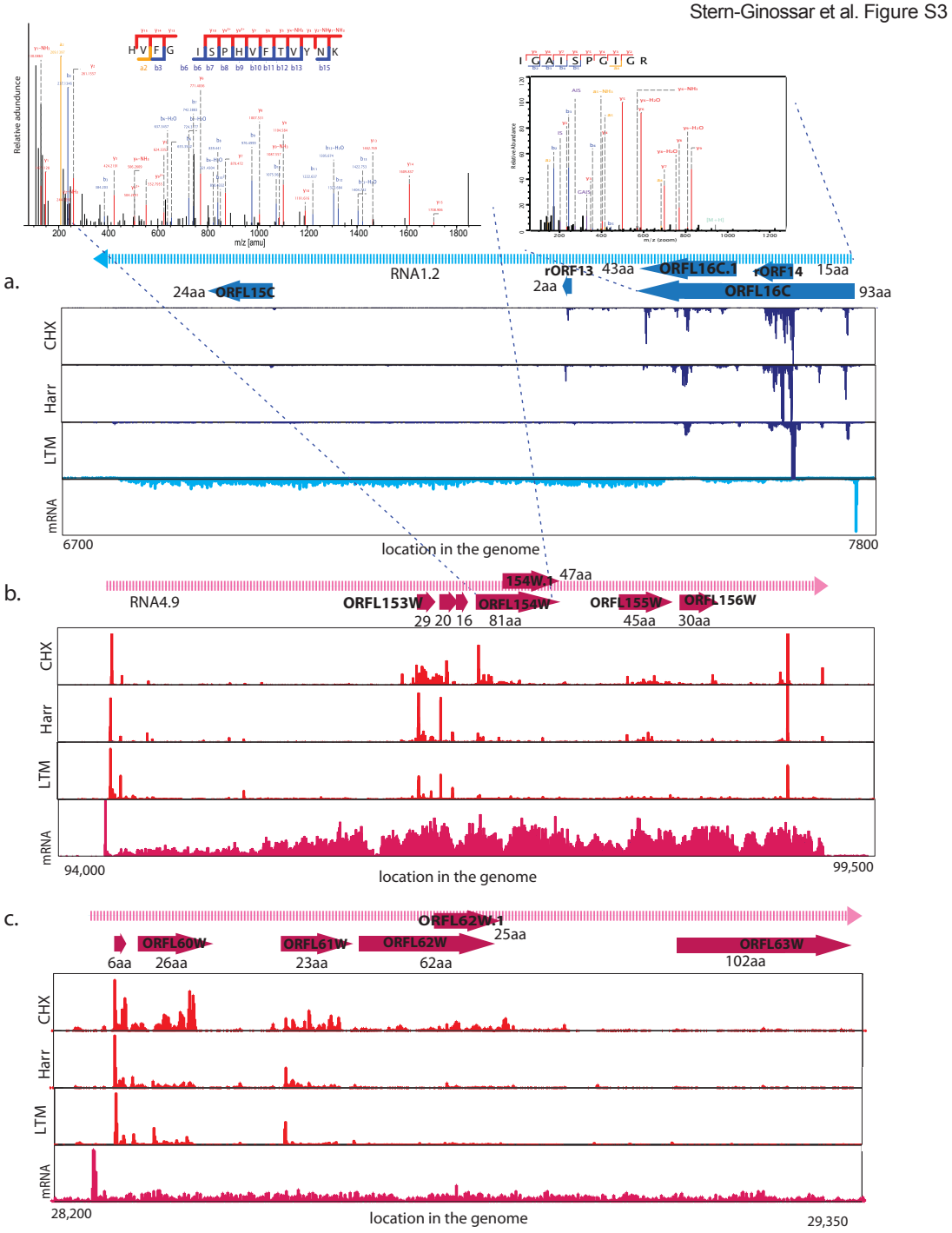


Fig. S3

Examples of translation from long RNAs that lack canonical ORFs. The ribosome occupancies and mRNA profiles at 72 hpi are shown around three different genomic loci (a, b, c). Note that from each of the RNAs there are multiple short ORFs that are being translated. The upper panels show the annotated MS/MS spectra of two unique peptides,

(a) IGAISPGIGR, originating from the 93aa protein encoded by RNA1.2 and (b) ISPHVFTVYNK that originated from the 81aa protein encoded by RNA4.9.

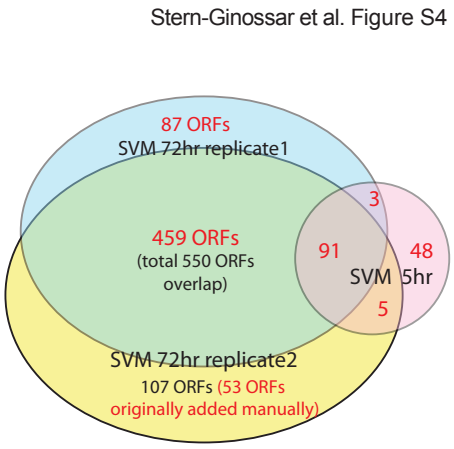


Fig. S4

Venn diagram summarizing the reproducibility of the SVM predictions. ORFs labeled in red are the ORFs that were included in our annotations. Note the diagram presents only the 751 ORFs that were supported by the LTM analysis (excluding the 12 ORFs that were predicted by the SVM but were not supported by the LTM data).

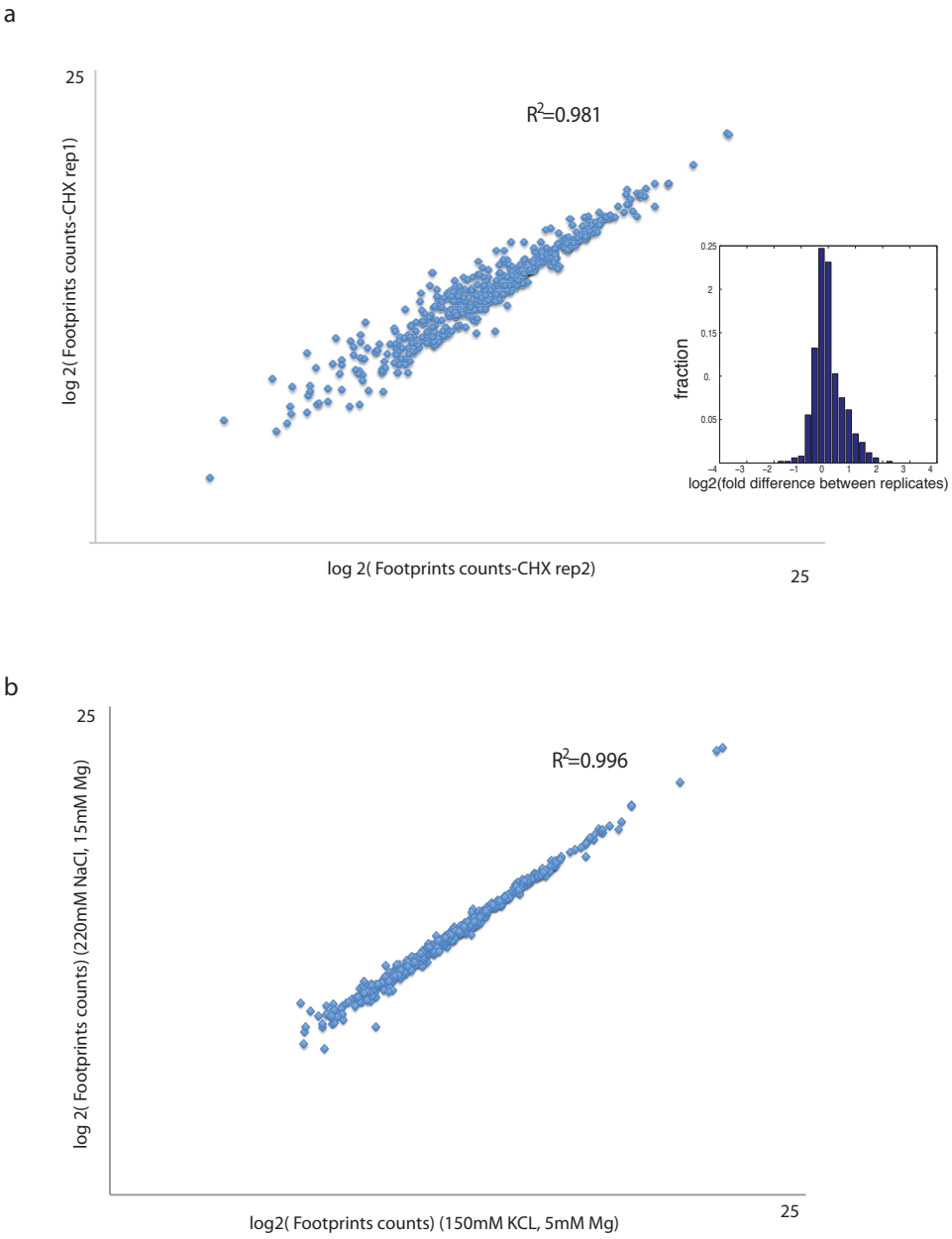


Fig. S5

Analyses of independent biological replicates and different lysis conditions reveal high reproducibility in footprints densities. Correlation between footprints densities obtained for HCMV ORFs from (a) independent biological replicates generated from cells 72hpi treated with CHX (Inset) histogram of ratios between replicates (calculated for each ORF) or (b) cells 72hpi treated with CHX and then lysed with different lysis buffers containing different concentration of Magnesium and salts.

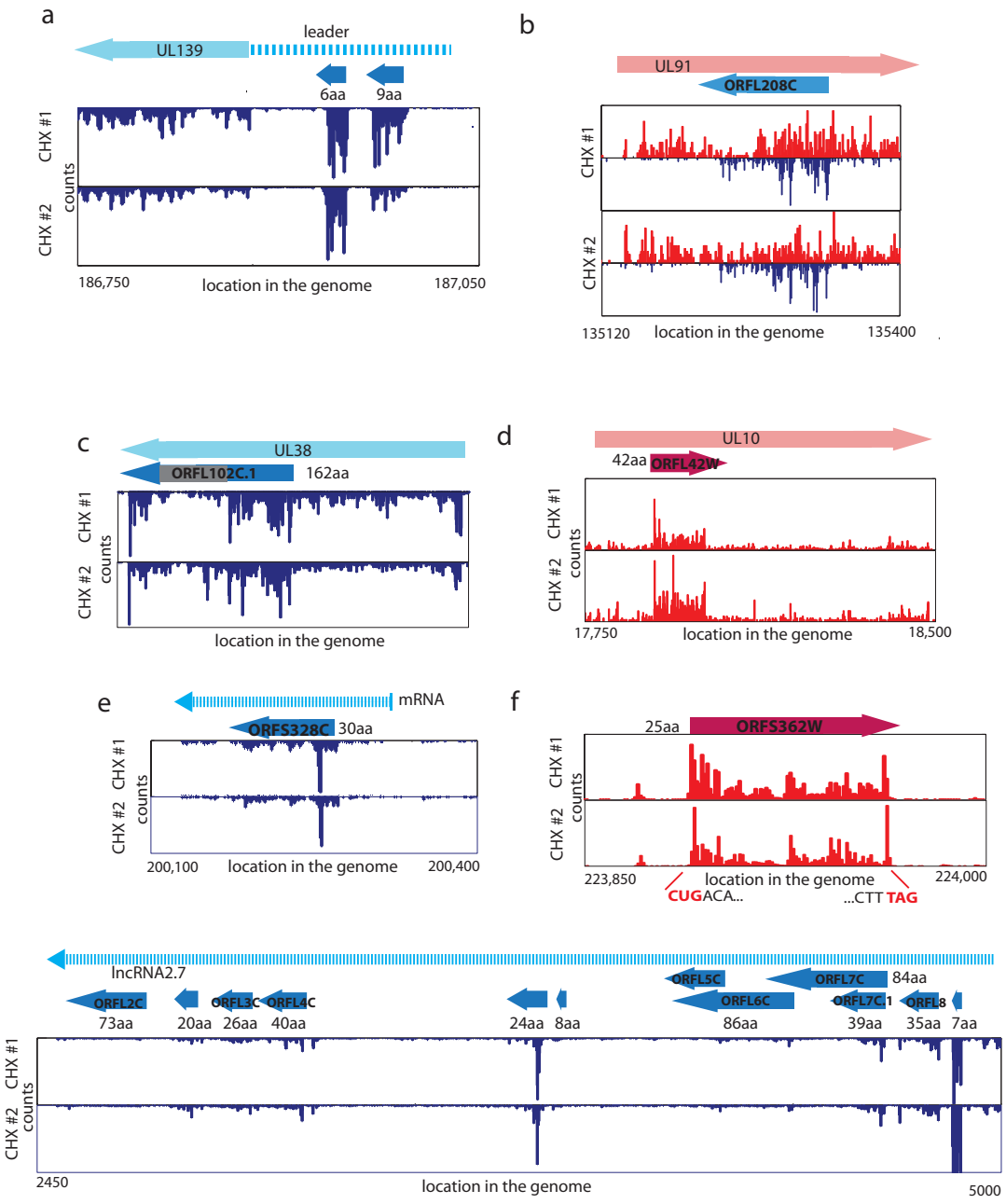


Fig. S6
Reproducibility between biological replicates of footprints that do not correspond to previously annotated viral ORFs. The ribosome occupancy profiles of two independent biological replicates are shown for (a) the leader region of UL139 gene, (b) the region of UL91 gene, (c) and regions of UL38 and (d) UL10 genes. (e) The ribosome occupancies around a short ORF that initiates at a CUG. (f) The ribosome occupancy profiles are shown around RNAβ 2.7.

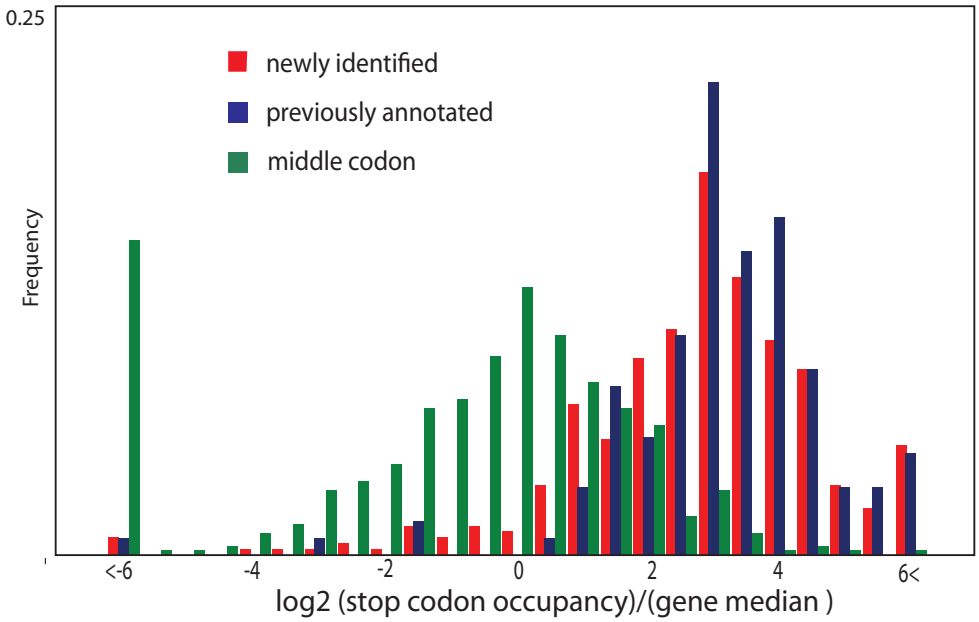


Fig. S7

The distribution of ribosome occupancy at stop codons. The ribosome footprints density at each stop codon (relative to the median density across the gene) is depicted for the previously annotated ORFs (blue) and for the newly identified ORFs (red). As a control, the occupancy of the middle codon of each ORF was calculated in the same way (green). The analysis was done on the no-drug sample.

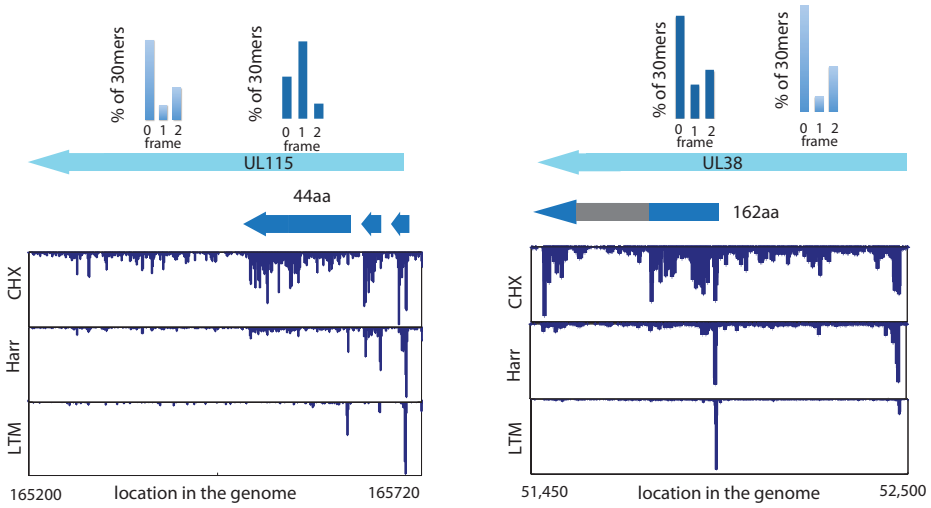


Fig. S8

The position of 30-nt ribosome footprints relative to the reading frame within (a) an ORF that contains an out frame overlapping ORF, where a shift in the translated frame is observed (compare light blue to dark blue upper panels) or (b) within an ORF that contains an in-frame overlapping ORF in which no shift is observed.

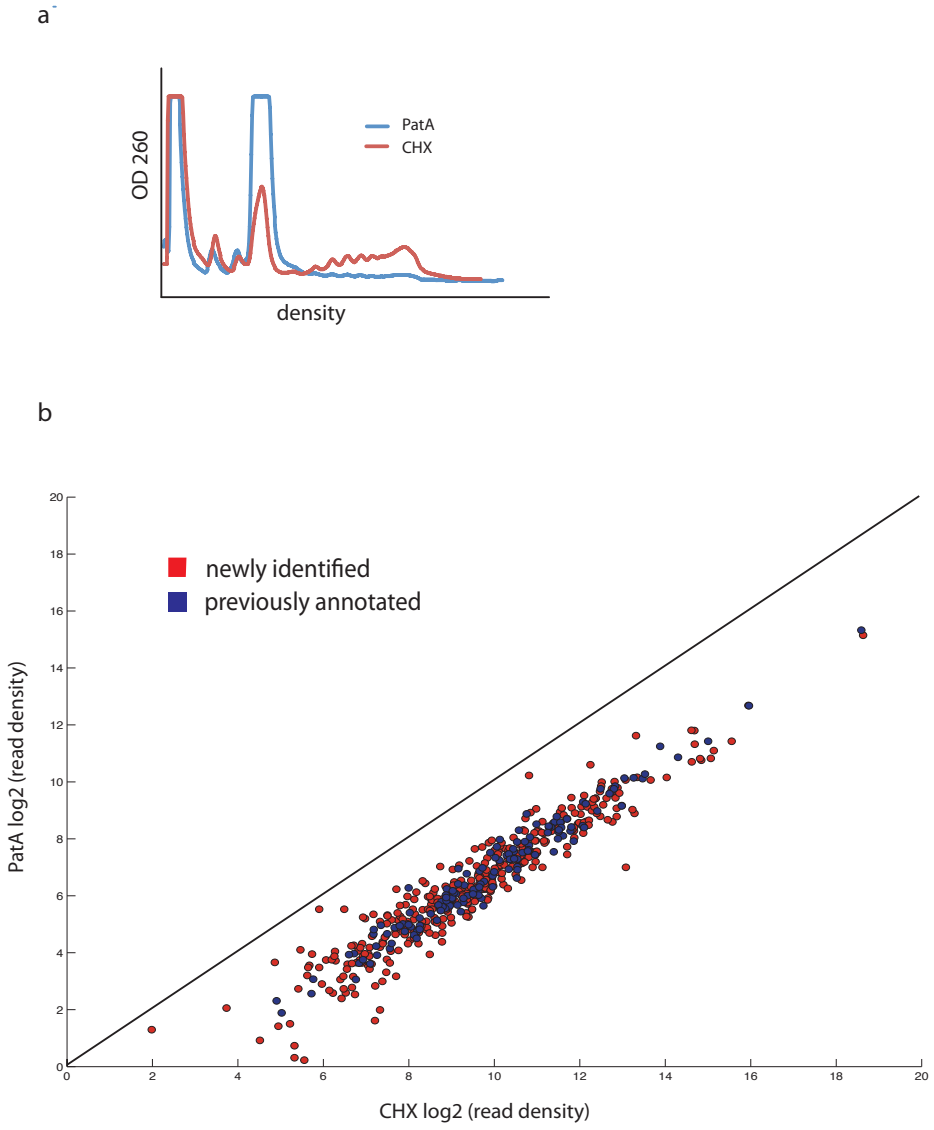


Fig. S9

Ribosomes run-off after a brief Pateamine A treatment**.** (a) Cells were treated with CHX (100 μM) or Pateamine A (PatA) (200nM) for 5 min before lysis. Cell lysates were subjected to centrifugation through a 10–50% sucrose gradient and their polysome profiles are presented showing that after treatment with PatA the polysomes collapse. (b) Correlation between absolute footprints densities of libraries prepared from cells treated with either CHX or with PatA. The absolute footprints densities were calculated relative to rRNA levels in the sample. Both newly identified (red) and previously annotated (blue) HCMV ORFs show a drop in footprints densities after PatA treatment demonstrating that the majority of the ORFs chase following a short inhibition of translation initiation.

