SI Appendix

Cellular responses to human cytomegalovirus infection: induction of a mesenchymal-to-epithelial transition (MET) phenotype.

Adam Oberstein and Thomas Shenk

SI Materials and Methods

Cells and viruses. MRC-5 embryonic lung fibroblasts, ARPE-19 adult retinal pigment epithelial cells, and MDA-MB-231 cells were from the American Type Culture Collection. SUM1315MO2 breast cancer cells (1) were generously provided by Stephen Ethier (Medical University of South Carolina). All cell types were grown in DMEM supplemented with 10% fetal bovine serum, 1 mM sodium pyruvate, 2 mM glutamax (Gibco), 10 mM Hepes pH 7.4, 0.1 mM MEM Non-Essential Amino Acids (Gibco), 100 units/ml Penicillin G, and 100 µg/ml Streptomycin Sulfate. G166 and G179 glioma stem cell lines were obtained from Biorep (Milan, Italy) and cultured on laminin coated plates, as described previously (2). HCMV strains TB40/E-BAC4 (3), Merlin 1120 (4), and recombinant BAC derivatives were reconstituted by electroporation into ARPE-19 cells. Virus stocks were titered on MRC-5 and ARPE-19 cells by IE1 fluorescent focus assay (5).

Construction of recombinant TB40 and Merlin viruses. GFP and a porcine teschovirus-1 2A peptide (P2A) was inserted upstream of UL83 using galK recombineering (6). In the first step, TB40/E-BAC4 or Merlin BAC 1120 was transformed into *E.coli* strain SW102 and the UL83 start codon was replaced with an *em7*-promoter:*galK* cassette by homologous recombination. The *galK* cassette was amplified from plasmid pGalK and contained 50-bp homology arms flanking the UL83 start site. In the second step, *galK* was replaced with a GFP-P2A cassette amplified from a synthetic gene-block (IDT), restoring the UL83 start codon. GalK-negative clones were counterselected on 2-Deoxy-D-galactose (Sigma), and confirmed by restriction analysis and sequencing. Recombinants showed similar growth properties as parental strains. Recombineering primers and the GFP:P2A sequence are listed in *SI Appendix*, Table S6.

RNA sample preparation. For RNA-seq, cells were infected at a multiplicity of 3 infectious units/cell based upon the cognate titers determined on each cell type. At 24, 72, and 120 hpi cells were collected in Trizol reagent (Applied Biosystems) and total RNA was isolated using an RNeasy RNA isolation kit (Qiagen). DNA was removed using Turbo DNase (Thermo Fisher Scientific) and RNA was quantified using an Agilent 2100 Bioanalyzer. Sequencing libraries were prepared by the Penn State College of Medicine Genome Sciences Facility using a KAPA Stranded RNA-Seq Kit with RiboErase (Kappa Biosytems) and sequenced on an Illumina HiSeq 2500 DNA sequencer for 100 cycles in paired-end, rapid mode (100 bp x2).

RNA-seq data analysis. Concatenated human/HCMV fasta and annotation (.gtf) files were created for mapping and feature counting by combining sequences and annotations from gencode human genome Release 24 (GRCh38.p5) and TB40E-BAC4 (EF999921.1). HCMV long non-coding RNA's were annotated in TB40/E by querying the TB40/E genome with the corresponding sequences from HCMV strain Merlin (NC_006273.2) with tblastn (7). Sequencing fragments were mapped using HISAT2 version 2.0.3 (8). Greater than 93% of the reads were mapped for all samples. Strand-specific fragment counts were generated at the gene level using featureCounts (9), requiring a minimum

mapping quality score of 10 and concordantly mapped fragment ends on the same chromosome. To eliminate compositional bias, HCMV counts were removed prior to normalization and ratio determination. Ratio determination and significance calling was performed using DESeq2 (10) with dispersion estimates fit with local regression. For re-analysis of data from Tirosh et. al (11), fastq files from GEO accession GSE69906 were processed as described above except HCMV strain merlin (GU179001.1) was used and DESeq (12) was used with following settings for ratio modeling in the absence of biological replication: method="blind", sharingMode="fit-only", fitType="local".

