Decoding Human Cytomegalovirus
Noam Stern-Ginossar *et al.*
Science **338**, 1088 (2012);
DOI: 10.1126/science.1227919

11. J. G. Kingsolver, *Am. Nat.* **174**, 755 (2009).
12. R. W. Eppley, *Fish Bull.* **70**, 1063 (1972).
13. T. L. Martin, R. B. Huey, *Am. Nat.* **171**, E102 (2008).
14. J. Norberg, *Limnol. Oceanogr.* **49**, 1269 (2004).
15. S. A. H. Geritz, É. Kisdi, G. Meszéna, J. A. J. Metz, *Evol. Ecol.* **12**, 35 (1997).
16. P. A. Abrams, *Ecol. Lett.* **4**, 166 (2001).
17. J. C. Stegen, R. Ferriere, B. J. Enquist, *Proc. Biol. Sci.* **279**, 1051 (2011).
18. R. W. Reynolds *et al.*, *J. Clim.* **20**, 5473 (2007).
19. R. W. Reynolds, N. A. Rayner, T. M. Smith, D. C. Stokes, W. Wang, *J. Clim.* **15**, 1609 (2002).
20. M. R. Kearney, W. Porter, *Ecol. Lett.* **12**, 334 (2009).
21. Intergovernmental Panel on Climate Change (IPCC), *Climate Change 2007: The Physical Science Basis. Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change*, S. Solomon *et al.*, Eds. (Cambridge Univ. Press, Cambridge, 2007).
22. N. Nakićenović *et al.*, *Special Report on Emissions Scenarios: A Special Report of Working Group III of the Intergovernmental Panel on Climate Change*, N. Nakićenović, R. Swart, Eds. (Cambridge Univ. Press, Cambridge, 2000); www.osti.gov/energycitations/product.biblio.jsp?osti_id=15009867.
23. T. Delworth *et al.*, *J. Clim.* **19**, 643 (2006).
24. D. W. McKenney, J. H. Pedlar, K. Lawrence, K. Campbell, M. F. Hutchinson, *Bioscience* **57**, 939 (2007).
25. D. U. Hooper *et al.*, *Nature* **486**, 105 (2012).
26. D. Tilman, D. Wedin, J. Knops, *Nature* **379**, 718 (1996).
27. P. B. Reich *et al.*, *Science* **336**, 589 (2012).
28. C. A. Deutsch *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **105**, 6668 (2008).
29. K. Härnström, M. Ellegaard, T. J. Andersen, A. Godhe, *Proc. Natl. Acad. Sci. U.S.A.* **108**, 4252 (2011).
30. A. F. Bennett, R. E. Lenski, in *In the Light of Evolution. Volume 1. Adaptation and Complex Design*, J. C. Avise, F. J. Ayala, Eds. (National Academies Press, Washington, DC, 2007), pp. 225–238.
31. J. L. Knies, R. Izem, K. L. Supler, J. G. Kingsolver, C. L. Burch, *PLoS Biol.* **4**, e201 (2006).
32. I. E. Huertas, M. Rouco, V. López-Rodas, E. Costas, *Proc. Biol. Sci.* **278**, 3534 (2011).
33. S. L. Chown *et al.*, *Clim. Res.* **43**, 3 (2010).
34. M. J. Angilletta Jr., R. S. Wilson, C. A. Navas, R. S. James, *Trends Ecol. Evol.* **18**, 234 (2003).

Acknowledgments: This research was supported by NSF (grants DEB-0845932, DEB-0845825, and OCE-0928819), including the BEACON Center for the Study of Evolution in Action (grant DBI-0939454) and a Graduate Research Fellowship to C.T.K., as well as a grant from the James S. McDonnell Foundation. K. Edwards and N. Swenson provided statistical advice, and J. Lennon, G. Mittelbach, and E . Miller provided comments on the manuscript. The trait data presented are shown in table S5, and the collated growth rate data are in table S6. Also thanks to NOAA/OAR/ESRL PSD, Boulder, CO, USA, for NOAA_OI_SST_V2 data provided at their Web site, www.esrl.noaa.gov/psd/. This is W. K. Kellogg Biological Station contribution number 1694. E.L. and M.K.T. conceived the original idea; M.K.T. collected the growth-temperature data; M.K.T. and C.T.K. analyzed the data; C.T.K. and C.A.K. developed and C.T.K. analyzed the eco-evolutionary model; M.K.T. and C.T.K. ran the mechanistic species distribution models; and M.K.T., C.T.K., and E.L. wrote and C.A.K. commented on the manuscript.

Supplementary Materials

www.sciencemag.org/cgi/content/full/science.1224836/DC1
Materials and Methods
Figs. S1 to S12
Tables S1 to S6
References (*35*–*135*)

17 May 2012; accepted 10 October 2012
Published online 25 October 2012;
10.1126/science.1224836

Decoding Human Cytomegalovirus

Noam Stern-Ginossar,[1] Ben Weisburd,[1] Annette Michalski,[2]* Vu Thuy Khanh Le,[3] Marco Y. Hein,[2] Sheng-Xiong Huang,[4] Ming Ma,[4] Ben Shen,[4,5,6] Shu-Bing Qian,[7] Hartmut Hengel,[3] Matthias Mann,[2] Nicholas T. Ingolia,[1]† Jonathan S. Weissman[1]*

The human cytomegalovirus (HCMV) genome was sequenced 20 years ago. However, like those of other complex viruses, our understanding of its protein coding potential is far from complete. We used ribosome profiling and transcript analysis to experimentally define the HCMV translation products and follow their temporal expression. We identified hundreds of previously unidentified open reading frames and confirmed a fraction by means of mass spectrometry. We found that regulated use of alternative transcript start sites plays a broad role in enabling tight temporal control of HCMV protein expression and allowing multiple distinct polypeptides to be generated from a single genomic locus. Our results reveal an unanticipated complexity to the HCMV coding capacity and illustrate the role of regulated changes in transcript start sites in generating this complexity.