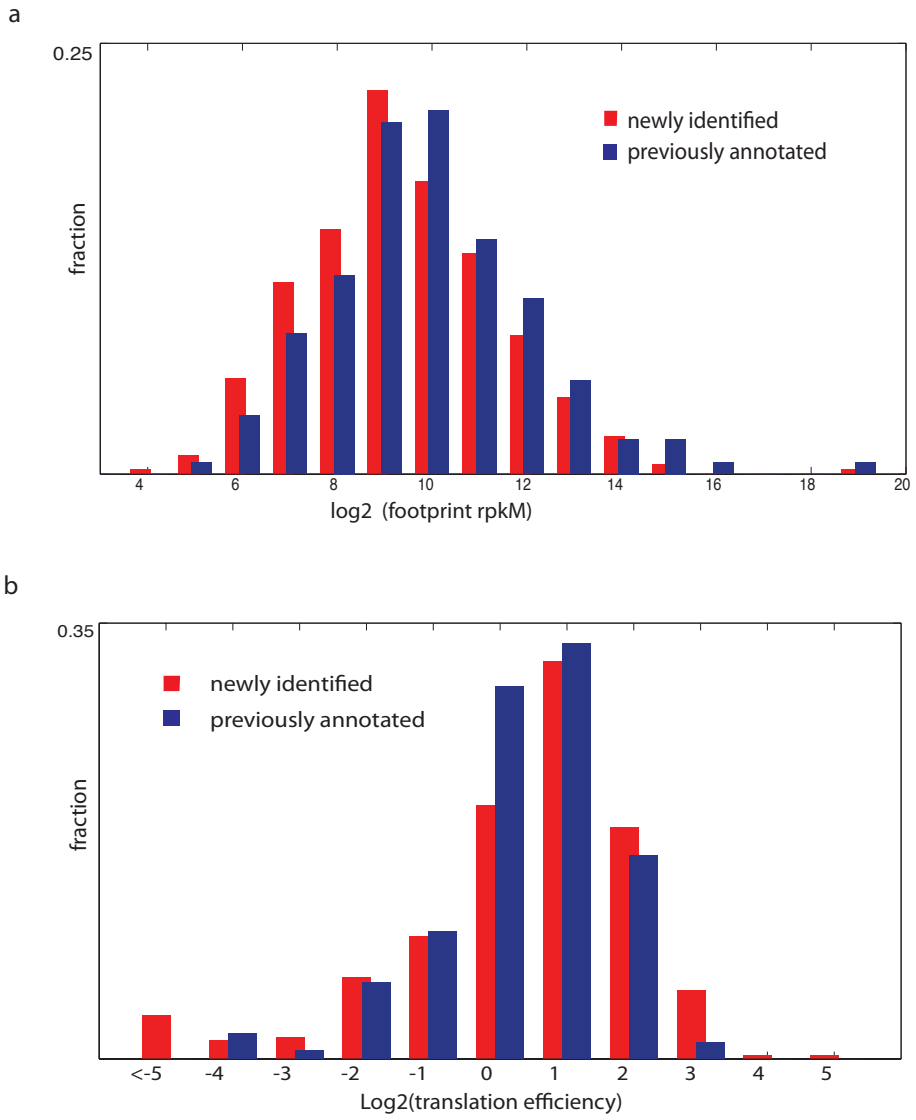


Fig. S10

The expression and translation efficiency of HCMV translated ORFs. (a) The expression distribution of HCMV translated ORFs at 72hpi. The newly identified ORFs (red) and the previously annotated ORFs (blue) are binned by their normalized footprints density, reads per kilobase per million (rpkM). The difference between distributions was not statistically significant ($P = 0.199$, *Ks* test).

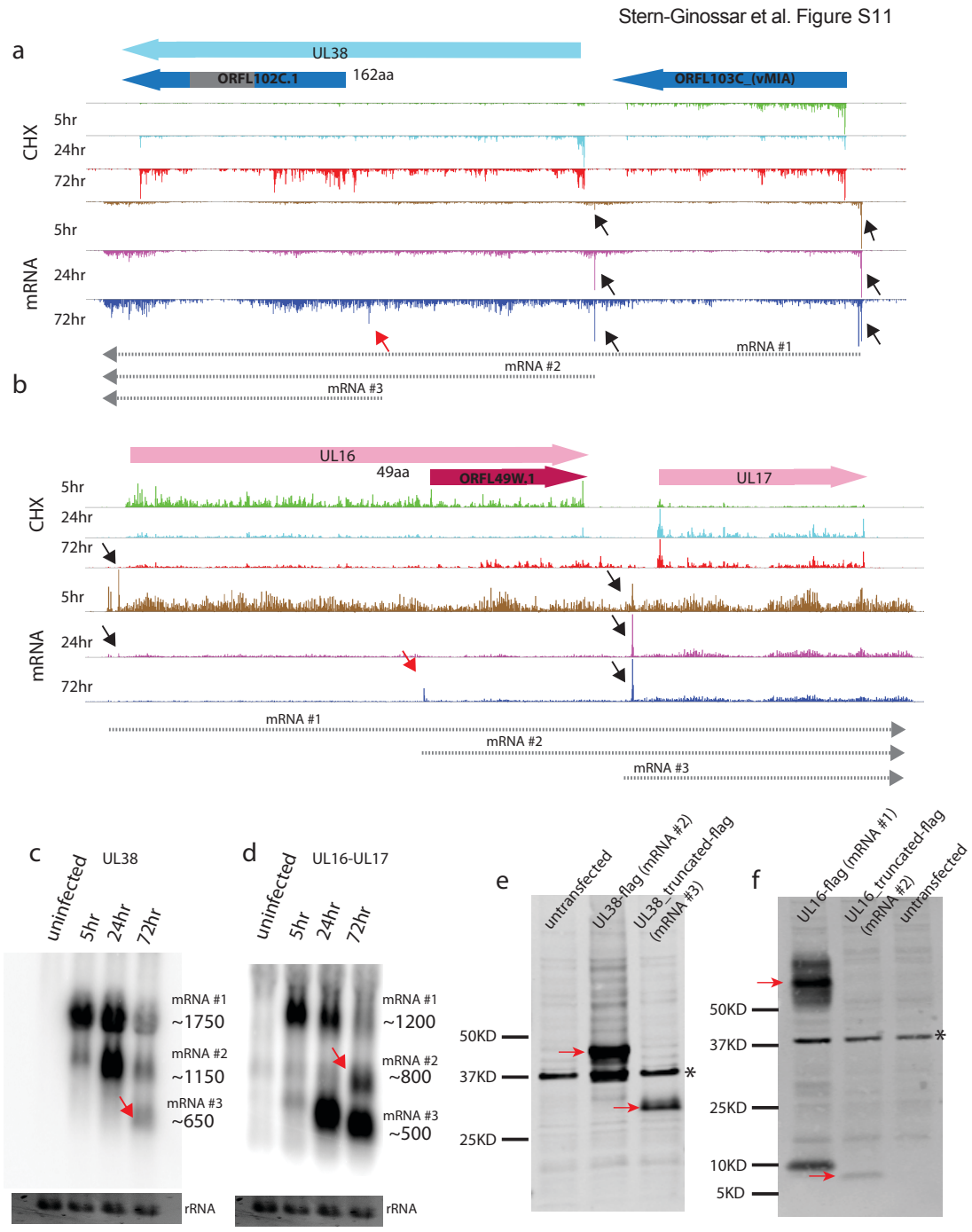


Fig. S11

Expression of N-terminally truncated viral proteins. (a, b) Changes in transcripts 5' ends during infection that lead to N-terminally truncated protein product. The mRNA and ribosome occupancy profiles are shown around (a) UL38 and (b) UL16-UL17 loci at different time points during infection (marked on the left). Small arrows denote the

different mRNA starts and the corresponding mRNAs are illustrated (lower part). Note that the changes in the transcript 5' ends (marked by red arrows) expose internal ORFs. (c, d) Total RNA extracted at different time points along infection was subjected to Northern blotting using a probe against the 3' end of the message; the transcripts in which the 5' end is just upstream of the internal ORFs are marked with red arrows. (e, f) Expression of the transcripts with a c-terminal tag results in the expected full and truncated protein products (marked by red arrows). Unspecific products are marked with a black asterisk.

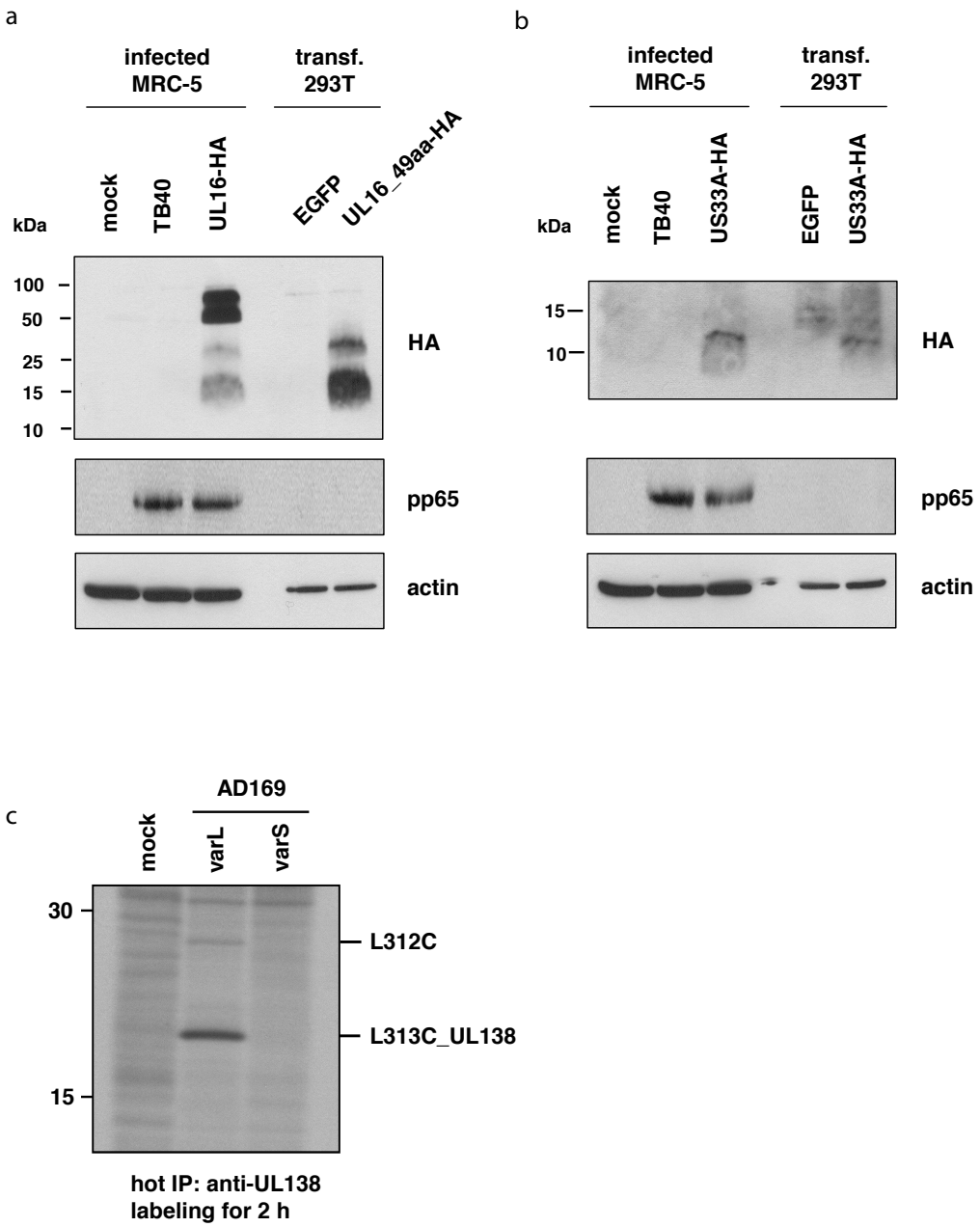


Fig. S12

MRC-5 cells were infected with the parental TB40 and (a) TB40-UL16-HA or (b) TB40-US33A-HA for 72 hr. Protein lysates were subjected to westernblotting for detection of indicated proteins. Control samples were generated by transfection of 293T cells. (c) MRC-5 cells were infected with AD169-varL (an AD169 variant that contain the UL*b'* region [including UL138]) or AD169-varS (an AD169 variant in which the ULb' is

deleted [ΔUL138]) and labeled with 35S methionine/cysteine for 2hr before lysis. The IP was done with anti-UL138 serum. The observed protein sizes of ORFL312C and ORFL313C_UL138 fit the predicted sizes (Table S5).

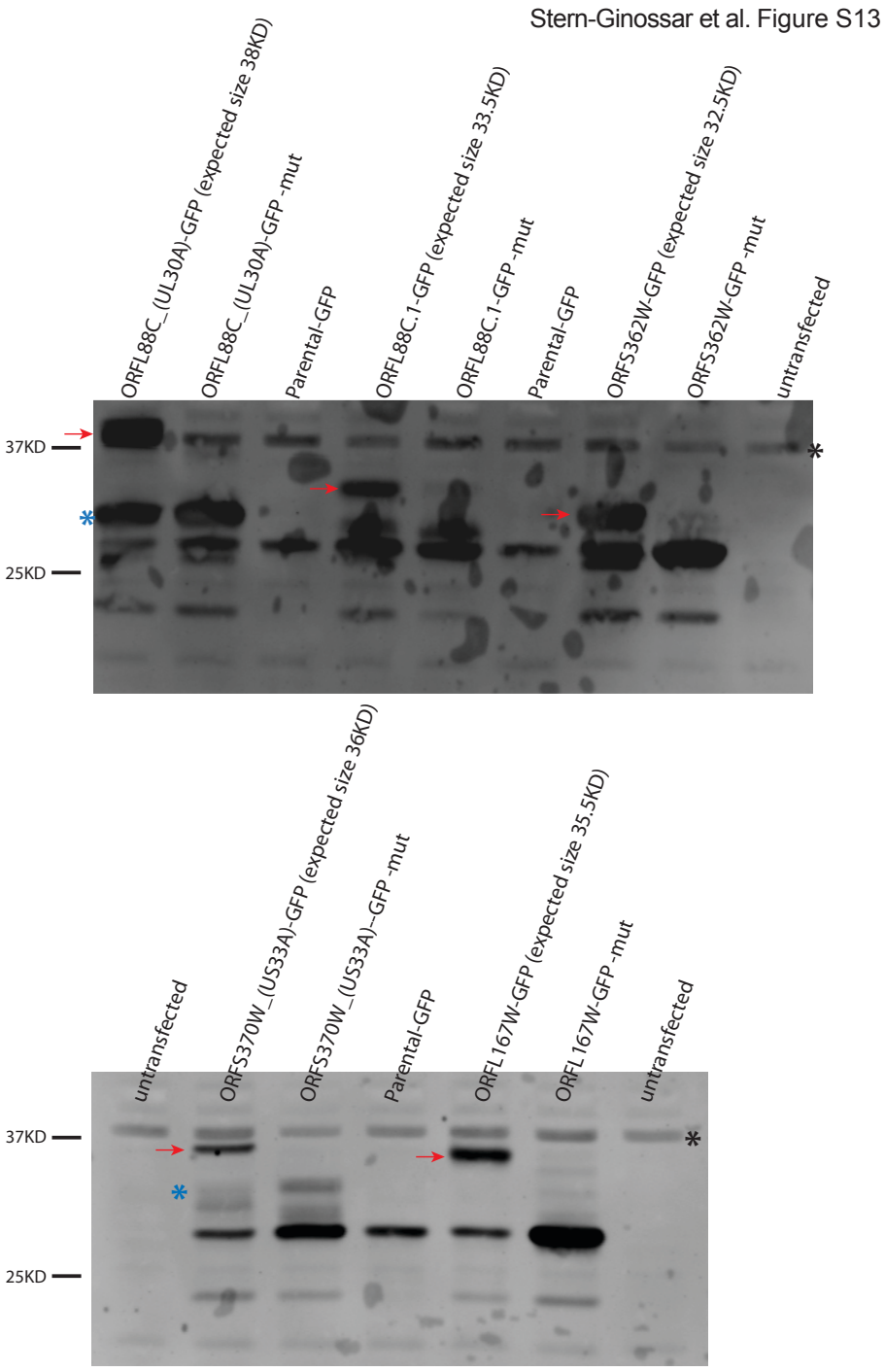


Fig. S13
Fusion of HCMV short ORFs to a GFP-tag. Five newly annotated short ORFs were tagged with a GFP using their natural transcript context or a mutated version in which the predicted initiation site was out of frame (by insertion of an extra nucleotide). The expected protein fusion products are marked with red arrows. Additional expected products that result from alternative in frame initiation are marked with blue asterisk and unspecific products are marked with a black asterisk.