Gene set enrichment analysis. Selected gene sets from the MSigDB v5.2 (13) and Harmonizome (14) were re-formatted and concatenated into one large .gmt file for gene set testing. Gene sets with less than 25 or greater than 1500 genes were removed from analysis. Gene set tests were implemented in the statistical computing language *R* version 3.3.1 (15). For each gene set, infected/mock ratios were subsetted for both MRC-5 and ARPE-19 at each time point. Each gene set distribution was then compared to the parent distribution, for each cell type, using the Wilcoxon rank-sum test (16); the null hypothesis being that the two distribution medians were equal. Wilcoxon *p*-values were corrected for multiple hypothesis testing using Storey's *q*-value (17). Gene sets with both MRC-5 and ARPE-19 *q*-values ≤ 0.05 were considered concordantly regulated. Gene set trends were assigned by comparing parental and gene set population means estimated from a two-sample *t* test with Welch's correction. Hierarchical clustering of gene sets was performed using the heatmap.2 function from the *R* package gplots version 3.0.1 (18).

Gene set overlap network generation and analysis. Our method for creating gene set overlap networks was based on the “Enrichment Map” method (19), except we implemented it *de novo* in *R* for greater flexibility. At each infection time point, overlap coefficients for all pair-wise concordant gene sets were calculated and converted into an adjacency matrix. The overlap coefficient was defined as in (19): $OC = A \cap B \,/\, \mathrm{Min}(A,B)$. Initial networks were generated in *R* using igraph version 1.1.0 (53) by creating nodes for each gene set and converting the overlap matrix into a list of weighted edges. Node sizes were proportional to their rank, which was determined by ranking the the product of each gene set's MRC-5 and ARPE-19 *q*-values. Edges with OC's less than 0.1 were removed and networks were exported for analysis using the rgexf *R* package version 0.15.3 (20). Networks were rendered and analyzed using Gephi version 0.9.1 (21). Force-directed networks were generated by filtering edges with overlap coefficients ≤ 0.4 and applying the Fruchterman-Reingold force-directed layout (22). Communities were detected by assigning each node to a modularity class using the Louvain algorithm (23), as implemented in Gephi. Communities were manually annotated by inspecting “pathway” oriented gene set libraries such as MSigDB Hallmarks, Kegg, and Reactome.

Creation of MET gene sets. For the Pattabiraman et. al (24) gene set, GMT files were created from RNA-seq ratios appearing in the authors' supplementary tables S7 and S8. For the comparison of chlolera toxin treated N8 cells (CTx-N8) vs. N8 cells, a list of regulated MET genes was extracted by filtering genes with fold-change ratios ≥ 4 and *p*-values ≤ 0.01. Remaining genes (3394) were then ranked by their *p*-values and the top 10% were used as an MET gene set for GSEA testing. For the two studies by Roca et al. (25, 26), GMT files were created from normalized expression matrices deposited in the Gene Expression Omnibis (accessions GSE48230 and GSE51975). For these data sets, gene sets were created by removing infinite fold-change values and extracting genes with fold-change ratios ≥ 4.

Quantitative reverse transcription polymerase chain reaction (qRT-PCR). RNA was isolated as described for RNA-seq above. cDNA was prepared using Superscript III reverse transcriptase (Invitrogen) and random hexamer primers. *q*RT-PCR reactions were performed using Power SYBR

Green Master Mix (Applied biosystems) and data was collected on a QuantStudio 6 Flex digital PCR machine (Life Technologies). Data was analyzed in biological triplicate using the ΔΔCT method (59) with *cyclophin A* (*PPIA*) as the reference gene. *CDH1* and *VIM* primers were from (27). All other primers were designed with either QuantPrime (28) or GETPrime (29). *q*RT-PCR primer sequences are listed in *SI Appendix*, Table S7.

UV and ganciclovir treatments. Virus was UV inactivated by three, 2 min-exposures to 1200 mJ/cm^2 of 254-nm light in a Stratalinker 2400 (Stratagene). Ganciclovir (Sigma) was dissolved at 40 mM in DMSO and added to cultures after the viral innoculation period at final concentration of 100 μM in complete medium. The vehicle, 0.25% DMSO, had no effect on viral replication.