The herpesvirus human cytomegalovirus (HCMV) infects the majority of humanity, leading to severe disease in newborns and immunocompromised adults (*1*). The HCMV genome is ~240 kb with estimates of between 165 and 252 open reading frames (ORFs) (*2*, *3*). These annotations likely do not capture the complexity of the HCMV proteome (*4*) because HCMV has a complex transcriptome (*5*, *6*), and genomic regions studied in detail reveal noncanonical translational events, including regulatory (*7*) and overlapping ORFs (*8*–*11*). Defining the full set of translation products—both stable and unstable, the latter with potential regulatory/antigenic function (*12*)—is critical for understanding HCMV.

To identify the range of HCMV-translated ORFs and monitor their temporal expression, we infected human foreskin fibroblasts (HFFs) with the clinical HCMV strain Merlin and harvested cells at 5, 24, and 72 hours after infection using four approaches to generate libraries of ribosome-protected mRNA fragments (Fig. 1A and table S1). The first two measured the overall in vivo distribution of ribosomes on a given message; infected cells were either pretreated with the translation elongation inhibitor cycloheximide or, to exclude drug artifacts, lysed without drug pretreatment (no-drug). Additionally, cells were pretreated with harringtonine or lactimidomycin (LTM), two drugs with distinct mechanisms, which lead to strong accumulation of ribosomes at translation initiation sites and depletion of ribosomes over the body of the message (Fig. 1A) (*13*–*15*). A modified RNA sequencing protocol allowed quantification of RNA levels as well as identification of 5′ transcript ends by generating a strong overrepresentation of fragments that start at the 5′ end of messages (fig. S1) (*16*).

The ability of these approaches to provide a comprehensive view of gene organization is illustrated for the UL25 ORF: A single transcript start site is found upstream of the ORF (Fig. 1A, mRNA panel). Harringtonine and LTM mark a single translation initiation site at the first AUG downstream of the transcript start (Fig. 1A, Harr and LTM). Ribosome density accumulates over the ORF body ending at the first in-frame stop codon (Fig. 1A, CHX and no-drug). In the no-drug sample, excess ribosome density accumulates at the stop codon (Fig. 1A, no-drug) (*14*).

Examination of the full range of HCMV translation products, as reflected by the ribosome footprints, revealed many putative previously unidentified ORFs: internal ORFs lying within existing ORFs either in-frame, resulting in N-terminally truncated translation products (Fig. 1B), or out of frame, resulting in entirely previously unknown polypeptides (Fig. 1C); short uORFs (upstream ORFs) lying upstream of canonical ORFs (Fig. 2A); ORFs within transcripts antisense to canonical ORFs (Fig. 2B); and previously unidentified short ORFs encoded by distinct transcripts (Fig. 2C). For all of these categories, we also observed ORFs starting at near-cognate codons (codons differing from AUG by one nucleotide), especially CUG (Fig. 2D).

HCMV expresses several long RNAs lacking canonical ORFs, including β2.7, an abundant RNA, which inhibits apoptosis (*17*). In agreement with β2.7's observed polysome association (*18*), multiple short ORFs are translated from this RNA (Fig. 2E and fig. S2), and the corresponding proteins for two of these ORFs were detected by means of high-resolution MS (Fig. 2E). Although the translation efficiency

[1]Department of Cellular and Molecular Pharmacology, Howard Hughes Medical Institute, University of California, San Francisco, San Francisico, CA 94158, USA. [2]Department of Proteomics and Signal Transduction, Max Planck Institute of Biochemistry, Martinsried D-82152, Germany. [3]Institut für Virologie, Heinrich-Heine-Universität Düsseldorf, 40225 Düsseldorf, Germany. [4]Department of Chemistry, Scripps Research Institute, 130 Scripps Way #3A2, Jupiter, FL 33458, USA. [5]Department of Molecular Therapeutics, Scripps Research Institute, 130 Scripps Way #3A2, Jupiter, FL 33458, USA. [6]Natural Products Library Initiative at The Scripps Research Institute, Scripps Research Institute, 130 Scripps Way #3A2, Jupiter, FL 33458, USA. [7]Division of Nutritional Sciences, Cornell University, Ithaca, NY 14853, USA.

*To whom correspondence should be addressed. E-mail: michalsk@biochem.mpg.de (A.M.); weissman@cmp.ucsf.edu (J.S.W.)

†Present address: Department of Embryology, Carnegie Institute for Science, Baltimore, MD 21218, USA.

of these ORFs is low, four of them are highly conserved across HCMV strains (table S2). We found three similar polycistronic coding RNAs (including RNA1.2 and RNA4.9), and two short proteins encoded by these RNAs were confirmed with MS (fig. S3).

To define systematically the HCMV-translated ORFs using the ribosome profiling data, we first annotated HCMV splice junctions, identifying 88 splice sites (table S3). We then exploited the harringtonine-induced accumulation of ribosomes at translation start sites so as to identify ORFs using a support vector machine (SVM)–based machine learning strategy (*14*, *19*). We observed a strong enrichment for AUG (33-fold) and near cognate codons in the translation initiation sites identified with this analysis (Fig. 3A). Visual inspection of the ribosome profiling data confirmed the SVM-identified ORFs and suggested an additional 53 putative ORFs (table S4). The large majority (86%) of the SVM-identified ORFs, and all of the manually identified ones, were identified by means of SVM analysis of an independent biological replicate (table S5 and fig. S4). The observed initiation sites were not caused by harringtonine because LTM treatment also induced ribosome accumulation at the vast majority (>98%) of these positions (Fig. 3B).