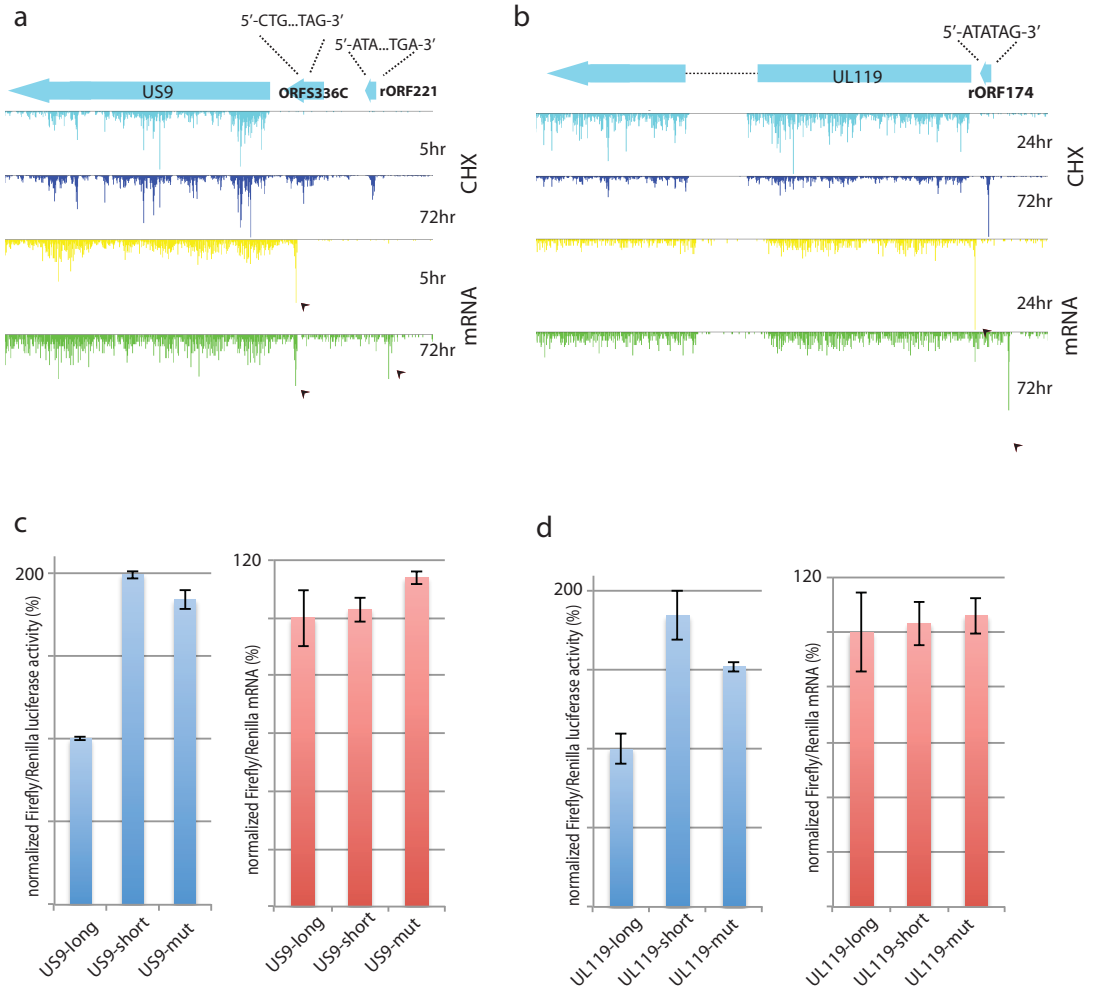
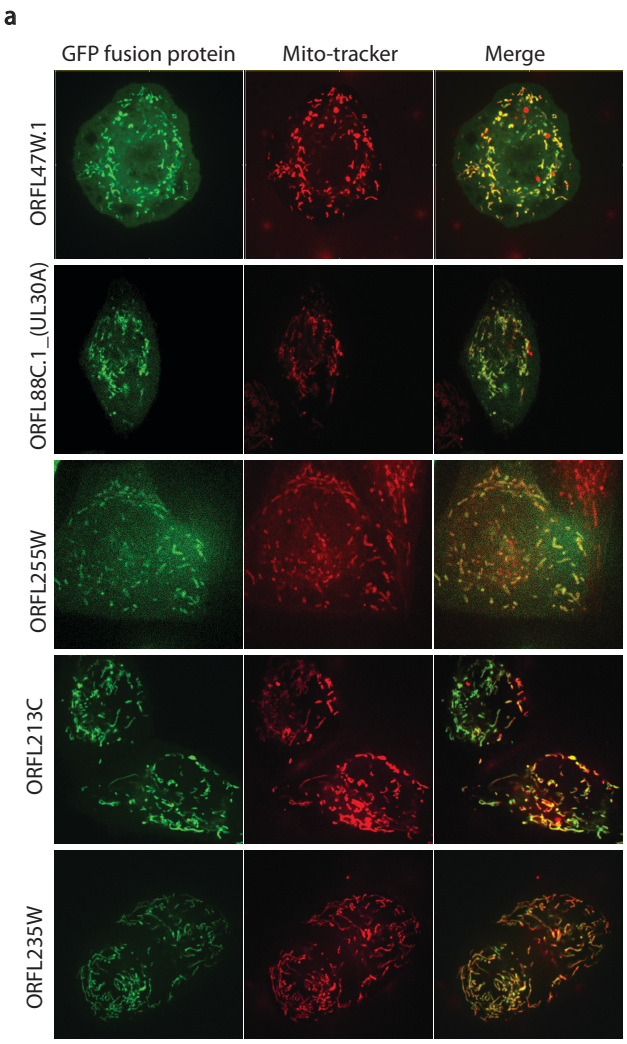


Fig. S14

Near cognate uORFs affect downstream translation. (a, b) Changes in transcripts 5' ends during infection that lead to inclusion of uORFs upstream of (a) US9 or (b) UL119. (c, d) The long, short (the mRNA isoform that does not include uORFs) and mutated (a long UTR in which the near cognate codons were mutated) versions of the 5'UTR together with a portion of the coding region of (c) US9 or (d) UL119 were fused to a *firefly* luciferase ORF. HeLa cells were transiently co-transfected with the cloned *firefly* reporter plasmid together with a *Renilla* plasmid (as a control for transfection efficiencies). The left panels demonstrate luciferase activity normalized to the average activity of the long 5'UTR. The right panels show the relative mRNA levels measured by real time PCR. Values are mean ±s.d. for triplicate samples.


a
GFP fusion protein
Mito-tracker
Merge
ORFL47W.1
ORFL88C.1_(UL30A)
ORFL255W
ORFL213C
ORFL235W

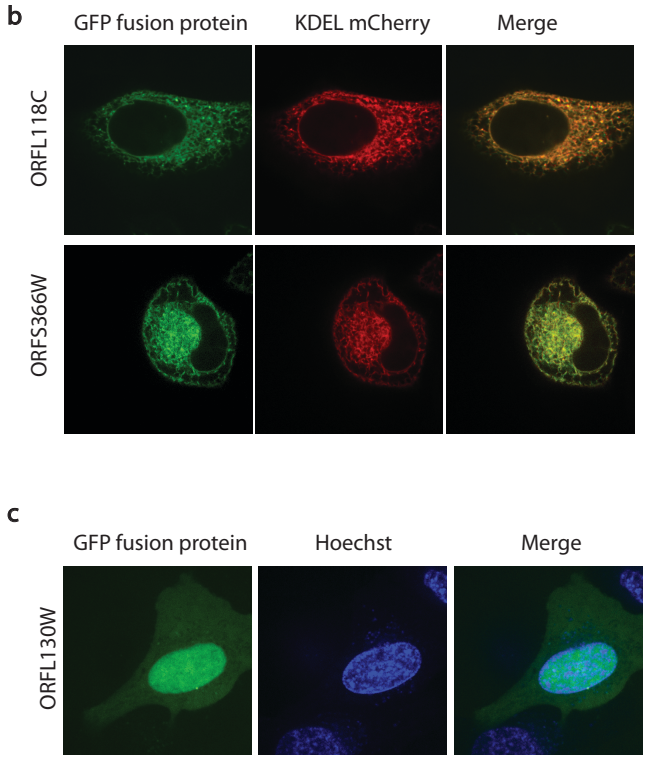


Fig. S15
Subcellular localization of HCMV GFP-tagged short ORFs. HeLa cells were transfected with GFP fusion proteins together with an ER marker (KDEL-mCherry) or stained with either MitoTracker Red or with Hoechst and imaged by spinning disk confocal microscopy.

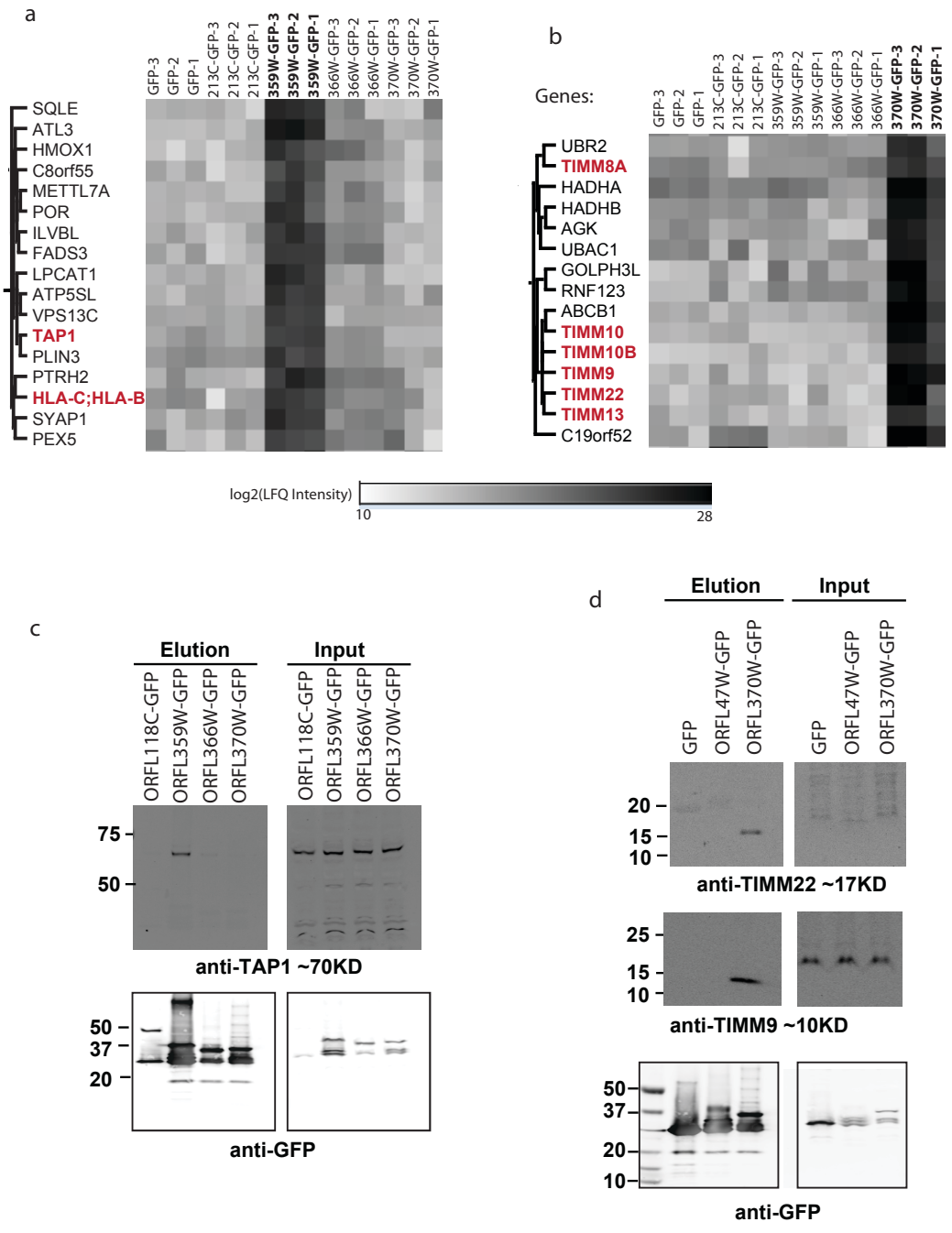


Fig. S16

HCMV short proteins Co-IP specific host proteins. (a,b) Heat map representation of interacting proteins identified by IP-MS. Four short HCMV proteins fused to GFP were used as baits; a GFP-only construct served as control. Each IP was done in triplicate. The heat map shows proteins specifically enriched along with (a) ORFL359W (ER-localized)

and (b) ORFS370W_(US33A) (mitochondria-localized). Colors in the heat map represent the intensities of the proteins in the respective samples as calculated by the label-free quantification (LFQ) algorithm of the MaxQuant software (*15*). The genes labeled in red are part of a complex for which the interaction was confirmed by western.

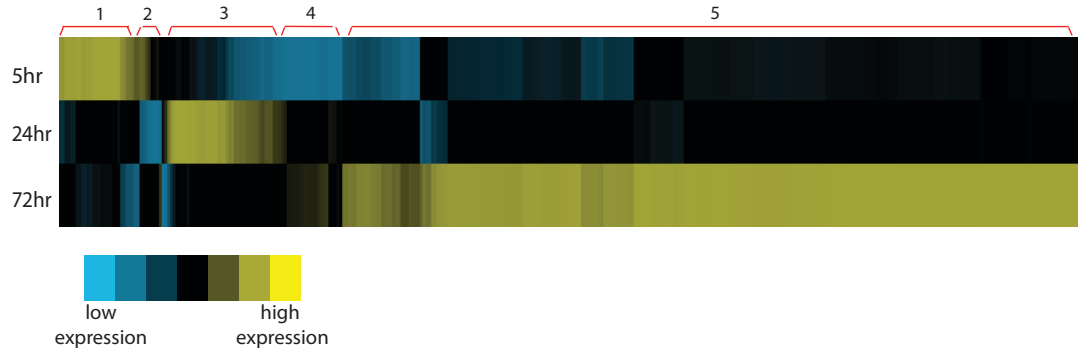


Fig. S17

Strong temporal control of viral proteins synthesis during infection. Normalized ribosome counts were calculated for each previously annotated ORFs and newly identified ORFs (for which we had non-overlapping measurements of ribosome densities). The footprint counts of each ORF across the three time points was normalized and then the ORFs were subjected to hierarchical average clustering based on uncentered correlation. Data is represented using colorimetric scale. The five main clusters are marked.

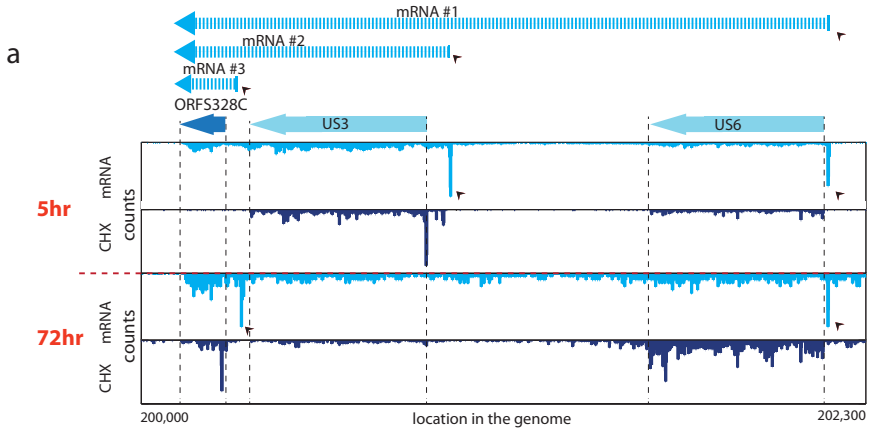


Fig. S18

The mRNA and ribosome occupancy profiles are shown around US3-US6 loci at different time points during infection (marked on the left). Small arrows denote the different mRNA starts and the corresponding mRNAs are illustrated (upper part). Note that in the change in the transcript 5' end reveals a downstream short ORF (ORFS328C).

a

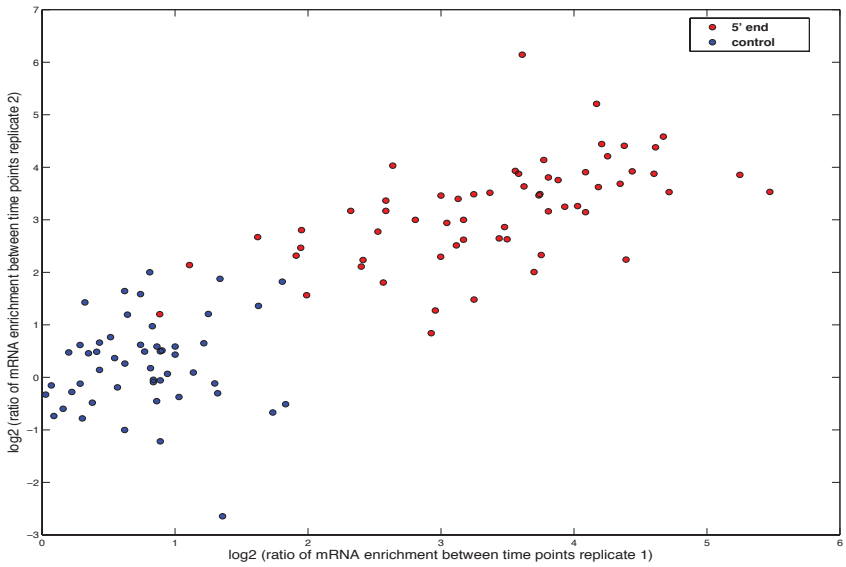
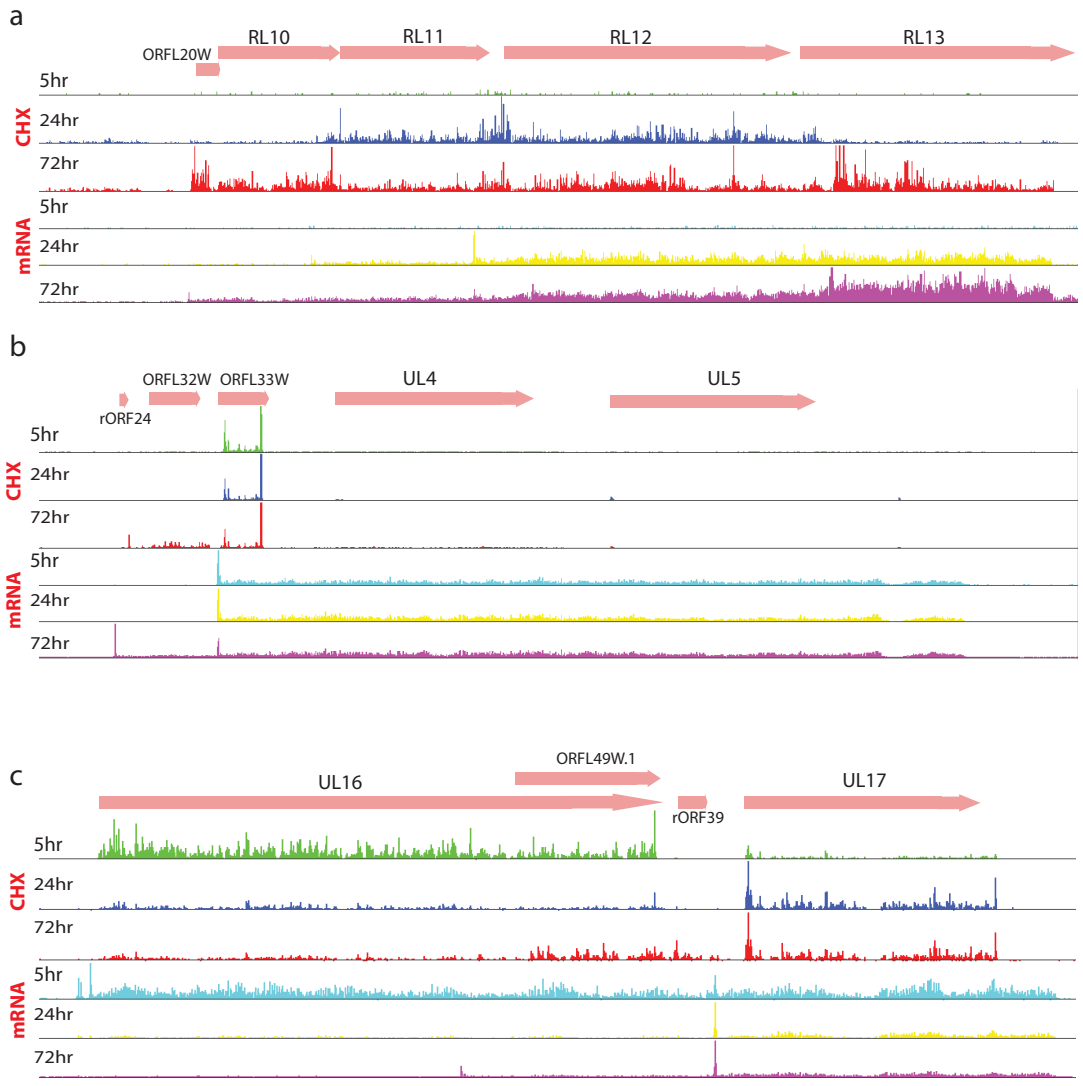
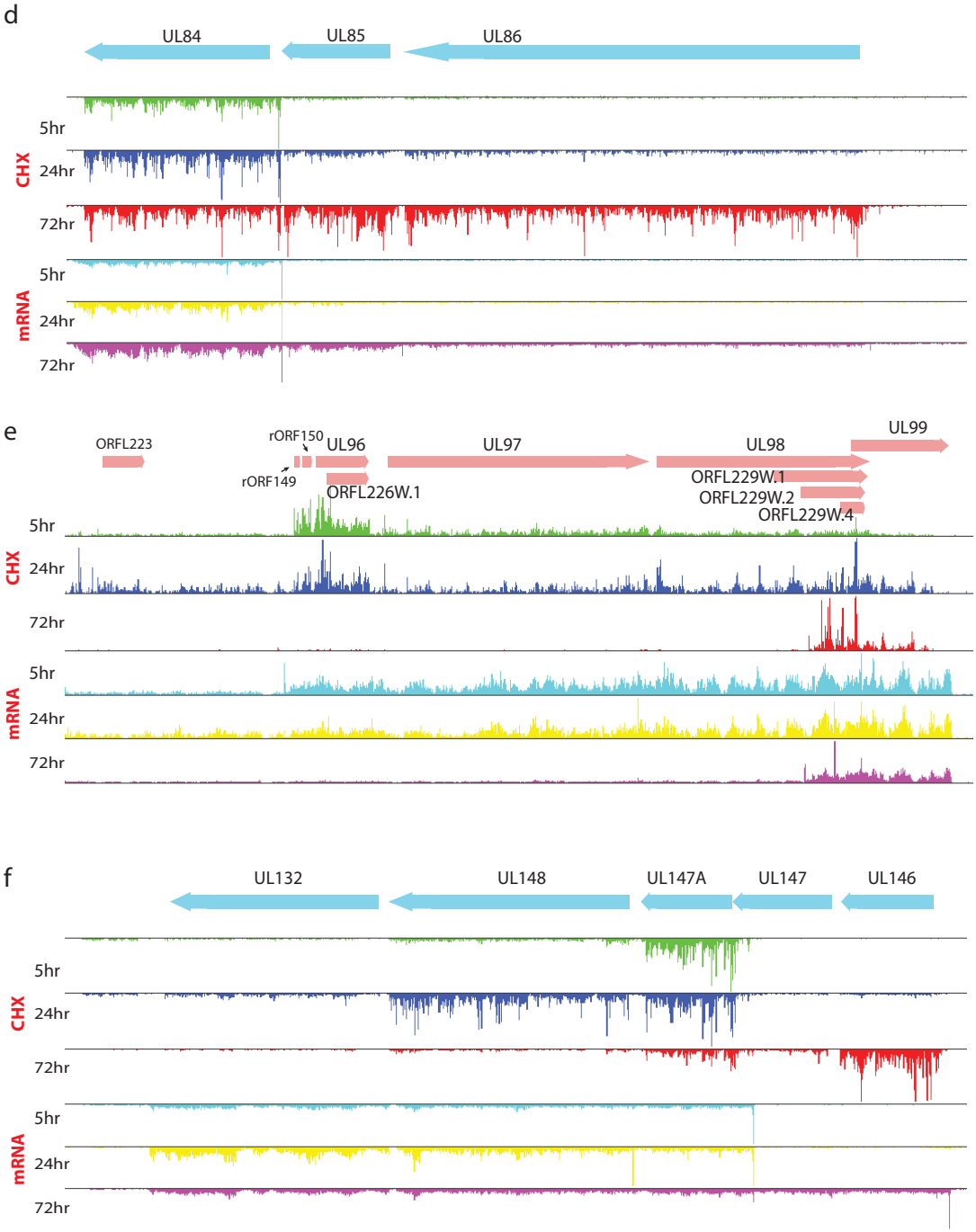