Western blotting and immunofluorescence staining of E-cadherin. For western analysis cell monolayers were collected in RIPA buffer (50 mM Hepes, 150 mM NaCl, 1% NP-40, 0.5% sodium deoxycholate, 0.1% SDS, 5 ug/ml aprotinin, 10 ug/ml leupeptin, 1mM PMSF, pH 7.4), sonicated, and cleared by centrifugation. Protein concentration was determined using the BCA assay (Pierce) and 50 ug total protein per sample was separated by SDS-PAGE. Proteins were transferred to PVDF and blocked with 5% BSA in HBST (50 mM Hepes, 100 mM NaCl, 0.05% tween-20, pH 7.4). Primary antibodies (*SI Appendix*, Table S8) were diluted in 1% BSA. Dilutions were as follows: anti-E-cadherin (BD Biosciences) 1:500, anti-β-tubulin 1:2000, anti-EpCAM 1:500, anti-Vimentin 1:1000, anti-Fibronectin 1:10,000, anti-IE1 1:1000, anti-pp28 1:1000. Primary antibodies were detected using HRP-conjugated secondary antibodies. For immunofluorescence staining, monolayers were grown in 96-well tissue culture plates, fixed with 2% paraformaldehyde/PBS, and analyzed by indirect immunofluorescence using a goat antibody to the extracellular domain of E-cadherin (R & D Sytems), without permeabilization. Cells were blocked and primary antibody was diluted at 1:50 in PBS containing 5% donkey serum. Primary antibody was detected using an AlexaFluor-conjugated anti-goat secondary antibody and imaged on an Operetta high content imaging system (PerkinElmer).

Migration assays. SUM1315 or MDA-231 cells were infected with TB40*epi* or mock-infected at a multiplicity of 10 IU/cell based on the stock's ARPE-19 titer. At 120 hpi, cells were trypsinized and plated on 8.0 μm transwell filters or prepared for intracellular flow-cytometry analysis (see below). For the transwell migration assay, 5 x 10^4 cells were seeded onto each filter and induced to migrate towards medium containing 20% FBS. After 16 h at 37°C/5% CO_2, filters were fixed/stained with 0.2% crystal-violet in ethanol, non-migrated cells were removed with a cotton swab, and the bottom surface of the filter (migrated cells) was imaged by bright-field microscopy. Migrated cell number was quantified from 5 random fields, from three independent infections (n = 15), using CellProfiler (30) software.

Intracellular flow cytometry analysis. Cell monolayers were detached and stained with LIVE/DEAD Fixable Blue Dead Cell Stain (L/D Blue; ThermoFisher) for 20 min at room-temperature, fixed with 2% paraformaldehyde in PBS for 20 min at RT, and post-fixed/permeabilized in MeOH for 10 min on ice. Cells were then stained for 30 min at RT with 20 ul mouse anti-IE1 antibody 1B12 (5) in 100 ul permeabilization medium (ThermoFisher). IE1-stained cells were detected with an anti-mouse Alexa Fluor 647 (AF647) conjugated secondary antibody (ThermoFisher), and analyzed for L/D Blue, GFP:P2A:UL83, and AF647 on a LSRII Flow Cytometer (BD Biosciences). L/D Blue gates were established by comparing unstained with boiled/stained populations. GFP and AF647 gates were established by analyzing stained, mock-infected, cells.

Data Availability. RNA-seq data has been deposited in the NCBI Gene Expression Omnibus (GEO) under accession number GSE99454.

SI References

1. Neve RM, et al. (2006) A collection of breast cancer cell lines for the study of functionally distinct cancer subtypes. *Cancer Cell* 10(6):515–527.

2. Pollard SM, et al. (2009) Glioma Stem Cell Lines Expanded in Adherent Culture Have Tumor-Specific Phenotypes and Are Suitable for Chemical and Genetic Screens. *Cell Stem Cell* 4(6):568–580.

3. Sinzger C, et al. (2008) Cloning and sequencing of a highly productive, endotheliotropic virus strain derived from human cytomegalovirus TB40/E. *J Gen Virol* 89(2):359–368.

4. Stanton RJ, et al. (2010) Reconstruction of the complete human cytomegalovirus genome in a BAC reveals RL13 to be a potent inhibitor of replication. *J Clin Invest* 120(9):3191–3208.

5. Zhu H, Shen Y, Shenk T (1995) Human cytomegalovirus IE1 and IE2 proteins block apoptosis. *J Virol* 69(12):7960–7970.

6. Warming S (2005) Simple and highly efficient BAC recombineering using galK selection. *Nucleic Acids Res* 33(4):e36–e36.

7. Camacho C, et al. (2009) BLAST+: architecture and applications. *BMC Bioinformatics* 10(1):421.

8. Kim D, Langmead B, Salzberg SL (2015) HISAT: a fast spliced aligner with low memory requirements. *Nat Methods* 12(4):357–360.