In total, we identified 751 translated ORFs that were supported by both the LTM and harringtonine data (tables S5 and S6 and file S1). The footprint density measurements for these ORFs were reproducible between biological replicates (figs. S5 and S6). Of these ORFs, 147 were previously suggested to be coding (Fig. 3C). We did not find strong evidence of translation for 24 previously annotated ORFs (table S7), although these proteins may well be expressed under different conditions.

Many newly identified ORFs are very short (245 ORFs ≤ 20 codons) (Fig. 3C) and are found upstream of longer ORFs. We also identified 239 short ORFs (21 to 80 codons) (Fig. 3D). Last, we identified 120 ORFs that are longer than 80 amino acids. These are primarily ORFs that contain splice junctions or alternative 5′ ends of previous annotations.

Several lines of evidence support the validity of the ORFs we identified. First, as seen for the previously annotated ORFs, newly identified ORFs showed a significant [$P < 10^{-70}$; Kolmogorov-Smirnov (K-S) test] excess of ribosome footprints at the predicted stop codon (Fig. 1A and fig. S7). Because our ORF predictions

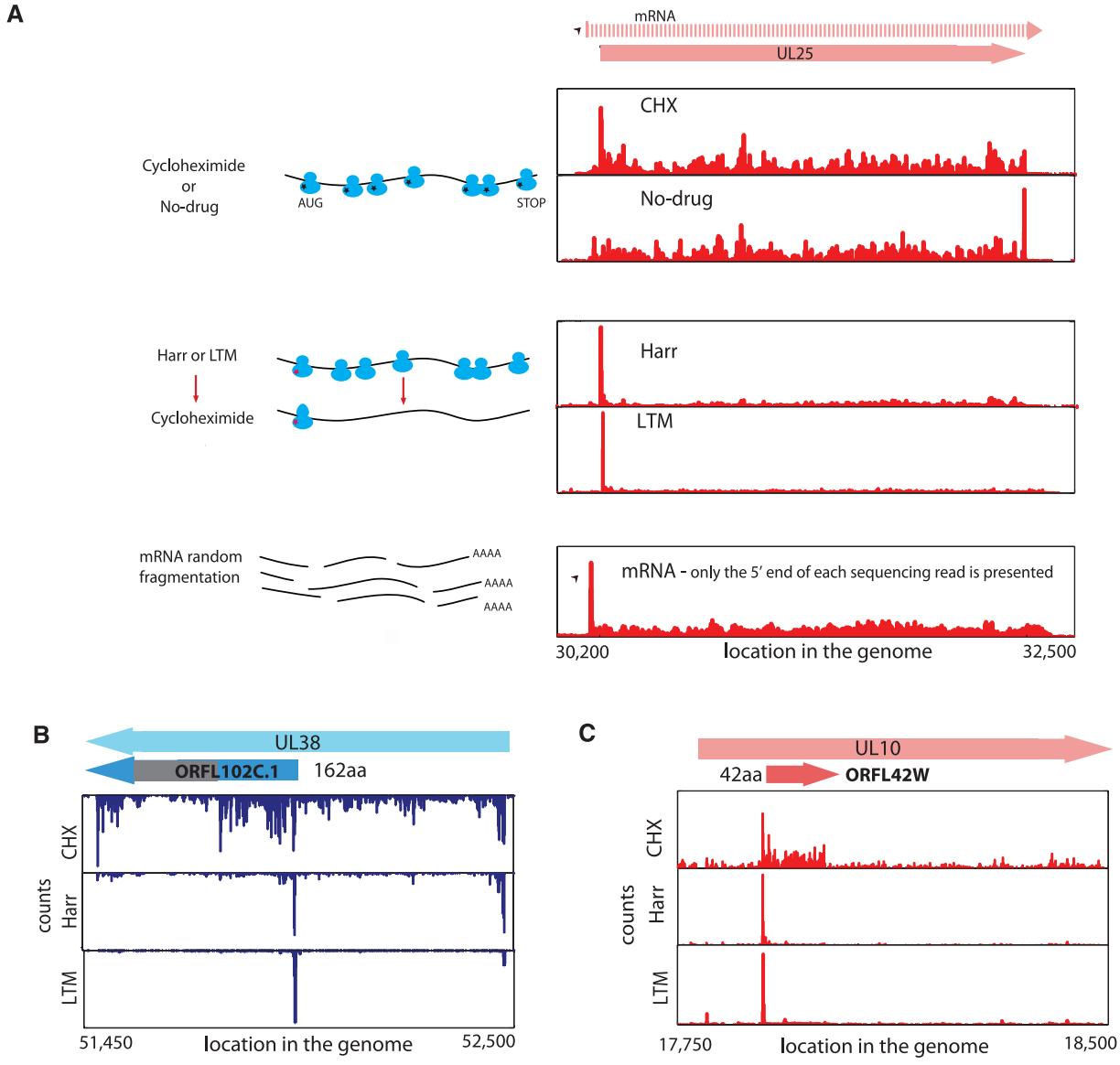

Fig. 1. Ribosome profiling of HCMV-infected cells. (**A**) Ribosome occupancies after various treatments (illustrated to left); cycloheximide (CHX), no-drug, harringtonine (Harr), and LTM together with mRNA profiles of the UL25 gene at 72 hours after infection. An arrow marks the mRNA start. (**B** and **C**) Ribosome occupancy profiles for (B) UL38 and (C) UL10 genes that contain internal initiations. The gray area symbolizes a low-complexity region.