Fig. S19

Analysis of reproducibility in changes of mRNA 5' ends during infection. The ratio of mRNA 5' end enrichment between two time points is depicted for 61 5' end positions we report as changing (blue, table S10) and for a control position (10bp downstream) for the two mRNA biological replicates.


a
ORFL20W
RL10
RL11
RL12
RL13
5hr
CHX
24hr
72hr
5hr
mRNA
24hr
72hr
b
ORFL32W
ORFL33W
UL4
UL5
rORF24
5hr
CHX
24hr
72hr
5hr
mRNA
24hr
72hr
c
ORFL49W.1
UL16
UL17
rORF39
5hr
CHX
24hr
72hr
5hr
mRNA
24hr
72hr


d
UL84
UL85
UL86
CHX
5hr
24hr
72hr
mRNA
5hr
24hr
72hr
e
ORFL223
rORF150
rORF149
UL96
ORFL226W.1
UL97
UL98
UL99
ORFL229W.1
ORFL229W.2
ORFL229W.4
CHX
5hr
24hr
72hr
mRNA
5hr
24hr
72hr
f
UL132
UL148
UL147A
UL147
UL146
CHX
5hr
24hr
72hr
mRNA
5hr
24hr
72hr

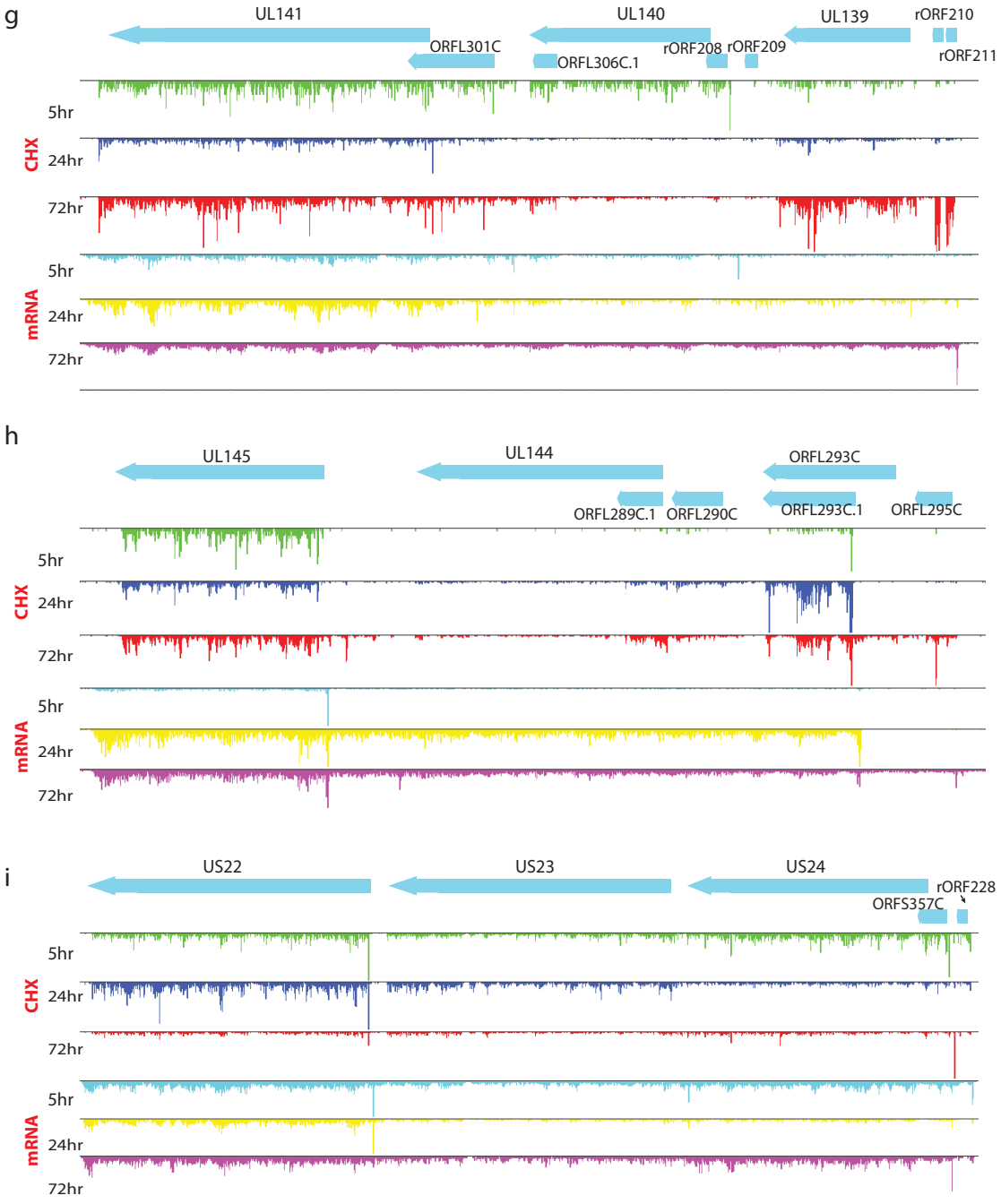


Fig. S20

Examples of nine genomic loci that show changes in the transcripts 5' ends during infection. (a-i) The mRNA and ribosome occupancy (CHX) profiles at 5, 24 and 72 (hpi) for each genomic locus are shown. The most highly expressed ORFs in each loci are illustrated in each panel.

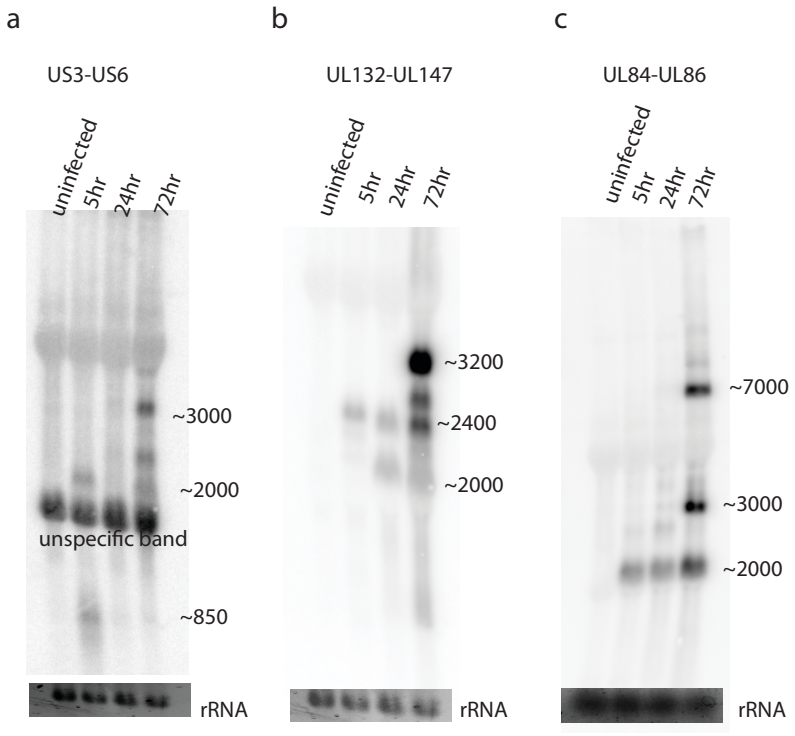


Fig. S21

Northern blot analysis showing the changes in transcripts 5' ends during infection. Total RNA extracted at the different time points along infection was subjected to Northern blotting using a probe against the 3' end of (a) US3, (b) UL132 and (c) UL84.

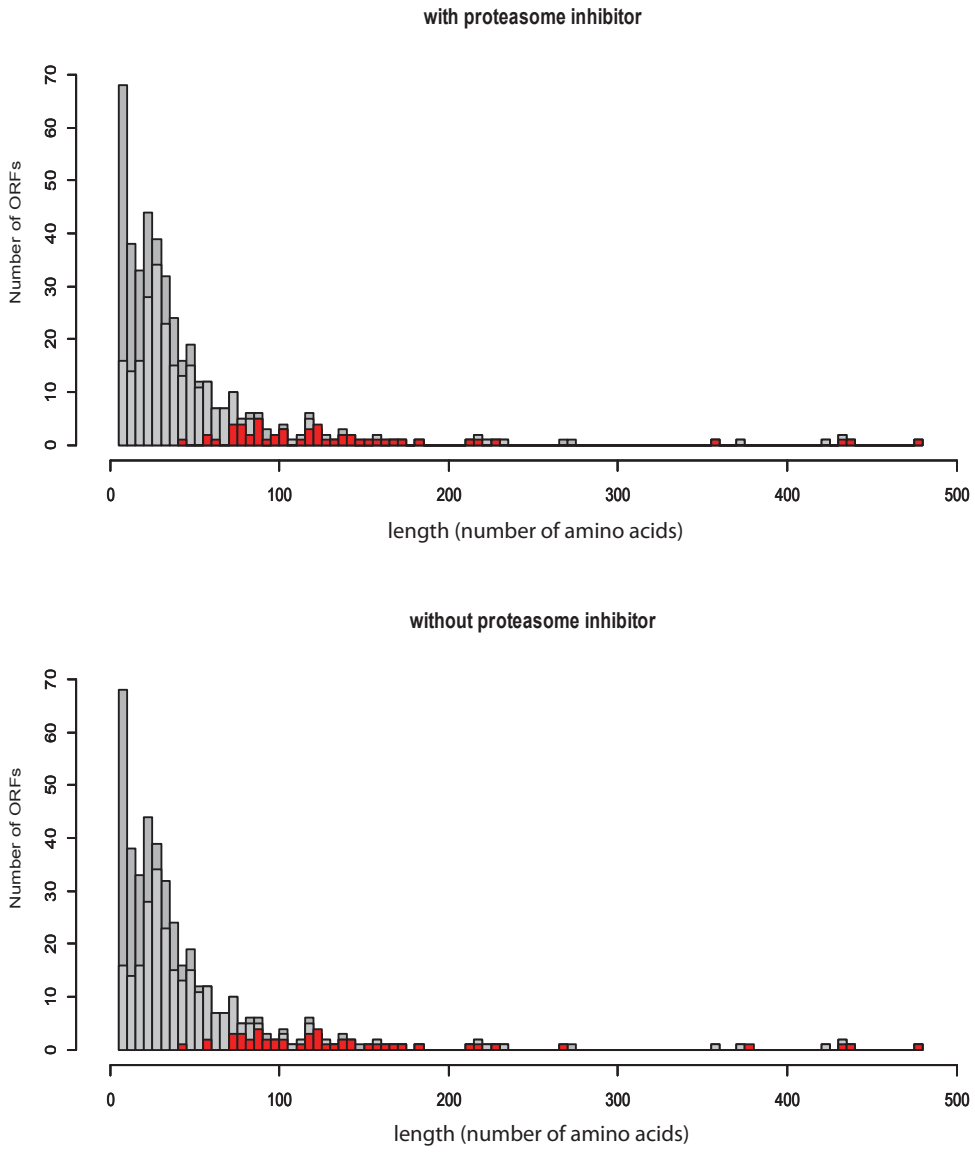


Fig. S22

All non-overlapping novel ORFs were binned by length (dark grey) and filtered according to presence of at least one tryptic peptide (> 6 aa) after *in-silico* digestion of the protein sequence (light grey); Non-overlapping ORFs that were identified by mass-spectrometry are marked in red.

sample	total reads	reads >= 25bp after linker & polyA removed	reads after rrna removed	total viral reads	subset of viral reads that aligned uniquely	total human reads	length of rawreads (mean and stdev)	length aligned reads (mean and stdev)
mRNA 5hr	25171983	25099258	15830591	882080	879063	12487354	μ=56, σ=7.7	μ=54.94, σ=6.52
mRNA 24hr	23852532	23758082	14543666	1288522	1283493	10864037	μ=56, σ=7.3	μ=54.95, σ=5.77
mRNA 72hr	19704160	19335706	10974346	3933717	3926633	4771572	μ=57, σ=7.8	μ=55.5, σ=5.84
Footprints CHX 5hr	23804050	21828252	11139771	507841	505173	7718906	μ=31, σ=3.2	μ=30.5, σ=2.12
Footprints CHX 24hr	23192119	18803120	12961846	864060	856464	7436294	μ=33, σ=3.6	μ=31.5, σ=1.84
Footprints CHX 72hr	26395689	25687165	18667511	8142886	8127748	8565081	μ=31, σ=2.5	μ=31.2, σ=2.02
Footprints Harr 5hr	25038467	24534819	6586917	191974	190857	4563399	μ=31.8, σ=2.52	μ=31.8, σ=2.52
Footprints Harr 72hr	26145020	25583705	9822847	3132481	3122138	5345658	μ=31.43, σ=2.29	μ=31.43, σ=2.29
Footprints No_drug 5hr	21624068	21164526	15097553	669201	665577	12292199	μ=31.28, σ=2.35	μ=31.03, σ=1.93
Footprints No_drug 72hr	20673613	19401500	8957410	2755833	2750338	4677701	μ=30.22, σ=2.48	μ=29.8, σ=1.90
Footprints LTM 5hr	55983543	53921028	10534812	729639	724531	6019751	μ=29.44, σ=2.97	μ=29.48, σ=2.02
Footprints LTM 72h	46160824	43927410	12819689	4745464	4731572	4680448	μ=30.45, σ=3.1	μ=30.57, σ=1.98
Footprints CHX 72h no rRNA_substraction	73590982	71201049	25070597	5962970	5949176	6177136	μ=32.54, σ=4.6	μ=30.44, σ=2.25
Footprints PatA 72hr no rRNA_substraction	19643696	19286069	2101388	262250	261586	1054893	μ=32.35, σ=3.5	μ=32.69, σ=3.5

Table S1.

A list of the number of sequencing reads we obtained from each sample. The first column indicates the sample identity, the second column lists the total number of reads we obtained from the sequencer, the third column lists all the reads that passed the process of linker sequence and polyA removal and that were greater then 25bp, the fourth column lists the number of reads after removing reads that aligned to the rRNA, the fifth column lists the reads that aligned to the Merlin genome, the sixth column lists the number of reads that mapped uniquely to viral genome, the seventh column lists the reads that aligned to the human genome, the eighth lists the mean (μ) and standard deviation (σ) of raw reads length, the ninth column lists the mean (μ) and standard deviation (σ) of aligned reads length.

Name	5' position in the genome	3' position in the genome
rORF20	13086	13139
ORFL33W	13612	13680
ORFL35W_(UL5)	14358	14858
ORFL36W_(UL6)	14965	15807
ORFL37W_(UL7)	15865	16533
ORFL40W_(UL8)	16573	16941
ORFL45W_(UL11)	18642	19460
ORFL66W.iORF1	29324	29416
rORF49	29900	29929
rORF59	38009	38029
ORFL97W	46408	46566
ORFL116W_(UL47)	61677	64628
ORFL120W_(UL48)	64625	71350
ORFL137W.iORF2	76968	77093
ORFL141W_(UL53)	77086	78216
rORF91	78202	78216
ORFL189W_(UL77)	112227	114137
ORFL199W.iORF1	117410	117520
rORF130	118582	118626
ORFL215W_(UL88)	132484	133773
ORFL229W.iORF2	145365	145769
ORFL243W_(UL105)	153233	156103
rORF215	191906	191938
ORFL321W.iORF2	192023	192325
ORFS362W	223890	223967
rORF174	169344	169349
rORF221	205109	205120
ORFL122C	66722	66814
ORFL129C.iORF1	71792	71872
ORFL143C_(UL54)	78194	81922
ORFL145C_(UL55)	82066	84789
rORF0094	85108	85113
ORFL185C	111134	111490
ORFL212C_(UL86)	125490	129602
ORFL213C	131254	131436
ORFL224C.iORF2_(UL89)	133770	139697
ORFL237C	149340	149699
ORFL255C_(UL115)	165022	165858
ORFL265C.iORF2	170689	171873
ORFL297C_(UL142)	183406	184323
ORFL313C_(UL138)	187441	187950
ORFL319C_(UL148B)	191190	191432
ORFS326C.iORF1_(US1)	198543	199013
ORFS329C_(US3)	200345	200905
ORFS347C_(US19)	213551	214273
ORFS351C_(US21)	215191	215922
ORFL104C_(UL37)	50262	53060
ORFL253W_(UL112)	161917	164128
ORFL248W_(UL111A)	161003	161692
ORFL53W_(UL20)	25640	26656
ORFL107C_(UL41A)	54401	54637
ORFL184C_(UL75)	109224	111452
ORFL229W_(UL98)	144015	145769

Table S4.