9. Liao Y, Smyth GK, Shi W (2014) featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics* 30(7):923–930.

10. Love MI, Huber W, Anders S (2014) Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol* 15(12). doi:10.1186/s13059-014-0550-8.

11. Tirosh O, et al. (2015) The Transcription and Translation Landscapes during Human Cytomegalovirus Infection Reveal Novel Host-Pathogen Interactions. *PLOS Pathog* 11(11):e1005288.

12. Anders S, Huber W (2010) Differential expression analysis for sequence count data. *Genome Biol* 11(10):R106.

13. Liberzon A, et al. (2011) Molecular signatures database (MSigDB) 3.0. *Bioinformatics* 27(12):1739–1740.

14. Rouillard AD, et al. (2016) The harmonizome: a collection of processed datasets gathered to serve and mine knowledge about genes and proteins. *Database* 2016:baw100.

15. R Development Core Team (2011) *R: A language and environment for statistical computing.* (R Foundation for Statistical Computing., Vienna, Austria) Available at: http://www.R-project.org/.

16. Mann HB, Whitney DR (1947) On a Test of Whether one of Two Random Variables is

Stochastically Larger than the Other. *Ann Math Stat* 18(1):50–60.

17. Storey JD, Tibshirani R (2003) Statistical significance for genomewide studies. *Proc Natl Acad Sci* 100(16):9440–9445.

18. Warnes GR, et al. (2016) *gplots: Various R Programming Tools for Plotting Data* Available at: https://CRAN.R-project.org/package=gplots.

19. Merico D, Isserlin R, Stueker O, Emili A, Bader GD (2010) Enrichment Map: A Network-Based Method for Gene-Set Enrichment Visualization and Interpretation. *PLoS ONE* 5(11):e13984.

20. Yon GV, Lacoa JF, Kunst JB (2015) *rgexf: Build, Import and Export GEXF Graph Files* Available at: https://CRAN.R-project.org/package=rgexf.

21. Bastian M, Heymann S, Jacomy M, others (2009) Gephi: an open source software for exploring and manipulating networks. *ICWSM* 8:361–362.

22. Fruchterman TM, Reingold EM (1991) Graph drawing by force-directed placement. *Softw Pract Exp* 21(11):1129–1164.

23. Blondel VD, Guillaume J-L, Lambiotte R, Lefebvre E (2008) Fast unfolding of communities in large networks. *J Stat Mech Theory Exp* 2008(10):P10008.

24. Pattabiraman DR, et al. (2016) Activation of PKA leads to mesenchymal-to-epithelial transition and loss of tumor-initiating ability. *Science* 351(6277):aad3680-aad3680.

25. Roca H, et al. (2013) Transcription Factors OVOL1 and OVOL2 Induce the Mesenchymal to Epithelial Transition in Human Cancer. *PLoS ONE* 8(10):e76773.

26. Roca H, et al. (2014) A bioinformatics approach reveals novel interactions of the OVOL transcription factors in the regulation of epithelial–mesenchymal cell reprogramming and cancer progression. *BMC Syst Biol* 8(1):1.

27. Mani SA, et al. (2008) The Epithelial-Mesenchymal Transition Generates Cells with Properties of Stem Cells. *Cell* 133(4):704–715.

28. Arvidsson S, Kwasniewski M, Riano-Pachon DM, Mueller-Roeber B (2008) QuantPrime - a flexible tool for reliable high-throughput primer design for quantitative PCR. *BMC Bioinformatics* 9(1):465.

29. Gubelmann C, et al. (2011) GETPrime: a gene- or transcript-specific primer database for quantitative real-time PCR. *Database* 2011(0):bar040-bar040.

30. Carpenter AE, et al. (2006) CellProfiler: image analysis software for identifying and quantifying cell phenotypes. *Genome Biol* 7(10):R100.

31. Silva MC, Yu Q-C, Enquist L, Shenk T (2003) Human Cytomegalovirus UL99-Encoded pp28 Is Required for the Cytoplasmic Envelopment of Tegument-Associated Capsids. *J Virol* 77(19):10594–10605.