were based on translation initiation sites found in the harringtonine and LTM samples, the observation that these accurately predicted downstream stop codons in an untreated sample provides independent support for our approach. Second, ribosome-protected footprints displayed a 3-nucleotide (nt) periodicity that was in phase with the predicted start site both globally (Fig. 3E) and in specific ORFs that contain internal

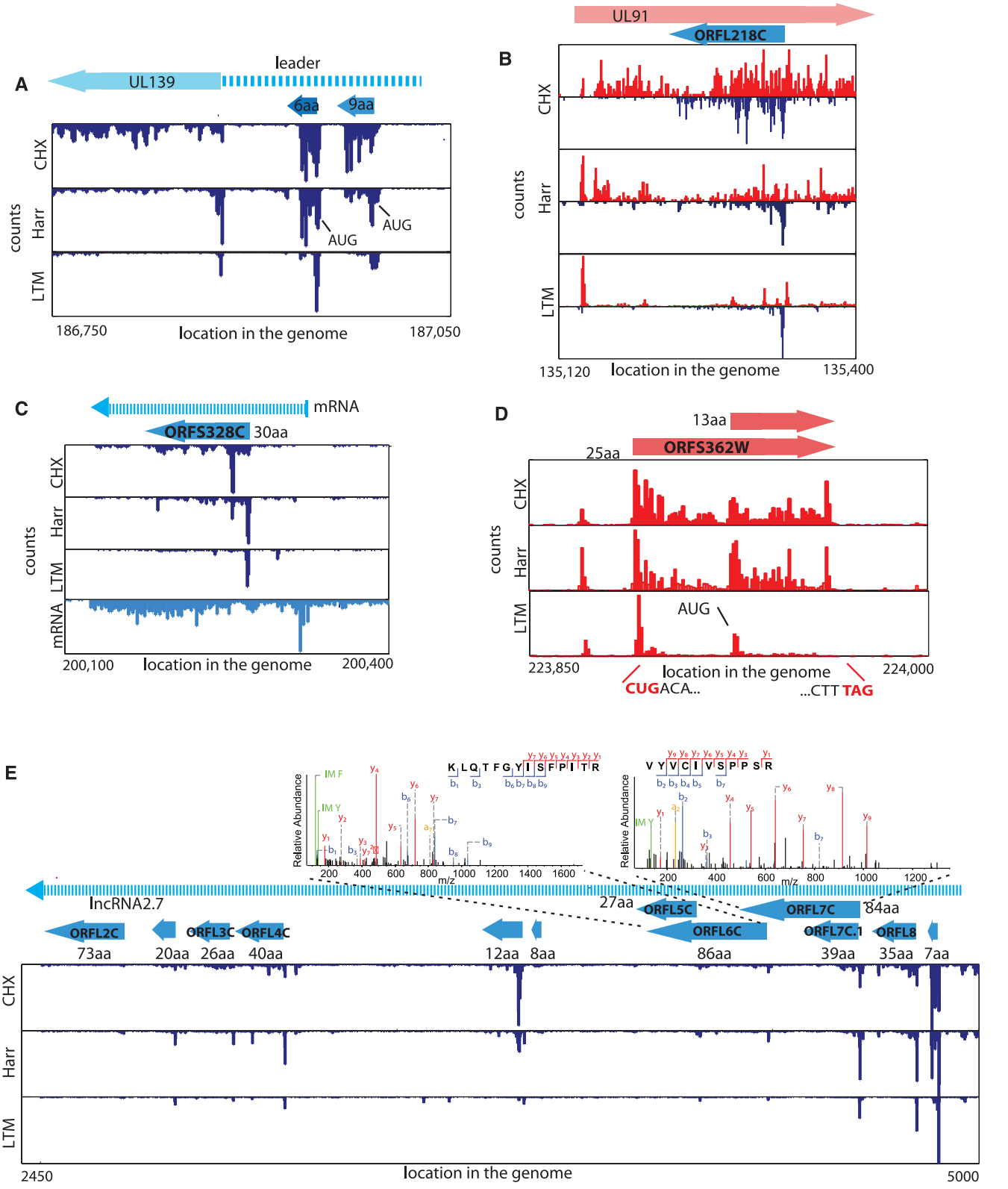

Fig. 2. Many ribosome footprints do not correspond to previously annotated ORFs. (**A**) Ribosome occupancy profiles for the leader region of UL139 gene. (**B**) Ribosome occupancy profiles of plus and minus strands (red and blue, respectively) for the UL91 gene. (**C**) mRNA and ribosome occupancy profiles for a previously unidentified short ORF. (**D**) Ribosome occupancies around a short ORF that initiates at a CUG codon. (**E**) Ribosome occupancy profiles for RNA β2.7. (Top) The annotated MS/MS spectra of two distinct peptides originating from ORFL6C and ORFL7C.

out-of-frame ORFs (fig. S8). Third, brief inhibition of translation initiation using an eIF4A inhibitor Pateamine A (*20*) led to depletion of ribosome density from the body of the large majority of the predicted ORFs (fig. S9), indicating that the ribosomes were engaged in active elongation. The newly identified ORFs also exhibited a distribution of expression levels similar to that of previously annotated canonical ORFs (fig. S10). Last, many of the newly identified ORFs are conserved in other HCMV strains (table S2).