ORFs that were manually added to the SVM-based annotations are listed by their names and positions in the genome.

ORF name	Comments
RL5A	ORFL11C has an alternative 5' end
RL6	Weak expression, the LTM treatment marks a clear initiation at the predicted ATG
RL9A	Weak expression, the LTM treatment marks a clear initiation at the predicted ATG
UL1	Very weak expression.
UL9	Very weak expression
UL10	Very weak expression
UL18	Very weak expression
UL24	ORFL71 has an alternative 5' end
UL31	ORFL87 has an alternative 5' end
UL49	
UL70	ORFL166C has an alternative 5' end
UL79	Very weak expression
UL87	Very weak expression
UL92	Very weak expression
UL93	ORFL221W has an alternative 5' end
UL95	Very weak expression
UL102	Very weak expression
UL117	ORFL257C has an alternative 5' end
UL120	
UL148C	Very weak expression
IRS1	Repeat region
US29	Very weak expression
US34A	
TRS1	Repeat region

Table S7.

This table lists all the ORFs that were previously suggested to be coding and we did not detect in the current work. The first column lists the ORF name, the second column contains comments regarding these ORFs; indicating whether a weak expression is observed, pointing out to an ORF with an alternative 5’ end that contains the main translation initiation or indicating whether this ORF is located in repeat regions and was therefore not detected due to alignment issues.

Without proteosome inhibitor

	Name	Protein sequence	Length (aa)	Sequence coverage [%]	Identified peptides	Andromeda Score	PEP
1	ORFL103C_(vMIA)	MSPVYVNLLGSVGLLAFWYFSYRWIQRKRLEDPLPPWLRKKKACALTRRSRHRLRRQHGVIDGENSETE RSVDLVAALLAEAGEESVTEDTEREDTEEEREDEEEENEARTPEVNPMDAEGLSGLAREACEALKKALRR HRFLWQRRRRARLLQHNGPQQSHQ*	163	9.3	IIQHNGPQQSHQ	200	3.71E-24
2	ORFL106C	MANSLWHRLRTLVSNSRQNEQIQQHSYRLHLRGYGSADFNSDGWTPVYEDQEFIVFSCRDDGNNRRR DVRSRSAMSARVPGLGVRRVPPRRHGPHDARRVLLPPDLQQHRRRVLASSCR*	119	31.9	ANSIWHR QNEQIQQHSYR RVIIPPDIQQHR VIIPPDIQQHR VIIPPDIQQHRR VPGIGVR	177 92 211 147 45 167	2.41E-05 1.11E-03 5.55E-26 3.28E-07 1.55E-02 6.42E-04
3	ORFL114W	MTLFAHQHAHSLYRRRYFAERSPEIVLLLDTVQLHPGRAGRIKGAQTATHASTATGHGGTAVTRQRQHR RRAVGRRRLGTPTDDAATTRKTRYTCYTARYVGISSGRFLVTPGQRPLLEAVISPDSISGDLCRRFLRLGF RAAAWNPTWPAPMCGHRGHPLFLVRVSRIHKFTRKITKKSGRFQ*	185	16.8	IKGAQTATHASTATGHGGTAVTR GAQTATHASTATGHGGTAVTR GHPIFIVR	102 122 93	9.98E-13 3.07E-13 8.20E-03
4	ORFL118C	MHQKHRVQFGVGVAPKQHVGHERAGPVDEPLVGQVQEFGHGATIDVGPRLRGSRFQLRDHPADQRL LAPGTRTALLAAALGRRVVTGDQVRRRRGAAALFGYYGLDLAQGLQQLLESRQRGPVVVERQPVLVSA AVVLVSQEAVQLFVQLLGGVQRVARRLYRGVA*	166	27.7	AGPVDEPIVGQVQEFGHGATIDVGPR TAIIAAAIGR VQFGVGVAPK	198 107 102	4.36E-74 4.82E-04 7.29E-04
5	ORFL119C	MSQIVSQTAGAAAHHVLDQRVDGGVQVMRYGARHELGQQVHIGRDGYTAEVLESRGDVEDAIAYRL KNRGVDVLDEGGQVVDGDQVLVQRHLLGQQYTAQLNECEKTRRTHGLKQALRDTKDLRELSYHHQLRE DAGELGMMAEG*	145	65.5	EISYHHQIR GDVEDAIAYR GVDVIDEGGQVVDGDQVIVQR HEIGQQVHIGR HIIGQQYTAQINECEK SQIVSQTAGAAAHHVIDQR VDGGVQVMR	97 112 125 165 233 172 118	1.46E-03 2.93E-04 1.54E-14 2.33E-09 6.40E-68 4.78E-36 5.67E-04
6	ORFL121C	MDDELLETIGEESFVEGRRRGRCRVSAAFGTRRVQRPGAGVGRQRRQRQSLQLLLYLVDQLLVAAQGR QGPFAAAVGRWYRGRRRGRRYR*	90	20.0	VSAAFGTR QGPFAAAVGR	136 98	3.30E-05 2.90E-03
7	ORFL130W	MALALESTQTPERQVDDAADEAAAQRGAECKEDDLGGLYVMRLLSGARVSYRRSCRRRASSARLVVQT QCPQRHQVSRTRVAKRLHRLSGQYVERVSAGYLPGVPTQGCDEAPQLRRPRLPLVTPLVTVVYGQISAG VEALGA*	143	72.7	AIAIESTQTPER AIAIESTQTPERQVDDAADEAAAQR EDDIGGIYVMR GAECKEDDIGGIYVMR IPIVTPIVTVVYGQISAGVEAIGA ISGQYVER IVVQTQCPQR QVDDAADEAAAQR VSAGYIPGVPTQGCDEAPQIR	161 236 85 190 120 101 210 149 132	2.91E-12 3.20E-101 1.76E-03 2.50E-37 1.95E-18 5.05E-03 5.83E-20 2.00E-11 1.43E-16
8	ORFL131W	MKDARTWHVMHQLLLEDRRRLRRRLRRWLRRFYRRFLRLQRTAADHQLHMEMVLVNAEGREETAQSL VELISRGVRSVSETTIGHDRATAEPAPAEEPAPAGPESQSARAATAQTAVAMARSALATTMTVPNDNQ SAPRTACTAETADDGDGSTPLADDSRAPPEADARTLRCPGPPLPPSLLPPEPATPPPLHQRLPRLRIRIV PYRPHNDTAATTPPPWVAAAAEARQRRRQRRPWWESNSDDEEEEEGEMRSEGTAFPPFA*	266	19.5	ATAEPAPAEEPAPAGPESQSAR AATAQTAVAMAR ATAEPAPAEEPAPAGPESQSAR SAIATTMTVPNDNQSAPR	137 222 137 118	9.97E-21 6.65E-32 9.97E-21 1.12E-08
9	ORFL137W.iORF1	MARAPASSTYGRSAGRRGGRLIPGESSTPRQPLPSLRIPDAQPRHLGGRRTAQLLCGRNHVRRHAIHRT HQCDQHH*	76	34.2	IIPGESSTPR QPIPSIR RTAQIICGR TAQIICGR	67 91 56 129	1.65E-02 4.24E-02 4.76E-02 7.90E-04
10	ORFL142W	MSLESTDESTRSSTRTRCRYRSSRACTRRRKIRQKPRGTPYFLTSSSERRDRRTAASETSRSSKSSERTSRVT PLTNSQAVLRTRSTPFMLRPEAPSTLPM*	101	49.5	GTPYFITSSSER SIESTDESTR STPFMIRPEAPSTIPM VTPITNSQAVIR	123 66 118 125	1.70E-05 1.83E-02 8.97E-08 1.35E-06
11	ORFL147C	MSLAGARPDDSVSYVSESSHGDEFVTETMRSVFEMQRIRHGAGVSKVLRSERVTGGVHAVQEKRGGYS	476	42.6	AIPVQNEIR AAAGAPPSPMR CVSPAVPGAHER DDMRPMIR DPGAGGRPGGIR DVHQNQAYSR DVHQNQAYSRGDSTSAAR GADHHPESGPIAEQAAAGIAVQPYSGPPSVKPVR GAPVYGQRPFQR GQPGVQQAPR GYGFAGSGADGIIAPAGPGGIIPYRPPFGSIR HIAGGASGSVR HIAGGASGSVRR IRHGAGVSK SERVTGGVHAVQEK SERVTGGVHAVQEKR SVFEMQR VRDPGAGGRPGGIR VTGGVHAVQEK VTGGVHAVQEKR	125 131 148 101 138 59 106 230 126 81 199 123 112 107 107 152 114 150 180 150	3.98E-04 1.99E-05 4.12E-09 6.37E-03 1.33E-07 4.22E-02 3.10E-06 5.40E-124 1.21E-06 4.03E-03 2.35E-88 4.53E-05 5.09E-05 1.19E-03 2.53E-05 1.27E-14 1.63E-02 3.96E-12 1.09E-11 1.78E-09
12	ORFL149C.iORF2	MRPFLERGDRSSVGGAHGGGGFRRLCAYACALLRRRGAHESHVLLSFCALLSSYARHRGFYRGSRRRAS AVRGRPPEVRVGGLHARGGSRAYRSGGVVRGGGLSGVRDDCARGGGQRLEGVSLCRSADSMRGGGD HGATARRRGGARAVVPAYAL*	155	37.4	AVVPAYAI GGGISGVRDDCAR GGGQRIEGVSICR ICAYACAIIR IEGVSICR SSVGGAHGGGGFR SSVGGAHGGGGFRR	110 103 104 81 124 73 64	3.79E-03 1.50E-04 1.39E-04 3.70E-03 1.02E-03 3.86E-03 7.74E-03
13	ORFL151C	MALVDQESCLRGFYSVFLRHCDVEPAIATFDGVRSPPDVVRLYPKPRYDYTLASAFLGTAVSGYTTSRRD NAPNKNTGHRYVYTTSILIAVVTSVDTQHRSPSRCFMAMGESYFLQDEKCVFW*	123	14.6	AIVDQESCIR GFYSVFIR	133 108	4.58E-05 9.35E-04
14	ORFL154W	MYTELSLLYKKKKLSIFISSEFLLDRHVFGTLSPHVFTVYNKNMCFSKDLEVFENGGNLRLGALSPGLGRSR RPAPIFFYFFTKYMMNQE*	90	24.4	DIEVFENGGNIR IGAISPGIGR	68 89	7.27E-03 1.97E-03
15	ORFL159W	IIVFAVAGSSCGDAGSVMMRGGADGIFREMVFAGGCSVPCRRRGWTSLASELPWHGGGDPAPSVCL VVWFACVYSLLILVVLLIYRCCIGFQDDLVSRTLAVYRACIQGPICNQTHNSTS*	121	41.3	EMVFAGGCSVPCRR GGADGIFR	71 102	8.01E-06 5.02E-03
16	ORFL161C.iORF2	MIPKAIRTTILPFHPPPATHSGAPADTSSWALRASVADIGRPPLSSDAWPRPPRCPLTPTTVSLTTTMTA NRTITAAIRRIFVAEFPVVLRLSVLLTPPQPIIITDRPPDRNNNRCWRLLTTNWTPWTRTNYSNSAV*	138	84.1	ASVADIGRPPISSDAWPRPPR CPITPTTVSITTTMTANR IITTNWTPWTR ISVIITPPQPIIITDRPPDR RIFVAEFPVVIR TITAAIRR TTIIIPFHPPPATHSGAPADTSSWAIR	195 123 135 63 84 104 195	1.79E-49 2.99E-09 1.15E-05 4.78E-03 7.99E-04 3.12E-03 3.56E-71
17	ORFL174C.iORF1	MPGRAIRLHARAAGIRAASPAVGCRQPQRHFSVHLSGHHRRRYRRFPLFCARTLAAHELAAAQRTPFLFI TYPSDAGDRSGSHANRLF*	88	38.6	FPIFCAR TIAAHEIAAAQR TPFIFITYPSDAGDR	75 119 120	7.37E-02 3.80E-05 2.13E-07
18	ORFL174C.iORF2	MKRRRHYRRRSKFPTSGCLPVSSSNNATRSWPFTINTSFGYHASAKCPPRWASASISPKVFSALPFTNA WTRNSFACPSCWNPGCKSRSWMSSTSTTLFSPSFRAPSKAISAFSPVLCQNPGSS*	124	70.2	AISAFSPVICQNPGSS NSFACPSCWNPGCK SKFPTSGCIPVSSSNNATR SWMSSTSTTIFSPSFR VFSAIPFTNAWTR WASASISPK	87 128 132 114 167 57	3.44E-04 1.91E-08 1.97E-13 1.10E-07 4.42E-16 6.83E-02
19	ORFL185C	LDHAAEPSGAAALCGQASPPTSSSSPSVSSATYFRHDMAQKPYPNRWTKRFTYCSTPTGDPSASCVKIPP SVPTTAASVTARSSGKTPSVSTFSKAIINTMYSICLDVFLRVLWRSSF*	118	48.3	HDMAQKPYPNR FTYCSTPTGDPSASCVK IPPSVPTTAASVTAR RFTYCSTPTGDPSASCVK SSGKTPSVSTFSK TPSVSTFSK	82 110 138 86 54 130	2.47E-03 3.21E-06 1.34E-10 3.08E-05 3.55E-02 4.09E-04
20	ORFL18W	MRQSIGLDERDAATPLHGRLVVVVKEEHQQSQEEQQALTCPGGRGEGLSEMAVRHHVRQLRRNGEER KITVRGRGSY*	77	13.0	GEGISEMAVR	79	3.43E-03
21	ORFL190C	MSTSSRQPPPRIATAASSSNRRSTRSGQRSTFSSCGSKKTATGSRQKVCNRLILTPSQIVQNRIGTQVETA RAKSPVKRWAIGLLLKKNQTRHYAPDLTATARQTQKIKVHQGALELAQALGVLQAVGGVPPHGLIVGV*	139	23.0	HYAPDITATAR IGTQVETAR IIITPSQIVQNR	163 84 105	4.22E-09 5.43E-03 2.53E-04
22	ORFL206W	MARVTGLVFSVVFSVSLQSCGSVSGDASRMGSRSSSAMEAPSTESCCCDGEWDVDRECSVDLTWCVI ADRQPRFSVCFDEIK*	82	11.0	FSVCFDEIK	99	1.34E-03
23	ORFL211W	MSGVVEYVAQTAGLRGPGVVCGRESLCVRFAMVVDGLGGGLRVVRIDGQHAGGTPARVVGDLKVIN VHRHRRKD*	74	36.5	FAMVVDGIGGGIR SGVVEYVAQTAGIR	78 191	2.31E-03 2.05E-30
24	ORFL223W	LAFLGRVVVAVAAAEGHERQEGLVARVGQPEFGVHHLVALAAQRRHIGEGRGDAGGGVALALLQLRV ALLDLGGRISAQHDGGGGGASGG*	90	68.9	GDAGGGVAIAIIQIR VAIIDIGGR VGQPEFGVHHIVAIAAQR VVVAVAAAEGHER VVVAVAAAEGHERQEGIVAR	179 144 114 165 85	2.87E-66 3.41E-04 1.83E-08 1.51E-15 4.31E-06