High-resolution tandem mass spectrometric measurements on virally infected cells by using stringent criteria and manual validation (files S2 and S3) (*16*, *21*) unambiguously detected 53 previously unidentified proteins out of the 96 genomic loci that are not overlapping with annotated ORFs and contain at least one specific previously unidentified protein that is longer than 55 amino acids (table S8). For classes of new ORFs that were difficult to monitor with MS (truncated forms of longer proteins or short proteins), we used a tagging approach. For two N-terminally truncated proteins (derived from UL16 and UL38), we confirmed the appearance of alternative shorter transcripts and detected the expected full length and truncated tagged protein products (fig. S11). The truncated protein derived from UL16 was also observed in the context of the native virus (fig. S12), and we confirmed a splice variant of UL138 by using an antibody (fig. S12). For five short ORFs (including two initiated at near cognate start sites), we fused the ORFs in frame to a green fluorescent protein (GFP)–coding region in their otherwise native transcript context. We identified protein products of the expected sizes and confirmed that we correctly identified the translation start sites (fig. S13). We also showed that one of these short proteins (US33A-57aa), which was not identified with MS but was recently predicted by means of transcript analysis to be coding (*6*), is expressed in the context of the native virus (Fig. 3F and fig. S12). Additionally, we focused on the very short, near cognate driven uORFs that lie directly upstream of UL119 and US9, whose inclusion changes during infection as a result of changes in the 5′ end of the transcripts. We found that these uORFs modulated the translation efficiency of a downstream reporter gene (fig. S14).

Last, we examined the subcellular localization for 18 newly identified ORFs (11 of which were detected by means of mass spectrometry) (table S9) using transient expression of GFP-tagged proteins. We detected 15 proteins, 10 of which showed specific subcellular localization patterns: six in mitochondria, three in the endoplasmic reticulum (ER), and one in the nucleus (Fig. 3G and fig. S15). Immunoprecipation and MS experiments on two of these GFP-tagged proteins, ORF359W (ER localized) and US33A (mitochondrially localized), identified a few specific interacting proteins. Western blot analysis confirmed the interactions with TAP1 (ORF359W) and the mitochondrial inner membrane transport TIM machinery (US33A) (fig. S16).

HCMV genes are expressed in a temporally regulated cascade. Our data provides an opportunity to monitor viral protein translation throughout infection. Most of the viral genes, including newly identified ORFs, showed tight temporal regulation of protein synthesis levels; 82% of ORFs varied by at least fivefold. Hierarchical clustering of viral coding regions by their footprint densities during infection (a measure of the relative translation rates) revealed several distinct temporal expression patterns (fig. S17).

As was seen previously for a limited number of genomic loci (*8*–*11*, *22*), examination of viral transcripts during infection revealed a pervasive use of alternative 5′ ends that is critical to the

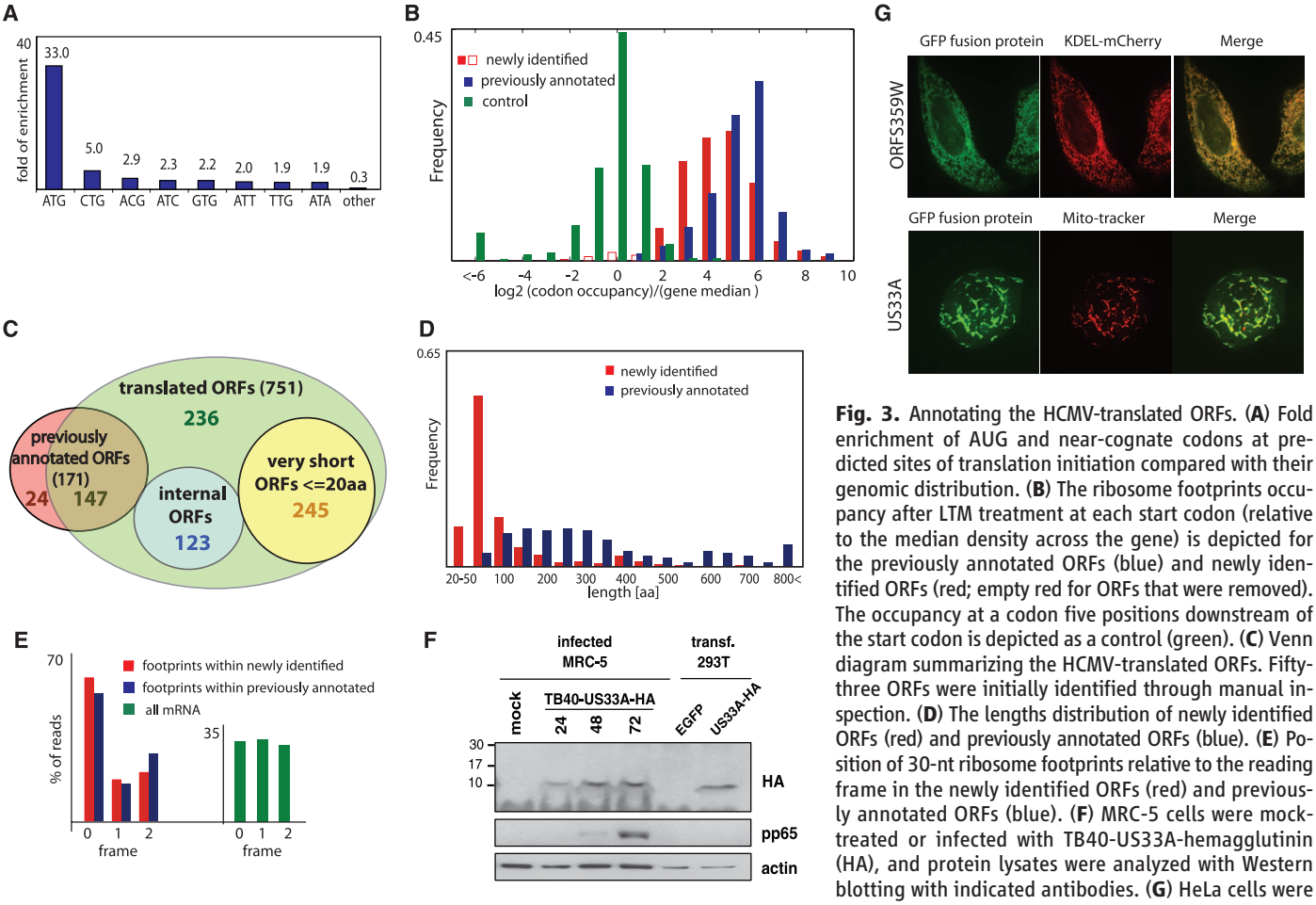

Fig. 3. Annotating the HCMV-translated ORFs. (**A**) Fold enrichment of AUG and near-cognate codons at predicted sites of translation initiation compared with their genomic distribution. (**B**) The ribosome footprints occupancy after LTM treatment at each start codon (relative to the median density across the gene) is depicted for the previously annotated ORFs (blue) and newly identified ORFs (red; empty red for ORFs that were removed). The occupancy at a codon five positions downstream of the start codon is depicted as a control (green). (**C**) Venn diagram summarizing the HCMV-translated ORFs. Fifty-three ORFs were initially identified through manual inspection. (**D**) The lengths distribution of newly identified ORFs (red) and previously annotated ORFs (blue). (**E**) Position of 30-nt ribosome footprints relative to the reading frame in the newly identified ORFs (red) and previously annotated ORFs (blue). (**F**) MRC-5 cells were mock-treated or infected with TB40-US33A-hemagglutinin (HA), and protein lysates were analyzed with Western blotting with indicated antibodies. (**G**) HeLa cells were transfected with GFP fusion proteins together with an ER marker (KDEL-mCherry) or stained with MitoTracker Red (Invitrogen, Grand Island) and imaged by means of confocal microscopy.

tight temporal regulation of viral genes expression and production of alternate protein products during infection. For example, at the US18-US20 locus, 5 hours after infection there is one main transcript starting just upstream of US20 enabling US20 translation. At 24 hours after infection, a shorter version of the transcript is detected starting immediately upstream of US18, enabling its translation. A third previously unknown transcript isoform starting within the US18 coding sequence emerges at 72 hours after infection, resulting in translation of a truncated version of US18 (ORFS346C.1) at this time point (Fig. 4, A and B). Another example is detailed in fig. S18, and we identified reproducible temporal regulation of 5′ ends in 61 viral loci (encompassing ~350 ORFs) (figs. S19 and S20 and table S10), six of which we confirmed with Northern blot analysis (Fig. 4B and figs. S11 and S21). Thus, our studies reveal a pervasive mode of viral gene regulation in which dynamic changes in 5′ ends of transcripts control protein expression from overlapping coding regions. Just as alternative splicing (a process in which a single gene codes for multiple proteins) expands protein diversity, alternative transcript start sites may provide a broadly used mechanism for generating complex proteomes.

The genomic era began with the sequencing of the bacterial DNA virus, phi X, in 1977 (*23*) and the mammalian DNA virus, Simian virus 40 (*24*), the following year. Since then, extraordinary advances in sequencing technology have enabled the determination of a vast array of viral genomes. Deciphering their protein coding potential, however, remains challenging. Here, we present an experimentally based analysis of translation of a complex DNA virus, HCMV, by using both next-generation sequencing and high-resolution proteomics. It is possible that many of the short ORFs we have identified are rapidly degraded and do not act as functional polypeptides. Nonetheless, these could still have regulatory function or be an important part of the immunological repertoire of the virus as major histocompatibility complex (MHC) class I bound peptides are