25	ORFL224C.iORF1	MRPLTCSTWSLTCAKSTCTASRRRATPRRVLATDCTSPPSSASTRRCARRPTCLCRALSWTRSSAVPIKSR RTPCSSRTRAAKSSIFCVTARSTPTPTIISARRFTCIWTRPSPPTARPRARAWRP*	126	53.2	FTCIWTR	100	3.20E-02
					FTCIWTRPSPPTAR	93	2.24E-04
					MRPITCSTWSITCAK	169	1.30E-25
					RVIATDCTSPPSSASTR	75	3.64E-04
					SSIFCVTAR	127	3.96E-04
					STPTPTIISAR	81	2.83E-03
					VIATDCTSPPSSASTR	131	2.37E-11
					VIATDCTSPPSSASTRR	75	3.82E-04
26	ORFL234C	MSTRCSVASRTCTVKVASSLGGCAPPVLLLTSTTLFSPSPSSRGSGGNSQRRSPPAGRSRKAVALRRTAPP PASVPKTCHSSSGTPSESAQGSTSSLPAEPTMSGRPHHESTSAAASNAVSRINVCRPWTA*	131	58.8	INVCRPWTA	126	4.94E-04
					RTAPPPASVPK	74	4.45E-03
					TAPPPASVPK	55	5.71E-02
					TCHSSSGTPSESAQGSTSSIPAEPTMSGR	92	9.41E-09
					VASSIGGCAPPVIIITSTTIFSPSPSSR	221	3.55E-95
27	ORFL235W	MRFVKSLLKICVFARDAPSGRSRVCWVTSPSTLRQPALRPAFCGFCRVATASFMSTTAIRVTDPSRPKCA CGLWSTVIGANFLAATIRVPPVVFKRARPACCWSGATSGWWLLLRGRWRLALGRGSCPGPAAGAVTG ARERVRAGSRVSRDPLSLRGSPQRRGLLVTALPAR*	172	8.7	GSCPGPAAGAVTGAR	69	2.78E-03
28	ORFL237C	MAVSDDAARATSATAMSRRETTHHPVERTTTGAAVAVEEVEGTHQPLVAPDQQHAGRALLNTTGGT RIVAAKKFAPITVDHSPHAHFGRDGSVTRIAVVDIKLAVATRQNPQKAGRSAG*	119	58.0	AIINTTGGTR	146	1.68E-05
					FAPITVDHSPHAHFGR	106	1.23E-05
					IAVVDIK	100	3.22E-02
					IAVVDIKIAVATR	68	6.76E-03
					KFAPITVDHSPHAHFGR	79	2.80E-04
					TTTGAAVAVEEVEGTHQPIVAPDQQHAGR	201	1.54E-87
29	ORFL251W	TASDERCARALRRRVMDLPTTVVRKYWTFTNPNRILHQSVNQTFDVRQFVFDNARLVNCVDGDGKVL HLNKGWLCATIMQHGEASAGAKTQQGFMSIDITGDGELQEHLFVRGGIVFNKSVSSVVGSSGPNESAL LTMISENGNLQVTYVRHYLKNHGESSGGGGGCGAASTASAVCVSSLGGSGGTRDGPSAEEQQRRRQE QRHEERRKKSSSSAGGGGGGGASGGGGGGGGSGGQHSSDSANGLLRDPRLMNRQKERRPPPSSENDG SPPLREAKRQKTTAQHEGHGGGGKNETEQQSGGAGGGGGGSGRMSLPLDTSEAVAFLNYSSSSSAV SSSSNNHHHHHHHHNAVTDVAAGTDDSCQRRRAGLSSSTTAASRGPPQLSPQPALAPQQSHVRHSSP PVLYPQVPSPVSRPLPPQSKHQLLLPGPCVKSSPTW*	438	55.7	VMDIPTTVVR	63	1.61E-02
					AGISSSTTAASR	79	2.85E-03
					GPPQJSPQPAIAPQQSHVR	183	6.50E-46
					HQIIIPGPCVK	115	1.06E-04
31	ORFL253W.iORF3	MARCCTSTRAGSALPLCSTARLRLAPRRSRASCPLTLRATGNFRSTSLYAAVSSLTNPSPRWWAPADPM RARCSL*	75	50.7	AGSAIPICSTAR	107	2.03E-04
					STSIYAAVSSITNPSPR	99	3.36E-05
					WWAPADPMR	57	6.73E-02
32	ORFL26W	MKDRNKILLCIIFICIMCLICIYFKRHCIPTPSPDKADFRVEFPSLLPCIGVQCAA*	56	46.4	HCIPTPSPDK	159	
33	ORFL271C.iORF1_(RL8A)	MLRIHKRQPPAGTLLRFQNVQSLHRRAAVSGRRSLLQSRENG*	42	47.6	FQNVQSIHR	174	3.80E-08
					FQNVQSIHRR	117	1.55E-04
					QPPAGTIIR	96	1.73E-03
					RQPPAGTIIR	98	3.72E-03
34	ORFL27C	MTFKLPYVMPHGHLRQALSPTSWTCEGLLLLLGLLVLFFHHHNQSAVERRRRVSFVEADRLPHESGWYS SDDDGDRDGDEETGESHNRNSVGLSAVFS*	98	30.6	IPYVMPHGHIR	83	1.91E-03
					NSVGISAVFS	166	6.25E-07
					RVSFVEADR	55	4.92E-02
					VSFVEADR	57	9.05E-02
35	ORFL294W	MDILTQRPKQRRDFGDGLVLTAAESVAGPARQQRATQTEYLEESYVCHRRARVSGGNDGGVVDTCAA CCALRKSAGLGDRGRKRERSSYRWRRRR*	95	71.6	MDIITQRPK	120	4.34E-04
					ARVSGGNDGGVVDTCAACCAIR	73	4.83E-05
					ATQTEYIEESYVCHR	198	3.86E-43
					ATQTEYIEESYVCHRR	121	2.54E-08
					DFGDGIVITAAESVAGPAR	269	1.15E-123
					RDFGDGIVITAAESVAGPAR	181	1.35E-38
					VSGGNDGGVVDTCAACCAIR	198	1.36E-60
					VSGGNDGGVVDTCAACCAIRK	180	5.55E-39
36	ORFL294W.iORF1	MDILTRPKQRRDFGDGLVLTAAESVAGPARQQRATQTEYLEESYVCHRRARVSGGNDGGVVDTCAAC CALRKSAGLGDRGRKRERSSYRWRRRR*	94	8.5	MDIITRPK	126	6.58E-04
37	ORFL298W	MRVSSNPRRRRGRYRSSCVENRGTSRSRDSVDARASATRSGLPASSDLTPACSKVRRAG*	59	25.4	SGIPASSDITPACSK	84	4.54E-04
38	ORFL299W	MQLREIPSGVRQAVLKSIHQPMAFDGVQDELVAAVVGTTETCANGNHGPGVCHDRAGDGNLLSLGD QHRRRRRGLIVLGPHLFGDVSGGEGERRGRIFEESRAAQQHPEQPRQSSERLPAAHDDSFLPFFAETLPA CSCSVCRPLKRRAGVVVTTVRRSFSPGMPATERPVFLQVFFAASPSPSRPTWWTSTAQELSRRRRYATCL IATSDV*	212	45.8	AGDGNIISIGDQHR	213	2.29E-42
					AGDGNIISIGDQHRR	138	1.16E-11
					AGVVVTTVR	105	1.03E-03
					AGVVVTTVRR	120	1.24E-04
					GIIVIGPHIFGDVSGGEGER	166	4.68E-29
					GIIVIGPHIFGDVSGGEGERR	109	4.01E-11
					GRIFEESR	72	3.72E-02
					IPAAHDDSFIPFFAETIPACSCSVCRPIK	168	9.31E-48
					RAGVVVTTVR	87	1.51E-03
					RAGVVVTTVRR	97	1.22E-03
					RGIIVIGPHIFGDVSGGEGER	84	7.65E-07
					RYATCIIATSDV	85	7.67E-04
					YATCIIATSDV	92	1.06E-03
39	ORFL301C	MRQVAYRRRRESSCAVLVHHVGRDGEGEAAKKTCKKTGRSVAGIPGEKLRRTVVTTTPARRLSGRHTEQ EQAGSVSAKKGRKESSCAAGSRSELCRGCSGCC*	102	40.2	ESSCAVIVHHVGR	112	2.75E-05
					ESSCAVIVHHVGRDGEGEAAK	44	4.79E-02
					RTVVTTTPAR	81	2.68E-03
					SVAGIPGEK	88	4.12E-03
					TVVTTTPAR	80	7.91E-03
					TVVTTTPARR	111	2.06E-04
40	ORFL321W	IDGRYKATRTRAAALAYGAAPGKLLFLLLFLLHRRDRFPRPGTYRSNYRNNLEESRHDGAQVCHHHDLSG QDQGAALVARQRHQEMVDQRCHRHLHRSLSGGPDVLYAAAGTQRERQRLDKSQAATAPSAVAGPPA DRDVVDHRTESHAYETPSSVQNSDATLGMASCSPPRRVSPDLLVSSELRRRNIWRRVTCVTGERGSPAV TTAVSSTRVRPAVLCGKAPVSETVDEKENGAATAGGGGVNAAVDVRRCEYSETAVRQKLDPSGNDSRC VVAAADSSTDSDGKSIIAGVQVVDHDEDIAPQSLWCTAFKEALWDVALLEVPRWAWQGWKRWRNS ESGRRWSAGSASASSLSDLAGEAVGELVGSVVAYVILERLWLAARGWVCETGVEAEEAMARRRQRML WRMFSRGGDGECSRRCSMEMACEEESAVL*	435	25.1	SQAATAPSAVAGPPADRDVVDHR	32	1.48E-02
					CVVAAADSSTDSDGK	125	1.77E-08
					EAIWDVAIIEVPR	158	4.19E-15
					GSPAVTTAVSSTR	188	3.96E-24
					GWVCETGVEAEEAMAR	124	7.96E-09
					RVSPDIIVSSEIR	114	1.64E-05
					VRPAVICGK	118	6.99E-04
					VSPDIIVSSEIR	90	8.76E-04
					WAWQGWK	114	1.63E-02
41	ORFL47W.iORF1	MRLRWCVRPVCSRCATATASGTRAHRSCCLTATVWRWPASSPRTTGQRSLSESTSATGPRSCVRWW WTHAAVRCYTTTPAVTSPAGAGVSRTAARRSG*	98	50.0	CYTTTPAVTSPAGAGVSR	122	4.64E-09
					SCCITATVWR	71	2.23E-02
					SISESTSATGPR	102	3.31E-04
					WCVRPVCSR	62	5.88E-02
42	ORFL71C	TTSSRLLRCVRLSCLPAYMEETRAGRYIVRAGSGGQTEDTASGAESVLATLAAVRTRRRSVVCDGPPGSP TDSARHMSDLASLALTAEFGLGCLEAYVRINAGQVLPVVWPPGWNLVLQEIETDEDFKPEDVKAWSHY LCCQTRLAFVGRFVNEGVLSPDQQKKTAVCLISDEGYVFCYVREDTAVYYLARNLMEFARVGLRAVETLH CMRYLTSSLVKRYFRPLLRAWSLGLDTMARFIIRHHGQFMPLTYPPGTELRLCNLRCFENSVEGGHLLRNI KTAFGMRVLGLGTVSLKGENAPFPHLRWPVDLIPIVVAYTGAVYACDVRDDRYIRVGDNLNTFMCLGLN LLFENRRFSGHNGIYDRVPDCPKGRQHR*	376	11.2	AGSGGQTEDTASGAESVIATIAAVR	234	5.41E-101
					SVVCDGPPGSPTDSAR	140	3.31E-16
43	ORFL73C	MSLSQKENSSPSAADGSPGATGLVLLLLLLVVRLLLPPLLLLSSPPPPAPFSGSEGTLGGRETLGDLPALGVA GEPRGAPAAICGCRRDDDRFFSLGFLAAPVPPVSRTSSKDDGASEETL*	121	21.5	ETIGDIPAIGVAGEPR	159	1.20E-22
					GAPAAICGCR	75	8.04E-03
44	ORFL7C	MQHARVYVCIVSPPSRLPSAVPLPPPPAPPPLPSPPRYPFLVGWSWGTGRRPVGFCFIFFFIFYLLLSLISDY HFPLLPTTNRR*	84	13.1	VYVCIVSPPSR	112	1.29E-04
45	ORFL82W	MVGMHQGHVRLQTRADGFVHRVAADHALGKKSGQIRGQLVQAVLLRVVGEHGAVRVHEHLVPAGG VRDRANEMQSSEFAEKVVLAVLQIDLSQRVRRMLAPDHVDALRGPAQRHARALEAPQVTPQALIVASH AQVGQVARAKDLGYQAAFATRPTRVPRQRKARRRRRRGGRRRGRRAALATAGHAGSPRRSTGDKKAE AVVVSATRSTPSSTRSAATERRGAPLLGFL*	230	47.0	AIEAPQVTPQAIIVASHAQVGQVAR	190	2.03E-62
					AKDIGYQAAFATRPTR	116	1.82E-07
					DIGYQAAFATRPTR	113	2.48E-06
					DRANEMQSSEFAEK	79	1.01E-03
					GQIVQAVIIR	151	1.00E-05
					KAEAVVVSATR	134	8.70E-06
					MIAPDHVDAIR	83	1.91E-03
					RGAPIIGFI	79	6.24E-03
					VHEHIVPAGGVR	108	1.57E-04
46	ORFL86W	MRPGSDAPKRTQSDEADPSSVTARRLERRAREEILLPGALGPARPTRFIYTVSDGGTPSPRRGRSSGDGG NRGAGGNGDRQRRRWRRPDYGGQAHACDPVDRRRVVAATVVAAAARQLHGAAVTAAVRRRRCGG GNSDERGGRENRGARGSGCGSGDQE*	159	61.6	AREEIIIPGAIGPAR	92	2.54E-04
					AREEIIIPGAIGPARPTR	104	7.92E-06
					EEIIIPGAIGPARPTR	198	5.62E-47
					FIYTVSDGGTPSPR	135	7.64E-09

					MRPGSDAPK	64	2.91E-02
					MRPGSDAPKR	71	1.44E-02
					QIHGAAVTAAVR	196	3.80E-20
					RPDYGGQAHACDPVDR	178	1.29E-29
					RPDYGGQAHACDPVDRR	134	3.63E-12
					RTQSDEADPSSVTAR	84	3.09E-04
					RVVAATVVAAAAR	128	1.46E-07
					TQSDEADPSSVTAR	75	2.19E-03
					VVAATVVAAAAR	126	1.00E-06
47	ORFL88C_(UL30A)	TACRDARWVRFIRLRTLWSVRPRAHVADLEDFAAFYRTLSDSEQQEFEQEAELASRSQRVRDLREARRQLKMDLMCHGG*	79	62.0	AHVADIEDFAAFYR	151	1.84E-12
					TISDSEQQEFEQEAEIASR	64	1.28E-07
					TIWSVRPR	128	9.89E-04
					MDIMCHGG	86	9.10E-03
48	ORFS343C.iORF1	MKRPWWPCWAGCIASRLSSESPACSSSKSPRRWPSWVASLWSSLQRLSSRVRAFLVTLSGRPRCSSWCLWPPRSSCISAMSDRFSLSETGTRAAVVRSDSWCYYSPASWRTSRR*	114	28.9	AFIVTISGRPR	138	5.30E-06
					FSISETGTR	71	1.63E-02
					SDSWCYYSPASWR	63	1.69E-02
49	ORFS365W	MYDRLRLLRGQLPDHPQRRTHARCFRDPAQRHQLLLLPHFQNRCGVSIAPQRSHCTGTYSGRACLYHLLAAVPPDAVCGHVKTPQMDLQQLRVRKIAQTCAYLDRVARLLSLLSQSAAVRLRGHQVSARTALSACRVSPATLFPRCILVPQHELFASGLAESKRDIFRHAVRRGVSRLTNYTVIKKRYLGLFIQTPCPPLFSPCPIYTILNPRPFPCD*	218	30.7	CIIVPQHEIFASGIAESK	94	1.51E-05
					IAQTCAYIDR	116	2.33E-04
					IISIISQSAAVR	178	2.10E-15
					ITNYTVIK	66	4.35E-02
					TPQMDIQQIR	120	1.91E-04
					VSPATIFPR	99	1.32E-03
50	ORFS366W	MAAQTLAVAIPTKCISTVYSVLSNWLSPPTAAWVCAAALPALVPPRPGPTGMTWRTSYMVCVWIFWCASFY*	71	16.9	AAQTIAVAIPTK	175	2.20E-15

With proteosome inhibitor

	Name	Protein sequence	Length (aa)	Sequence coverage [%]	Identified peptides	Andromeda Score	PEP
1	ORFL103C_(vMIA)	MSPVYVNLLGSVGLLAFWYFSYRWIQRKRLEDPLPPWLRKKKACALTRRSRHRLRRQHGVIDGENSETER SVDLVAALLAEAGEESVTEDTEREDTEEEREDEEEENEARTPEVNPMDAEGLSGLAREACEALKKALRRH RFLWQRRRRARLLQHNGPQQSHQ*	163	8.6	ARIIQHNGPQQSHQ IIQHNGPQQSHQ	77 221	1.74E-03 1.23E-40
2	ORFL106C	MANSLWHRLRTLVSNSRQNEQIQQHSYRLHLRGYGSADFNSDGWTPVYEDQEFIVFSCRDDGNNRRR DVRSRSAMSARVPGLGVRRVPPRRHGPHDARRVLLPPDLQQHRRRVLASSCR*	119	39.5	QNEQIQQHSYR ANSIWHR IRTIVSNSR RVIIPPDIQQHR VIIPPDIQQHR VIIPPDIQQHRR VPGIGVR	128 177 78 162 155 76 157	7.94E-06 2.42E-03 8.01E-03 7.81E-11 3.89E-14 3.37E-03 1.99E-03
3	ORFL114W	MTLFAHQHAHSLYRRRYFAERSPEIVLLLDTVQLHPGRAGRIKGAQTATHASTATGHGGTAVTRQRQHR RRAVGRRRLGTPTDDAATTRKTRYTCYTARYVGISSGRFLVTPGQRPLLEAVISPDSISGDLCRRFLRLGFR AAAWNPTWPAPMCGHRGHPLFLVRVSRIHKFTRKITKKSGRFQ*	185	23.2	GHPIFIVR IGTPTDDAATTR IKGAQTATHASTATGHGGTAVTR	108 66 161	3.95E-03 1.64E-02 1.74E-32
4	ORFL118C	MHQKHRVQFGVGVAPKQHVGHERAGPVDEPLVGQVQEFGHGATIDVGPRLRGSRFQLRDHPADQRL LAPGTRTALLAAALGRRVVTGDQVRRRRGAAALFGYYGLDLAQGLQQLLESRQRGPVVVERQPVLVSAA VVLVSQEAVQLFVQLLGGVQRVARRLYRGVA*	166	27.7	AGPVDEPIVGQVQEFGHGATIDVGPR VQFGVGVAPK TAIIAAAIGR	169 92 116	2.18E-41 1.64E-03 1.30E-04
5	ORFL119C	MSQIVSQTAGAAAHHVLDQRVDGGVQVMRYGARHELGQQVHIGRDGYTAEVLESRGDVEDAIAYRLK NRGVDVLDEGGQVVDGDQVLVQRHLLGQQYTAQLNECEKTRRTHGLKQALRDTKDLRELSYHHQLRED AGELGMMAEG*	145	73.1	DGYTAEVIESRGDVEDAIAYR EISYHHQIR GDVEDAIAYR GVDVIDEGGQVVDGDQVIVQR HEIGQQVHIGR HIIGQQYTAQINECEK SQIVSQTAGAAAHHVIDQR VDGGVQVMR	115 106 137 223 186 182 161 168	5.75E-11 6.69E-04 5.61E-05 8.07E-57 1.30E-29 3.70E-18 7.50E-21 2.05E-07
6	ORFL121C	MDDELLETIGEESFVEGRRRGRCRVSAAFGTRRVQRPGAGVGRQRRQRQSLQLLLYLVDQLLVAAQGRQ GPFAAAVGRWYRGRRRGRRYR*	90	23.3	QGPFAAAVGR RVQRPGAGVGR	108 58	3.26E-04 1.12E-01
7	ORFL130W	MALALESTQTPERQVDDAADEAAAQRGAECKEDDLGGLYVMRLLSGARVSYRRSCRRRASSARLVVQT QCPQRHQVSRTRVAKRLHRLSGQYVERVSAGYLPGVPTQGCDEAPQLRRPRLPLVTPLVTVVYGQISAG VEALGA*	143	72.7	AIAIESTQTPER AIAIESTQTPERQVDDAADEAAAQR EDDIGGIYVMR GAECKEDDIGGIYVMR IPIVTPIVTVVYGQISAGVEAIGA ISGQYVER IVVQTQCPQR QVDDAADEAAAQR VSAGYIPGVPTQGCDEAPQIR	207 219 144 266 142 87 208 179 132	5.44E-31 1.64E-76 5.68E-07 2.27E-81 3.19E-24 1.21E-02 5.64E-20 4.46E-24 2.96E-14
8	ORFL131W.iORF2	MARSALATTMTVPNDNQSAPRTACTAETADDGDGSTPLADDSRAPPEADARTLRCPGPPLPLPPSLLPP	148	63.2	AATAQTAVAMAR ATAEPAPAEEPAPAGPESQSAR EETAQSIVEIISR SVSETTIGHDR CPGPPIPIPPSIIPPEPATPPPIHQR IVPYRPHNDTAATTPPPWVAAAAEAR SAIATTMTVPNDNQSAPR SEGTAFPPFA TACTAETADDGDGSTPIADDSR TACTAETADDGDGSTPIADDSRAPPEADAR	177 253 111 93 85 129 213 68 200 192	5.35E-16 3.40E-86 1.10E-04 7.82E-04 1.40E-05 1.44E-17 7.37E-37 2.05E-02 7.26E-45 1.50E-71
9	ORFL137W.iORF1	MARAPASSTYGRSAGRRGGRLIPGESSTPRQPLPSLRIPDAQPRHLGGRRTAQLLCGRNHVRRHAIHRTH QCDQHH*	76	22.4	IIPGESSTPR QPIPSIR	85 124	3.45E-03 8.79E-03
10	ORFL142W	MSLESTDESTRSSTRTRCRYRSSRACTRRRKIRQKPRGTPYFLTSSSERRDRRTAASETSRSSKSSERTSRVT PLTNSQAVLRTRSTPFMLRPEAPSTLPM*	101	40.6	GTPYFITSSSER GTPYFITSSSERR STPFMIRPEAPSTIPM VTPITNSQAVIR	107 83 55 155	1.73E-04 1.41E-03 1.98E-02 7.42E-08
11	ORFL147C	MSLAGARPDDSVSYVSESSHGDEFVTETMRSVFEMQRIRHGAGVSKVLRSERVTGGVHAVQEKRGGYS VSVPEDLPGGGGAESYAEFAADSLSGDAAEGAARGYGFAGSGADGLLAPAGPGGLLPYRPPFGSLRSPS HRGAPVYGQRPFQRQSGHTQRHRALPVQNELRVRDPGAGGRPGGLRAAAGAPPSPMRHLAGGASGS VRRRDDMRPMLRGADHHPESGPLAEQAAAGLAVQPYSGPPSVKPVRCEYPDGGAGPAGPDNAHPPL GWSPFGPQKPILFFIGLPQLHPGGGGGAEGVQPVYGYTGTNIFLVGFYLLVPYLGGYRQAGGHHHPAAK CVSPAVPGAHERHQSPSVRGRGGHLRARGKGVGRGGAHVRGVGLCRPQQDHRPHHIPQHSSFRGQ PGVQQAPRKQRDVHQNQAYSRGDSTSAARWHGGRRPRGRGYSPAWTGGDVGDGYDFDDGQQQQ QQQYSQSEE*	476	38.4	AAAGAPPSPMR AIPVQNEIR CVSPAVPGAHER DDMRPMIR DPGAGGRPGGIR DVHQNQAYSR DVHQNQAYSRGDSTSAAR GADHHPESGPIAEQAAAGIAVQPYSGPPSVKPVR GAPVYGQRPFQR GYGFAGSGADGIIAPAGPGGIIPYRPPFGSIR HIAGGASGSVR SERVTGGVHAVQEKR SVFEMQR VRDPGAGGRPGGIR VTGGVHAVQEK VTGGVHAVQEKR	131 130 165 100 115 89 83 193 95 177 167 131 142 143 140 123	6.62E-06 5.01E-05 7.01E-11 6.18E-03 9.15E-05 1.92E-03 2.67E-04 1.67E-77 8.97E-04 4.20E-60 2.55E-10 1.31E-06 4.71E-03 8.84E-06 2.23E-06 3.82E-05
12	ORFL149C.iORF2	MRPFLERGDRSSVGGAHGGGGFRRLCAYACALLRRRGAHESHVLLSFCALLSSYARHRGFYRGSRRRASA VRGRPPEVRVGGLHARGGSRAYRSGGVVRGGGLSGVRDDCARGGGQRLEGVSLCRSADSMRGGGDH GATARRRGGARAVVPAYAL*	155	54.8	AVVPAYAI GAHESHVIISFCAIISSYAR GGGISGVRDDCAR GGGQRIEGVSICR ICAYACAIIR IEGVSICR MRPFIER RICAYACAIIR SSVGGAHGGGGFR	100 122 102 143 65 127 97 91 65	6.00E-03 2.27E-11 2.02E-04 8.10E-06 2.58E-02 1.74E-03 3.44E-02 7.35E-04 1.17E-02
13	ORFL151C	MALVDQESCLRGFYSVFLRHCDVEPAIATFDGVRSPPDVVRLYPKPRYDYTLASAFLGTAVSGYTTSRRDN APNKNTGHRYVYTTSILIAVVTSVDTQHRSPSRCFMAMGESYFLQDEKCVFW*	123	26.8	AIVDQESCIR GFYSVFIR HCDVEPAIATFDGVR	137 132 119	4.39E-05 1.39E-03 3.63E-06
14	ORFL154W	MYTELSLLYKKKKLSIFISSEFLLDRHVFGTLSPHVFTVYNKNMCFSKDLEVFENGGNLRLGALSPGLGRSRR PAPIFFYFFTKYMMNQE*	90	28.9	HVFGTISPHVFTVYNK IGAISPGIGR	192 64	6.56E-24 2.79E-02
15	ORFL159W	IIVFAVAGSSCGDAGSVMMRGGADGIFREMVFAGGCSVPCRRRGWTSLASELPWHGGGDPAPSVCLV VWFACVYSLLILVVLLIYRCCIGFQDDLVSRTLAVYRACIQGPICNQTHNSTS*	121	40.5	EMVFAGGCSVPCR GGADGIFR ACIQGPICNQTHNSTS CCIGFQDDIVSR	132 102 126 165	3.76E-46 5.01E-03 1.88E-06 7.01E-11
16	ORFL161C.iORF2	MIPKAIRTTILLPFHPPPATHSGAPADTSSWALRASVADIGRPPLSSDAWPRPPRCPLTPTTVSLTTTMTA	138	84.1	ASVADIGRPPISSDAWPRPPR CPITPTTVSITTTMTANR IFVAEFPVVIR IITTNWTPWTR ISVIITPPQPIITDRPPDR RIFVAEFPVVIR TITAAIR TTIIIPFHPPPATHSGAPADTSSWAIR	127 160 140 150 95 89 99 194	2.28E-12 1.27E-18 2.30E-06 3.16E-07 1.98E-06 1.01E-03 2.84E-02 2.36E-63
17	ORFL172W	LSRRRHRLPARVSVAATRARRRRPRTVRLPCGLVCLRVLTADSGRFAAHLPLLHRLPRRPSLPHLGCVPSLY TGPDM*	77	24.7	RPSIPHIGCVPSIYTGPDM	81	1.40E-04
	ORFL174C.iORF1	MPGRAIRLHARAAGIRAASPAVGCRQPQRHFSVHLSGHHRRRYRRFPLFCARTLAAHELAAAQRTPFLFI TYPSDAGDRSGSHANRLF*	88	39.8	FPIFCAR HFSVHISGHHR RFPIFCAR TIAAHEIAAAQR TPFIFITYPSDAGDR	90 41 89 132 87	4.06E-02 1.23E-01 1.96E-02 1.75E-05 2.65E-04
18	ORFL174C.iORF2	MKRRRHYRRRSKFPTSGCLPVSSSNNATRSWPFTINTSFGYHASAKCPPRWASASISPKVFSALPFTNAW TRNSFACPSCWNPGCKSRSWMSSTSTTLFSPSFRAPSKAISAFSPVLCQNPGSS*	124	83.9	AISAFSPVICQNPGSS NSFACPSCWNPGCK SKFPTSGCIPVSSSNNATR SWMSSTSTTIFSPSFR SWPFTINTSFGYHASAK VFSAIPFTNAWTR WASASISPK	128 158 129 106 90 152 67	9.47E-07 5.34E-08 2.81E-09 1.02E-04 1.45E-04 8.80E-07 2.60E-02
19	ORFL185C	LDHAAEPSGAAALCGQASPPTSSSSPSVSSATYFRHDMAQKPYPNRWTKRFTYCSTPTGDPSASCVKIPP SVPTTAASVTARSSGKTPSVSTFSKAIINTMYSICLDVFLRVLWRSSF*	118	44.9	HDMAQKPYPNR FTYCSTPTGDPSASCVK IPPSVPTTAASVTAR RFTYCSTPTGDPSASCVK TPSVSTFSK	55 120 77 93 148	7.30E-02 4.28E-06 8.66E-04 3.68E-05 1.41E-05
20	ORFL18W	MRQSIGLDERDAATPLHGRLVVVVKEEHQQSQEEQQALTCPGGRGEGLSEMAVRHHVRQLRRNGEER	77	13.0	GEGISEMAVR	89	1.42E-03