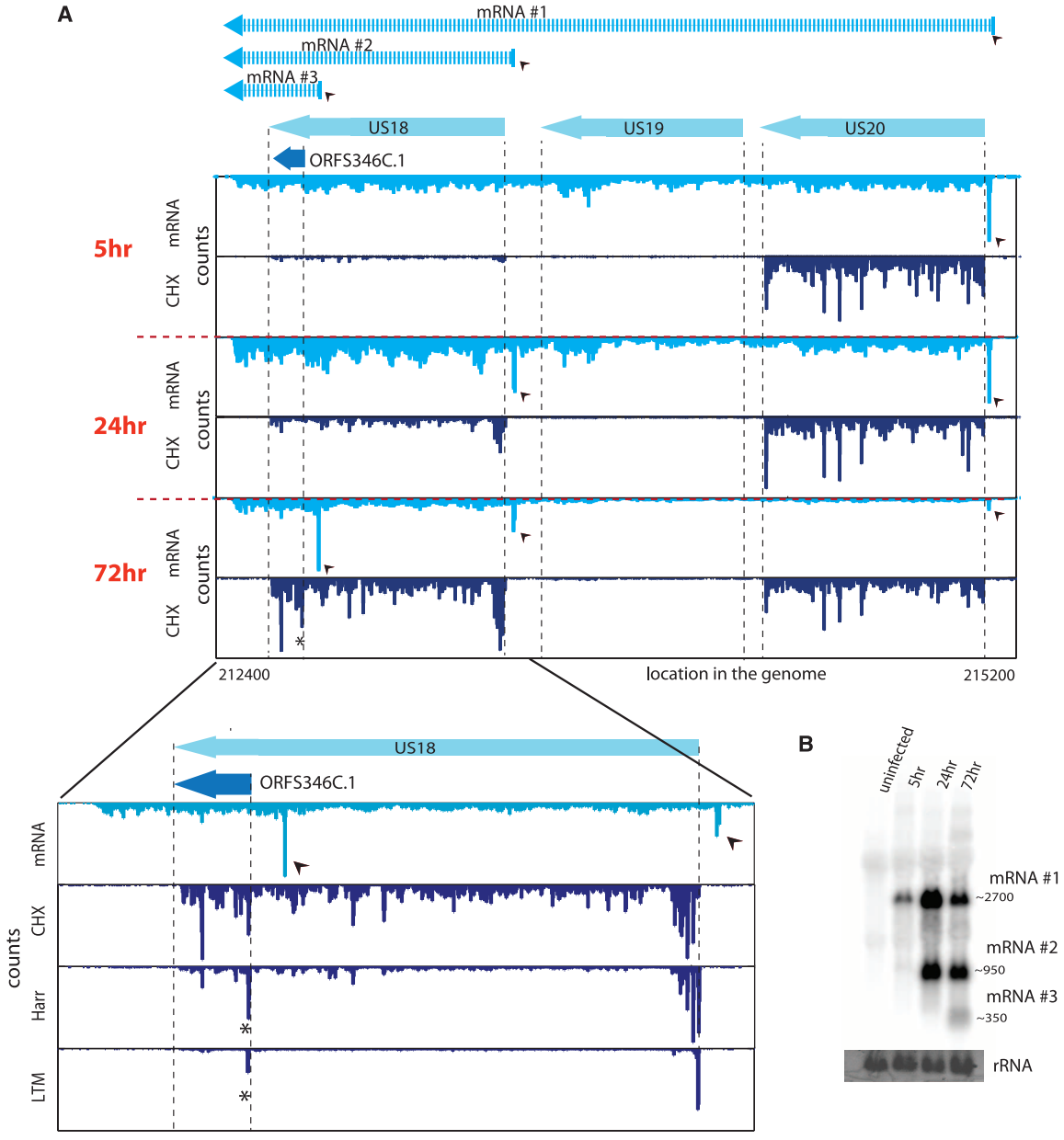

Fig. 4. A major source of ORFs' diversity during infection originates from alternative transcripts starts. (**A**) The mRNA and ribosome occupancy profiles around US18 to US20 loci at different infection times (marked left). Small arrows denote the different mRNA starts, and (top) the corresponding mRNAs are illustrated. (Bottom) An expanded view of the US18 locus at 72 hours after infection and includes the harringtonine and LTM profiles (asterisks indicate the internal initiation). (**B**) Total RNA extracted at different time points during infection was subjected to Northern blotting for ORFS346C.1.

generated at higher efficiency from rapidly degraded polypeptides (*25*). Our work yields a framework for studying HCMV by establishing the viral proteome and its temporal regulation, providing a context for mutational studies and revealing the full range of HCMV functional and antigenic potential.

References and Notes

1. E. S. Mocarski, T. Shenk, R. F. Pass, in *Fields Virology*, B. N. Fields, D. M. Knipe, P. M. Howley, Eds. (Lippincott, Williams & Wilkins, Philadelphia, PA, 2007), pp. 2701–2772.
2. A. J. Davison *et al.*, *J. Gen. Virol.* **84**, 17 (2003).
3. E. Murphy, I. Rigoutsos, T. Shibuya, T. E. Shenk, *Proc. Natl. Acad. Sci. U.S.A.* **100**, 13585 (2003).
4. E. Murphy, T. Shenk, *Curr. Top. Microbiol. Immunol.* **325**, 1 (2008).
5. G. J. Zhang *et al.*, *J. Virol.* **81**, 11267 (2007).
6. D. Gatherer *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **108**, 19755 (2011).
7. J. Cao, A. P. Geballe, *Mol. Cell. Biol.* **16**, 7109 (1996).
8. T. Stamminger *et al.*, *J. Virol.* **76**, 4836 (2002).
9. B. J. Biegalke, E. Lester, A. Branda, R. Rana, *J. Virol.* **78**, 9579 (2004).
10. Z. Qian, B. Xuan, T. T. Hong, D. Yu, *J. Virol.* **82**, 3452 (2008).
11. L. Grainger *et al.*, *J. Virol.* **84**, 9472 (2010).
12. S. R. Starck *et al.*, *Science* **336**, 1719 (2012).
13. F. Robert *et al.*, *PLoS ONE* **4**, e5428 (2009).
14. N. T. Ingolia, L. F. Lareau, J. S. Weissman, *Cell* **147**, 789 (2011).
15. S. Lee *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **109**, 2424 (2012).
16. Materials and methods are available as supplementary materials on *Science* Online.
17. M. B. Reeves, A. A. Davies, B. P. McSharry, G. W. Wilkinson, J. H. Sinclair, *Science* **316**, 1345 (2007).
18. P. C. Lord, C. B. Rothschild, R. T. DeRose, B. A. Kilpatrick, *J. Gen. Virol.* **70**, 2383 (1989).
19. T. Joachim, in *Making Large Scale SVM Learning Practical*, B. Schölkopf, C. J. C. Burges, A. J. Smola, Eds. (MIT Press, Cambridge, MA, 1998), pp. 169–184.
20. M. E. Bordeleau *et al.*, *Chem. Biol.* **13**, 1287 (2006).
21. J. Cox *et al.*, *J. Proteome Res.* **10**, 1794 (2011).
22. F. S. Leach, E. S. Mocarski, *J. Virol.* **63**, 1783 (1989).
23. F. Sanger, S. Nicklen, A. R. Coulson, *Proc. Natl. Acad. Sci. U.S.A.* **74**, 5463 (1977).
24. W. Fiers *et al.*, *Nature* **273**, 113 (1978).
25. J. W. Yewdell, *Curr. Opin. Immunol.* **19**, 79 (2007).

Acknowledgments: We thank O. Mandelboim, D. Wolf, M. Trilling, A. Lauring, S. Karniely, and Weissman lab members for critical reading of the manuscript; C. Chu for assistance with sequencing; and J. Pelletier for providing Pateamine A. N.S.-G. is supported by a human frontier science program postdoctoral fellowship. This work was supported by the Howard Hughes Medical Institute (J.S.W.) and the Max Planck Society (M.M.). The Gene Expression Omnibus accession number for the data is GSE41605.