		HQCRIQEPPCRRLVAGGI*			RIVAGGI	108	1.67E-02
23	ORFL211W	MSGVVEYVAQTAGLRGPGVVCGRESLCVRFAMVVDGLGGGLRVVRIDGQHAGGTPARVVGDLKVINV HRHRRKD*	74	17.6	FAMVVDGIGGGIR	110	1.21E-04
24	ORFL223W	LAFLGRVVVAVAAAEGHERQEGLVARVGQPEFGVHHLVALAAQRRHIGEGRGDAGGGVALALLQLRVA LLDLGGRISAQHDGGGGGASGG*	90	61.1	GDAGGGVAIAIIQIR	213	6.29E-31
					VAIIDIGGR	119	1.92E-04
					VGQPEFGVHHIVAIAAQR	136	1.30E-09
					VVVAVAAAEGHER	196	5.02E-22
25	ORFL224C.iORF1	MRPLTCSTWSLTCAKSTCTASRRRATPRRVLATDCTSPPSSASTRRCARRPTCLCRALSWTRSSAVPIKSR RTPCSSRTRAAKSSIFCVTARSTPTPTIISARRFTCIWTRPSPPTARPRARAWRP*	126	54.0	FTCIWTR	77	7.06E-02
					FTCIWTRPSPPTAR	135	2.70E-05
					MRPITCSTWSITCAK	110	1.38E-25
					RFTCIWTRPSPPTAR	120	8.69E-06
					RVIATDCTSPPSSASTR	103	6.25E-05
					SSIFCVTAR	122	1.55E-04
					STPTPTIISAR	85	1.93E-03
					VIATDCTSPPSSASTR	164	1.05E-13
					VIATDCTSPPSSASTRR	32	7.07E-02
26	ORFL234C	MSTRCSVASRTCTVKVASSLGGCAPPVLLLTSTTLFSPSPSSRGSGGNSQRRSPPAGRSRKAVALRRTAPP PASVPKTCHSSSGTPSESAQGSTSSLPAEPTMSGRPHHESTSAAASNAVSRINVCRPWTA*	131	40.5	INVCRPWTA	167	3.15E-07
					PHHESTSAAASNAVSR	144	9.33E-11
					VASSIGGCAPPVIIITSTTIFSPSPSSR	221	5.76E-85
27	ORFL235W	MRFVKSLLKICVFARDAPSGRSRVCWVTSPSTLRQPALRPAFCGFCRVATASFMSTTAIRVTDPSRPKCAC GLWSTVIGANFLAATIRVPPVVFKRARPACCWSGATSGWWLLLRGRWRLALGRGSCPGPAAGAVTGA RERVRAGSRVSRDPLSLRGSPQRRGLLVTALPAR*	172	8.7	GSCPGPAAGAVTGAR	101	1.09E-04
28	ORFL237C	MAVSDDAARATSATAMSRRETTHHPVERTTTGAAVAVEEVEGTHQPLVAPDQQHAGRALLNTTGGTR IVAAKKFAPITVDHSPHAHFGRDGSVTRIAVVDIKLAVATRQNPQKAGRSAG*	119	52.9	AIINTTGGTR	145	3.49E-06
					FAPITVDHSPHAHFGR	106	1.02E-04
					FAPITVDHSPHAHFGR	126	1.88E-06
					IAVVDIK	112	1.69E-02
					KFAPITVDHSPHAHFGR	116	5.90E-06
					TTTGAAVAVEEVEGTHQPIVAPDQQHAGR	259	2.07E-137
29	ORFL251W	TASDERCARALRRRVMDLPTTVVRKYWTFTNPNRILHQSVNQTFDVRQFVFDNARLVNCVDGDGKVLH LNKGWLCATIMQHGEASAGAKTQQGFMSIDITGDGELQEHLFVRGGIVFNKSVSSVVGSSGPNESALLT MISENGNLQVTYVRHYLKNHGESSGGGGGCGAASTASAVCVSSLGGSGGTRDGPSAEEQQRRRQEQR HEERRKKSSSSAGGGGGGGASGGGGGGGSGGQHSSDSANGLLRDPRLMNRQKERRPPPSSENDGSPP LREAKRQKTTAQHEGHGGGGKNETEQQSGGAGGGGGGGSGRMSLPLDTSEAVAFLNYSSSSSAVSSSS NNHHHHHHHHNAVTDVAAGTDDSCQRRRAGLSSSTTAASRGPPQLSPQPALAPQQSHVRHSSPPVLY PQVPSPVSRPLPPQSKHQLLLPGPCVKSSPTW*	438	65.1	VMDIPTTVVR	89	1.92E-03
					AGISSSTTAASR	73	7.83E-03
					GPPQISPQPAIAPQQSHVR	291	4.51E-121
					HQIIIPGPCVK	170	1.21E-10
					HSSPPVIYPQVPSPVSR	45	4.96E-02
					HSSPPVIYPQVPSPVSRPIPPQSK	126	1.82E-16
30	ORFL253W.iORF2	MDLPTTVVRKYWTFTNPNRILHQSVNQTFDVRQFVFDNARLVNCVDGDGKVLHLNKGWLCATIMQHG EASAGAKTQQGFMSIDITGDGELQEHLFVRGGIVFNKSVSSVVGSSGPNESALLTMISENGNLQVTYVRH YLKNHGESSGGGGGCGAASTASAVCVSSLGGSGGTRDGPSAEEQQRRRQEQRHEERRKKSSSSAGGG GGGGASGGGGGGGSGGQHSSDSANGLLRDPRLMNRQKERRPPPSSENDGSPPLREAKRQKTTAQHE GHGGGGKNETEQQSGGAGGGGGGSGRMSLPLDTSEAVAFLNYSSSSAVSSSSNNHHHHHHHHNA VTDVAAGTDGIRGASASSQSAF*	358	2.3	GASASSQSAF	59	4.41E-02
31	ORFL253W.iORF3	MARCCTSTRAGSALPLCSTARLRLAPRRSRASCPLTLRATGNFRSTSLYAAVSSLTNPSPRWWAPADPMR ARCSL*	75	34.7	STSIYAAVSSITNPSPR	66	2.26E-03
					WWAPADPMR	74	1.01E-02
32	ORFL26W	MKDRNKILLCIIFICIMCLICIYFKRHCIPTPSPDKADFRVEFPSLLPCIGVQCAA*	56	17.9	HCIPTPSPDK	143	1.15E-05
33	ORFL271C.iORF1_ (RL8A)	MLRIHKRQPPAGTLLRFQNVQSLHRRAAVSGRRSLLQSRENG*	42	47.6	FQNVQSIHR	155	4.65E-06
					FQNVQSIHRR	125	9.37E-05
					QPPAGTIIR	116	2.35E-04
					RQPPAGTIIR	142	1.29E-05
34	ORFL27C	MTFKLPYVMPHGHLRQALSPTSWTCEGLLLLLGLLVLFFHHHNQSAVERRRRVSFVEADRLPHESGWYSS DDDGDRDGDEETGESHNRNSVGLSAVFS*	98	30.6	IPYVMPHGHIR	96	5.87E-04
					NSVGISAVFS	140	3.22E-05
					RVSFVEADR	94	1.59E-03
					VSFVEADR	108	3.87E-03
35	ORFL294W	MDILTQRPKQRRDFGDGLVLTAAESVAGPARQQRATQTEYLEESYVCHRRARVSGGNDGGVVDTCAAC CALRKSAGLGDRGRKRERSSYRWRRRR*	95	68.4	MDIITQRPK	126	8.69E-05
					MDIITRPK	110	2.86E-03
					ATQTEYIEESYVCHR	231	9.58E-52
					DFGDGIVITAAESVAGPAR	256	1.05E-86
					RDFGDGIVITAAESVAGPAR	134	1.85E-13
					VSGGNDGGVVDTCAACCAIR	295	5.51E-126
					VSGGNDGGVVDTCAACCAIRK	255	2.89E-80
36	ORFL298W	MRVSSNPRRRRGRYRSSCVENRGTSRSRDSVDARASATRSGLPASSDLTPACSKVRRAG*	59	25.4	SGIPASSDITPACSK	104	8.59E-05
37	ORFL299W	MQLREIPSGVRQAVLKSIHQPMAFDGVQDELVAAVVGTTETCANGNHGPGVCHDRAGDGNLLSLGDQ HRRRRRGLIVLGPHLFGDVSGGEGERRGRIFEESRAAQQHPEQPRQSSERLPAAHDDSFLPFFAETLPACS CSVCRPLKRRAGVVVTTVRRSFSPGMPATERPVFLQVFFAASPSPSRPTWWTSTAQELSRRRRYATCLIA TSDV*	212	48.1	AGDGNIISIGDQHR	145	1.67E-06
					AGDGNIISIGDQHRR	196	3.17E-22
					AGVVVTTVR	117	2.28E-04
					AGVVVTTVRR	122	9.64E-05
					GIIVIGPHIFGDVSGGEGER	159	9.07E-23
					GIIVIGPHIFGDVSGGEGERR	162	6.00E-24
					IPAAHDDSFIPFFAETIPACSCSVCRPIK	159	9.93E-45
					PTWWTSTAQEISR	133	5.71E-05
					RAGVVVTTVR	104	3.70E-04
					RAGVVVTTVRR	146	1.26E-06
					RYATCIIATSDV	89	1.01E-03
					YATCIIATSDV	123	1.06E-05
38	ORFL301C	MRQVAYRRRRRESSCAVLVHHVGRDGEGEAAKKTCKKTGRSVAGIPGEKLRRTVVTTTPARRLSGRHTEQ EQAGSVSAKKGRKESSCAAGSRSELCRGCSGCC*	102	31.4	ESSCAVIVHHVGR	199	1.27E-22
					SVAGIPGEK	65	3.03E-02
					TVVTTTPAR	127	8.45E-05
39	ORFL321W	IDGRYKATRTRAAALAYGAAPGKLLFLLLFLLHRRDRFPRPGTYRSNYRNNLEESRHDGAQVCHHHDLSG QDQGAALVARQRHQEMVDQRCHRHLHRSLSGGPDVLYAAAGTQRERQRLDKSQAATAPSAVAGPPA DRDVVDHRTESHAYETPSSVQNSDATLGMASCSPRRRVSPDLLVSSELRRRNIWRRVTCVTGERGSPAVT TAVSSTRVRPAVLCGKAPVSETVDEKENGAATAGGGGVNAAVDVRRCEYSETAVRQKLDPSGNDSRCV VAAADSSTDSDGKSIIAGVQVVDHDEDIIAPQSLWCTAFKEALWDVALLEVPRWAWQGWKRWRNSES GRRWSAGSASASSLSDLAGEAVGELVGSVVAYVILERLWLAARGWVCETGVEAEEAMARRRQRMLWR MFSRGGDGECSRRCSMEMACEEESAVL*	435	40.2	SISGGPDVIYAAAGTQR	45	1.79E-02
					SQAATAPSAVAGPPADRDVVDHR	137	5.85E-24
					APVSETVDEKENGAATAGGGGVNAAVDVR	346	1.60E-255
					APVSETVDEKENGAATAGGGGVNAAVDVRR	207	5.90E-83
					CEYSETAVR	80	6.35E-03
					CVVAAADSSTDSDGK	100	1.18E-04
					EAIWDVAIIEVPR	211	7.17E-31
					ENGAATAGGGGVNAAVDVR	94	1.44E-39
					GSPAVTTAVSSTR	147	1.52E-06
					GWVCETGVEAEEAMAR	169	4.02E-14
					RCEYSETAVR	133	7.17E-05
					RVSPDIIVSSEIR	200	3.33E-30
					RVSPDIIVSSEIRR	134	6.73E-05
					RVTCVTGER	109	4.38E-04
					VRPAVICGK	129	6.89E-05
					VRPAVICGK	130	5.51E-05
					VSPDIIVSSEIR	164	7.78E-11
					VTCVTGERGSPAVTTAVSSTR	200	7.74E-42
					WAWQGWK	132	5.20E-03
40	ORFL47W.iORF1	MRLRWCVRPVCSRCATATASGTRAHRSCCLTATVWRWPASSPRTTGQRSLSESTSATGPRSCVRWW WTHAAVRCYTTTPAVTSPAGAGVSRTAARRSG*	98	50.0	CYTTTPAVTSPAGAGVSR	111	6.39E-06
					SCCITATVWR	90	1.80E-03
					SISESTSATGPR	102	2.37E-04
					WCVRPVCSR	60	4.71E-02
41	ORFL52C	MDVNRPQNPSRVRRVRNMTVLTGARGNRRRNSSYRALLVMTTGPCDSATQRPGYVGIVSRW*	61	18.0	MDVNRPQNPSR	93	6.01E-04
42	ORFL6C	MINKRVKRKKLQTFGYLSFPITRRIQSRRFPTRENRTKTRSFLFILIVNNNYRQQRNTTKNVNPQGCSCPLD APVAILMARTAISV*	86	16.3	KIQTFGYISFPITR	78	1.14E-03
43	ORFL73C	MSLSQKENSSPSAADGSPGATGLVLLLLLLVVRLLLPPLLLLSSPPPPAPFSGSEGTLGGRETLGDLPALGVA GEPRGAPAAICGCRRDDDRFFSLGFLAAPVPPVSRTSSKDDGASEETL*	121	21.5	ETIGDIPAIGVAGEPR	171	1.93E-14
					GAPAAICGCR	96	1.21E-03
44	ORFL7C	MQHARVYVCIVSPPSRLPSAVPLPPPPAPPPLPSPPRYPFLVGWSWGTGRRPVGFCFIFFIFYLLLSLISDY HFPLLPTTNRR*	84	13.1	YYVCIVSPPSR	97	1.08E-03
45	ORFL82W	MVGMHQGHVRLQTRADGFVHRVAADHALGKKSGQIRGQLVQAVLLRVVGEHGAVRVHEHLVPAGG VRDRANEMQSSEFAEKVVLAVLQIDLSQRVRRMLAPDHVDALRGPAQRHARALEAPQVTPQALIVASH AQVGQVARAKDLGYQAAFATRPTRVPRQRKARRRRRRGGRRRGRRAALATAGHAGSPRRSTGDKKAE AVVVSATRSTPSSTRSAATERRGAPLLGFL*	230	47.0	AEAVVVSATR	59	4.24E-02
					AIEAPQVTPQAIIVASHAQVGQVAR	175	1.85E-46
					AKDIGYQAAFATRPTR	105	2.84E-04
					DRANEMQSSEFAEK	81	1.30E-03
					GQIVQAVIIR	122	1.18E-04
					KAEAVVVSATR	144	5.29E-07
					MIAPDHVDAIR	97	3.50E-04
					RGAPIIGFI	52	8.49E-02
					VHEHIVPAGGVR	110	1.35E-04
46	ORFL86W	MRPGSDAPKRTQSDEADPSSVTARRLERRAREEIILLPGALGPARPTRFIYTVSDGGTPSPRRGRSSGDGG NRGAGGNGDRQRRRWRRPDYGGQAHACDPVDRRRVVAATVVAAAARQLHGAAVTAAVRRRRCGG GNSDERGGRENRGARGSGCGSGDQE*	159	62.3	AREEIIIPGAIGPAR	103	9.78E-05
					AREEIIIPGAIGPARPTR	210	1.23E-35
					EEIIIPGAIGPAR	98	3.85E-04