Supplementary Materials

www.sciencemag.org/cgi/content/full/338/6110/1088/DC1
Materials and Methods
Figs. S1 to S22
Tables S1 to S10
Files S1 to S3

25 July 2012; accepted 12 October 2012
10.1126/science.1227919

Egg Cell–Secreted EC1 Triggers Sperm Cell Activation During Double Fertilization

Stefanie Sprunck,[1]* Svenja Rademacher,[1]† Frank Vogler,[1] Jacqueline Gheyselinck,[2]‡ Ueli Grossniklaus,[2] Thomas Dresselhaus[1]

Double fertilization is the defining characteristic of flowering plants. However, the molecular mechanisms regulating the fusion of one sperm with the egg and the second sperm with the central cell are largely unknown. We show that gamete interactions in *Arabidopsis* depend on small cysteine-rich EC1 (EGG CELL 1) proteins accumulating in storage vesicles of the egg cell. Upon sperm arrival, EC1-containing vesicles are exocytosed. The sperm endomembrane system responds to exogenously applied EC1 peptides by redistributing the potential gamete fusogen HAP2/GCS1 (HAPLESS 2/GENERATIVE CELL SPECIFIC 1) to the cell surface. Furthermore, fertilization studies with *ec1* quintuple mutants show that successful male-female gamete interactions are necessary to prevent multiple–sperm cell delivery. Our findings provide evidence that mutual gamete activation, regulated exocytosis, and sperm plasma membrane modifications govern flowering plant gamete interactions.

Sexual reproduction depends on the successful union of two gametes of opposite sex at fertilization. In flowering plants such as *Arabidopsis thaliana*, sexual reproduction is distinct in that two gamete fusion events take place in a coordinated manner, a phenomenon termed "double fertilization" (*1*). Two nonmotile sperm cells are delivered by a pollen tube into the female gametophyte (embryo sac) that harbors two dimorphic female gametes (Fig. 1A). One sperm fuses with the egg cell, giving rise to the embryo, whereas the second sperm cell fuses with the central cell to develop the triploid endosperm. Although this distinct mode of reproduction was discovered more than a century ago (*2*, *3*) and recent live-cell imaging has shed some light on the behavior of *Arabidopsis* sperm nuclei during double fertilization (*4*), almost nothing is known about the molecules mediating gamete interactions. To date, the evolutionary conserved HAP2 (HAPLESS 2)/GCS1 (GENERATIVE CELL SPECIFIC 1) is the only sperm protein known to be essential at a late step in gamete interactions (*5*, *6*), and its role as a fusogen is strongly supported by the observation that HAP2-deficient gametes of *Chlamydomonas* and *Plasmodium berghei* are able to adhere but fail to fuse (*7*).

[1]Cell Biology and Plant Biochemistry, Biochemie-Zentrum Regensburg, University of Regensburg, Universitätsstrasse 31, D-93053 Regensburg, Germany. [2]Institute of Plant Biology and Zürich-Basel Plant Science Center, University of Zürich, Zollikerstrasse 107, CH-8008 Zürich, Switzerland.

*To whom correspondence should be addressed. E-mail: stefanie.sprunck@biologie.uni-regensburg.de
†Present address: Plant Breeding Center of Life and Food Sciences Weihenstephan, Technische Universität München, Emil-Ramann-Strasse 4, D-85354 Freising, Germany.
‡Present address: Département de Biologie Moléculaire Végétale, Université de Lausanne, Biophore 4403, CH-1015 Lausanne, Switzerland.

Based on a transcriptomics approach that uses isolated egg cells of wheat (*8*), we discovered a family of *Arabidopsis* genes with sequence similarity to the largest wheat egg cell–specific transcript cluster, termed EC-1 (Egg Cell 1) (see supplementary materials and methods). We found transcripts of the five *Arabidopsis EC1-like* genes (*EC1.1*, *EC1.2*, *EC1.3*, *EC1.4*, and *EC1.5*) only in female reproductive tissues (Fig. 1B). Egg cell–specificity of *EC1* was shown by expressing the β-glucuronidase (GUS) reporter under control of individual *EC1* promoters (Fig. 1, C and D) and by in situ hybridization to tissue sections. Transcripts of *EC1* genes are specifically present in the egg cell (Fig. 1, E to G) but are not detectable early after fertilization (Fig. 1H and fig. S1B), whereas GUS remains active in zygotes and early embryos (fig. S1, D to F).

EC1 proteins belong to the large and unexplored group of ECA1 (Early Culture Abundant 1) gametogenesis-related cysteine-rich proteins characterized by their conserved cysteine-spacing signature (*9*). Within 118 ECA1 proteins of *Arabidopsis*, the EC1 family forms a distinct subclade (fig. S2A). Notably, we identified *EC1*-like genes or transcripts only in flowering plant species, including the basal angiosperm *Amborella trichopoda*, and not in gymnosperms, ferns (*Adiantum sp.*), Bryophytes (*Physcomitrella patens*), or green algae (*Volvox sp.*; *Chlamydomonas sp.*). Protein-sequence analyses revealed common features of EC1 proteins, such as a predicted N-terminal signal peptide for secretion and a similar predicted intramolecular disulfide bond arrangement (Fig. 2A). Two conserved signature sequences, termed S1 and S2, were identified by multiple sequence alignments with representatives from monocots, dicots, and basal angiosperms (fig. S2B).

To investigate the subcellular localization of EC1, we stably expressed a translational fusion of EC1.1 and the green fluorescent protein (GFP) under control of the *EC1.1* promoter. The