					EEIIIPGAIGPARPTR	236	7.11E-53
					FIYTVSDGGTPSPR	162	9.31E-10
					MRPGSDAPK	118	1.74E-04
					MRPGSDAPKR	105	8.08E-04
					QIHGAAVTAAVR	179	4.05E-16
					RPDYGGQAHACDPVDR	186	5.90E-20
					RPDYGGQAHACDPVDRR	180	1.23E-21
					RTQSDEADPSSVTAR	99	1.03E-04
					RVVAATVVAAAAR	61	1.27E-02
					TQSDEADPSSVTAR	96	2.42E-04
					TQSDEADPSSVTARR	51	3.71E-02
					VVAATVVAAAAR	138	8.78E-06
47	ORFL88C_(UL30A)	TACRDARWVRFIRLRTLWSVRPRAHVADLEDFAAFYRTLSDSEQQEFEQEAELASRSQRVRDLREARRQL KMDLMCHGG*	79	62.0	AHVADIEDFAAFYR	208	5.02E-30
					TISDSEQQEFEQEAEIASR	216	3.24E-49
					TIWSVRPR	119	2.46E-03
					MDIMCHGG	86	1.31E-02
48	ORFS334W	MSGSVSPQTAVPSLRALPAAATRAAWTPAGNGARSGYRHCRGKRLTSRKRVVRPPVGAQFRNVEPRHF PAGVRRP*	75	33.3	SGSVSPQTAVPSIR	85	5.48E-04
					VVRPPVGAQFR	167	3.07E-10
49	ORFS343C.iORF1	MKRPWWPCWAGCIASRLSSESPACSSSKSPRRWPSWVASLWSSLQRLSSRVRAFLVTLSGRPRCSSWC LWPPRSSCISAMSDRFSLSETGTRAAVVRSDSWCYYSPASWRTSRR*	114	58.8	AFIVTISGR	81	6.14E-03
					AFIVTISGRPR	199	2.86E-22
					CSSWCIWPPR	79	5.76E-03
					FSISETGTR	80	6.35E-03
					RPWWPCWAGCIASR	89	3.11E-04
					SDSWCYYSPASWR	107	1.45E-04
					SSCISAMSDR	88	2.08E-03
					SSCISAMSDRFSISETGTR	66	3.43E-03
50	ORFS359W	MRHGRFGRGAALQHGGPGRPAGEHPALRRIYLVFLLCLFPILFRDGLFHHVQHAFLDALQGSWGGGSPS EASNVVLWIFGFLFLVVIVGRVVFLLILFFVRVVDA*	105	20.0	GAAIQHGGPGRPAGEHPAIR	176	2.94E-30
					GAAIQHGGPGRPAGEHPAIRR	82	2.73E-18
51	ORFS365W	MYDRLRLLRGQLPDHPQRRTHARCFRDPAQRHQLLLLPHFQNRCGVSIAPQRSHCTGTYSGRACLYHLL AAVPPDAVCGHVKTPQMDLQQLRVRKIAQTCAYLDRVARLLSLLSQSAAVRLRGHQVSARTALSACRVS PATLFPRCILVPQHELFASGLAESKRDIFRHAVRRGVSRLTNYTVIKKRYLGLFIQTPCPPLFSPCPIYTILNPR PFPCD*	218	30.7	CIIVPQHEIFASGIAESK	104	1.41E-05
					IAQTCAYIDR	146	3.17E-06
					IISIISQSAAVR	178	5.03E-16
					ITNYTVIK	92	9.16E-03
					TPQMDIQQIR	122	1.19E-04
					VSPATIFPR	63	3.51E-02
52	ORFS366W	MAAQTLAVAIPTKCISTVYSVLSNWLSPPTAAWVCAAALPALVPPRPGPTGMTWRTSYMVCVWIFWC ASFY*	71	16.9	AAQTIAVAIPTK	181	2.06E-16

Table S8.

Novel HCMV ORFs that were detected by mass spectrometric analysis. This table is divided to two parts and lists the ORFs that were detected by mass spectrometric analysis of samples treated or untreated with proteasome inhibitor. The first column gives the name of the ORF, the second column lists the protein sequence, the third column gives the length of the protein, the forth gives the protein sequence coverage, the fifth column lists the peptide sequences with Ile and Leu both represented by an "I" due to their indistinguishability to our MS/MS analysis, the sixth column provides the Andromeda score (*19*) of the peptide identification and the seventh column lists the PEP value giving the probability of an incorrect sequence assignment. Note we did not double count peptides mapping to overlapping novel ORFs and in these cases only the longest ORF is being reported in this table.

ORF name	sequence	length	mass spectrometry	GFP expression	localization
ORFL27C.iORF1_(RL8A)	MPHGHLRQALSPTSWTCEGLLLLLGLLVLFFHHHNQSAVERRRRVSFVEADRLPHESGWYSSDDDGDRDGDEETGESHNRNSVGLSAVFS*	90	detected	yes	not specific
ORFL47W.iORF1	MRLRWCVRPVCSRCATATASGTRAHRSCCLTATVWRWPASSPRTTGQRSLSESTSATGPRSCVRWWWTHAAVRCYTTTPAVTSPAGAGVSRTAARRSG*	98	detected	yes	Mitochondria
ORFL66W	MPQAASSSLVASQSTVLTSFGAVGHAIVDAVVAAETSVFKQQVKSQTHERVEIISHPDVAVVAHVAGVDGPGIGDDDGDKVDGPAQVRKGRVLALEADGSQAENAHSESGFDVAQQVTAFHAVFETPEVA*	130	not detected	yes	not specific
ORFL80W	MSSASPTRTKPLRSRDSRTNLSCPTSAAYTVPVSPQ	41	not detected	yes	not specific
ORFL88C_(UL30A)	TACRDARWVRFIRLRTLWSVRPRAHVADLEDFAAFYRTLSDSEQQEFEQEAELASRSQRVRDLREARRQLKMDLMCHGG*	79	detected	yes	not specific
ORFL88C.iORF1	MGPLHQTAYALERQTQGACSRLGGLCRLLSHPLGQ*	35	not detected	yes	Mitochondria
ORFL118C	MHQKHRVQFGVGVAPKQHVGHERAGPVDEPLVGQVQEFGHGATIDVGPRLRGSRFQLRDHPADQRLLAPGTRTALLAAALGRRVVTGDQVRRRRGAAALFGYYGLDLAQGLQQLLESRQRGPVVVERQPVLVSAAVVLVSQEAVQLFVQLLGGVQRVARRLYRGVA*	166	detected	yes	ER
ORFL129C	LLALRQSAGSAAARLRARLRDALWTALRTRGLESRAGPQRGLCRVRSPAGFRRHSALRGTHLRLHCGGTRRDGPASPASHFARARPAVRALSDGRVPRQLFGKRRRRPAVSRLRAG*	116	not detected	yes	not specific
ORFL130W	MALALESTQTPERQVDDAADEAAAQRGAECKEDDLGGLYVMRLLSGARVSYRRSCRRRASSARLVVQTQCPQRHQVSRTRVAKRLHRLSGQYVERVSAGYLPGVPTQGCDEAPQLRRPRLPLVTPLVTVVYGQISAGVEALGA*	143	detected	yes	nucleus
ORFL211W	MSGVVEYVAQTAGLRGPGVVCGRESLCVRFAMVVDGLGGGLRVVRIDGQHAGGTPARVVGDLKVINVHRHRRKD*	74	detected	yes	not specific
ORFL213C	MFLGSRGARTPRPDAAMDESAAFILITRLVLKALTAMRSRLTPVTSANSIPPMSLRRHGA*	60	not detected	yes	Mitochondria
ORFL234C	MSTRCSVASRTCTVKVASSLGGCAPPVLLLTSTTLFSPSPSSRGSGGNSQRRSPPAGRSRKAVALRRTAPPPASVPKTCHSSSSGTPSESAQGSTSSLPAEPTMSGRPHHESTSAAASNAVSRINVCRPWTA*	131	detected	no expression	
ORFL235W	MRFVKSLLKICVFARDAPSGRSRVCWVTSPSTLRQPALRPAFCGFCRVATASFMSTTAIRVTDPSRPKCACGLWSTVIGANFLAATIRVPPVVFKRARPACCWSGATSGWWLLLRGRWRLALGRGSCPGPAAGAVTGARERVRAGSRVSRDPLSLRGSPQRRGLLVTALPAR*	172	detected	yes	Mitochondria
ORFL255C.iORF1	MARYRNLSVTVPLRRRPPTPCCWTRLSWTLWPCCTTIRINCGPC*	44	not detected	yes	Mitochondria
ORFL245C	MVLTCAERVRLRIALRFCAAITVVPVITQSRLAASTWMLEVLPAPAVPVTSSARKGKKGQRSTQGTRVPPASDEAAAAAEEETEE*	85	not detected	no expression	
ORFS359W	MRHGRFGRGAALQHGGPGRPAGEHPALRRIYLVFLLCLFPILFRDGLFHHVQHAFLDALQGSWGGGSPSEASNVVLWIFGFLFLVVIVGRVVFLLILFFVRVVDA*	105	detected	yes	ER
ORFS366W	MAAQTLAVAIPTKCISTVYSVLSNWLSPPTAAWVCAAALPALVPPRPGPTGMTWRTSYMVCVWIFWCASFY*	71	detected	yes	ER
ORFS370W_(US33A)	MSLRFPERAGYEKLGYRPHAKRVWVHDPLGLTRFIMRQLMMYPLVLPFTFPFYVPRS*	57	not detected	yes	Mitochondria

Table S9.

A list of ORFs that were GFP-tagged and examined by microscopy. The first column gives the name of the ORF, the second column lists the protein sequence, the third column gives the length of the protein, the fourth column indicate whether the ORF was detected by mass spectrometry, the fifth column indicates whether GFP expression could be observed and the sixth column indicates the subcellular localization.

official and/or new ORFs within the transcript	An example of 5' position	strand
RL10-RL11-RL12-RL13	9692	+
UL2	13122	-
UL4-UL5	13390	-
UL11-UL13	19622	+
ORFL41C-ORFL43C	17265	-
UL16-UL17	22598	+
UL19-UL20	25629	+
UL23-UL24	30434	-
UL25	28247	+
UL26-UL27-UL28-UL29	33138	-
UL30	38174	-
UL31	38056	+
UL32	43448	-
UL33-UL34	43359	+
UL35	46389	+
UL37-UL38	52501	-
UL40-UL41-UL42-UL43-UL44-UL45-UL46	55490	-
ORFL123C-ORFL124C	69441	-
UL47-UL48	64552	+
UL148A-UL149-UL50-UL51	72165	-
UL52-UL53	76942	+
UL54	82306	-
UL55-UL56-UL57	85152	-
UL69-UL70	102545	-
UL71-UL73-UL74A	105894	+
UL72-UL74	107680	-
UL75	111563	-
UL76-UL77-UL78	114106	+
UL80-UL80.5	117324	+
UL82-UL83	120594	-
UL84-UL85-UL86	125471	-
UL89	135810	-
UL92-UL93-UL94	137321	+
UL95-UL96	140982	+
UL97-UL99	145473	+
UL100	147696	-
UL102	146559	+
ORFL236C-ORFL239C-UL103	151413	-
RNA5	161789	-
UL111	161475	+
UL112	161862	+
UL114	165089	-
UL115-UL116-UL117-UL119	169384	-
UL122-UL123	171924	-
UL124	165146	+
ORFL286	181420	+
UL128-UL130-UL131A	176999	-
UL132-UL148-UL147A-UL147-UL146	181366	-
UL144-UL145-UL142	183236	-
UL141-UL140-UL139	186339	-
UL138-UL136-UL135-UL133-UL148A-UL148B-UL	191100	-
UL148C-UL148D	187176	+
ORFL321W	192630	+
US3-US6-US7	200982	-
US8-US9	205148	-
US12-US13-US14-US15-US16-US17	209540	-
US18-US19-US20-US21	212759	-
US22-US23	217858	-
US27-US28	223831	+
US30-US31-US32	227818	+
US34-US34A	230686	+

Table S10.

A list of ORFs that were GFP-tagged and examined by microscopy. The first column gives the name of the ORF, the second column lists the protein sequence, the third column gives the length of the protein, the fourth column indicate whether the ORF was detected by mass spectrometry, the fifth column indicates whether GFP expression could be observed and the sixth column indicates the subcellular localization.

Additional Data table S2 (separate file)

Conservation of the Merlin ORFs in other HCMV strains. The first column gives the ORF name, the second column lists the ORF length, the third column indicates whether the ORF correspond throughout most of its length to a canonical ORF(5' end extension),columns 4-20 list the percentage of identity to orthologs in other HCMV strains and in chimp CMV (found by tblastn, see description in material and methods) and the corresponding genomic coordinates (in parentheses). N.C indicates that no ortholog was found. M, L, T, V indicates that the corresponding initiating codon is conserved, * indicates that the corresponding stop codon is conserved.
Although interpretation of conservation analysis is complicated since many of the newly identified ORFs are antisense or overlapping with canonical ORFs, 52% of the novel ORFs (excluding internal in-frame ORFs) were conserved from start to stop (with above 80% amino acid identity) across all the 16 sequenced HCMV strains and 12 ORFs showed high conservation (including conservation of the start and stop codon) in chimp CMV (labeled in yellow).

Additional Data table S3 (separate file)

A list of identified splice junctions. The first column lists the splice junction name, the second column lists any known canonical ORF that is affected by the given splice junction, the third column lists any newly-identified ORFs that is affected by this splice junction, the fourth and fifth columns list the genomic coordinates that denote the 5 prime and the 3 prime positions of the intron, the sixth column lists the strand in which the splice junction was identified, the seventh, eighth and ninth columns list the HMMSplicer score that was obtained for this junction at 5hpi, 24hpi and 72hpi respectively, the tenth, eleventh and twelve columns list the TopHat score that was obtained for the junction at 5hpi, 24hpi and 72hpi respectively. The thirteenth column indicates whether the splice junction was also supported by Gatherer et al. (*3*) findings.

Additional Data table S5 (separate file)

This table lists all the HCMV translated ORFs >20 amino acids and their expression levels. The first column gives the ORF name. The second column gives the strand on which the ORF is located, the third column lists the coordinates in the genome, the fourth column lists the ORF's length, the fifth column gives the start codon sequence, the sixth column indicates for internal ORFs whether they are in-frame (I) or out of frame (O), the seventh column indicates whether the ORF was added manually, the eighth, ninth and tenth columns indicate whether the ORF was detected by the SVM at 5hr, 72hr and the 72hr replicate respectively, the eleventh column indicates whether the ORF was unambiguously detected by mass spectrometry (this is indicated only for the new ORFs which are non overlapping), the twelfth, thirteenth and fourteenth columns give the mRNA reads density (rpkM) that was obtained for the whole ORF(including regions that overlap with other ORFs) at 5hpi, 24hpi and 72hpi respectively. The fifteenth, sixteenth and seventeenth columns give the number of footprints reads density (rpkM) that was obtained for the whole ORF (including regions that overlap with other ORFs) at 5hpi, 24hpi and 72hpi respectively, the eighteenth, nineteenth and twentieth columns give (where possible) the mRNA reads density (rpkM) for regions of the ORF which are not overlapping with any other ORF at 5hpi, 24hpi and 72hpi respectively. The twenty one,

twenty two and twenty three columns (where possible) give the footprints reads density for regions of the ORF which are not overlapping with any other ORF at 5hpi, 24hpi and 72hpi respectively.

Additional Data table S6 (separate file)

This table lists all the HCMV translated ORFs ≤20 amino acids. The columns description is similar to that table S5. The mRNA and footprints reads density (rpkM)are presented only for ORFs > 10aa.

Additional File S1 (separate file)

A genome browser file that includes our annotations together with the footprints (CHX, Harr, LTM and no drug), mRNA data and predicted 3' ends for each time point. We include explicit directions for how to load and view the file in the materials and methods section.

Additional File S2 (separate file)

This file contains the complete set of annotated MS/MS spectra of the identified peptides that correspond to the proteins presented in table S8.

Additional File S3 (separate file)

This file contains two sheets that list the full data about the ORFs that were detected by mass spectrometry analysis; of samples treated or untreated with proteasome inhibitor. The first column lists the protein sequence, the second column gives the name of the ORF, the third column gives the length of the protein, the forth gives the protein sequence coverage, the fifth column lists all the ORFs from which an identified peptide could be derived, the sixth column lists the RAW file that contains the identified MS/MS spectrum, the seventh column lists the peptide sequence with Ile and Leu both represented by an "I" due to their indistinguishability to MS/MS analysis , the eighth column provides the peptide length (number of amino acids), the ninth column lists the charge state of the peptide, the tenth column lists the m/z ratio, the eleventh column provides calculated mass of the peptide, the twelfth column lists the scan number of the MS/MS scan that the identification is based on, the thirteenth column gives the Andromeda score (*19*) of the peptide identification and the fourteenth column lists the PEP value giving the probability of an incorrect sequence assignment.

References

1. A. Dolan *et al.*, *J Gen Virol* **85**, 1301 (May, 2004).
2. R. J. Stanton *et al.*, *J Clin Invest* **120**, 3191 (Sep, 2010).
3. D. Gatherer *et al.*, *Proc Natl Acad Sci U S A* **108**, 19755 (Dec 6, 2011).
4. N. T. Ingolia, L. F. Lareau, J. S. Weissman, *Cell* **147**, 789 (Nov 11, 2011).
5. B. Langmead, C. Trapnell, M. Pop, S. L. Salzberg, *Genome Biol* **10**, R25 (2009).
6. C. Trapnell, L. Pachter, S. L. Salzberg, *Bioinformatics* **25**, 1105 (May 1, 2009).
7. M. T. Dimon, K. Sorber, J. L. DeRisi, *PLoS One* **5**, e13875 (2010).
8. M. J. de Hoon, S. Imoto, J. Nolan, S. Miyano, *Bioinformatics* **20**, 1453 (Jun 12, 2004).
9. A. J. Saldanha, *Bioinformatics* **20**, 3246 (Nov 22, 2004).
10. R. Atalay *et al.*, *J Virol* **76**, 8596 (Sep, 2002).
11. C. Sinzger *et al.*, *J Gen Virol* **89**, 359 (Feb, 2008).
12. M. Wagner, A. Gutermann, J. Podlech, M. J. Reddehase, U. H. Koszinowski, *J Exp Med* **196**, 805 (Sep 16, 2002).
13. V. T. Le, M. Trilling, H. Hengel, *J Virol* **85**, 13260 (Dec, 2011).
14. N. C. Hubner *et al.*, *J Cell Biol* **189**, 739 (May 17, 2010).
15. J. Cox, M. Mann, *Nat Biotechnol* **26**, 1367 (Dec, 2008).
16. J. R. Wisniewski, A. Zougman, N. Nagaraj, M. Mann, *Nat Methods* **6**, 359 (May, 2009).
17. S. Schandorff *et al.*, *Nat Methods* **4**, 465 (Jun, 2007).
18. J. Rappsilber, Y. Ishihama, M. Mann, *Anal Chem* **75**, 663 (Feb 1, 2003).
19. J. Cox *et al.*, *J Proteome Res* **10**, 1794 (Apr 1, 2011).
20. O. R. Homann, A. D. Johnson, *BMC Biol* **8**, 49 (2010).