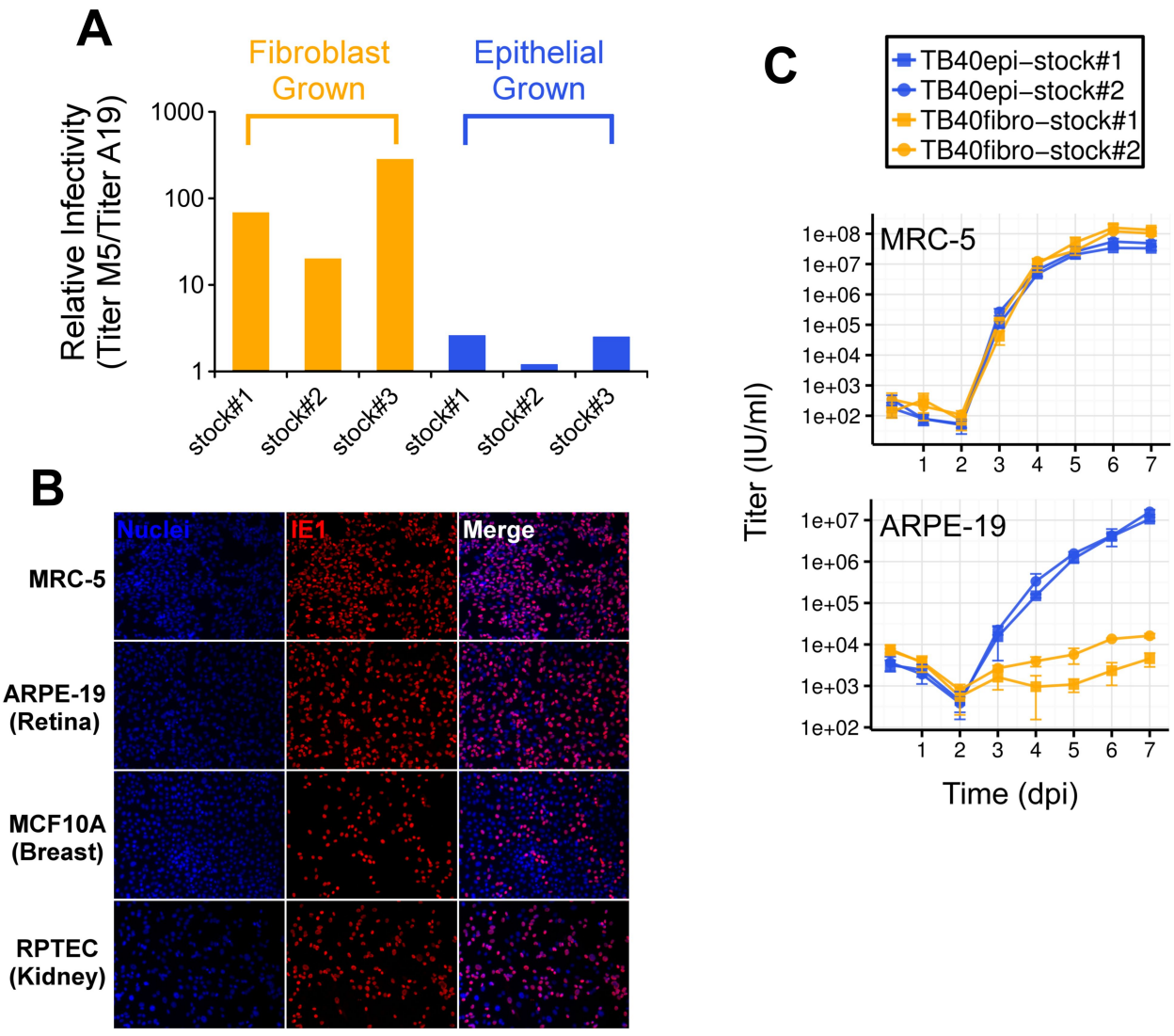

Figure S1. Growth of HCMV in ARPE-19 cells maintains epithelial tropism (Related to Figure 1). (A) Relative infectivity of three independently produced HCMV stocks grown in different hosts on MRC-5 (M5) and ARPE-19 (A19) cells. Titers were determined by assaying viral IE1 expression at 24 hpi on either MRC-5 or ARPE-19 cells. (B) Immunofluorescent staining of all nuclei (Hoescht; blue) and infected nuclei (IE1; red) showing efficient infection of various primary epithelial cell types with TB40*epi.* Infections were performed at a multiplicity of 3 infectious units/cell, based on titers determined on ARPE-19 cells. (C) Single-step growth curves of HCMV stocks grown in different host cells. Infections were performed at a multiplicity of 3 infectious units/cell, based on titers determined on MRC-5 cells. Cell-free virus was collected at indicated days post infection and titered by IE1 fluorescent focus assay on MRC-5 cells. Data are mean ± S.D. from three biological replicates.

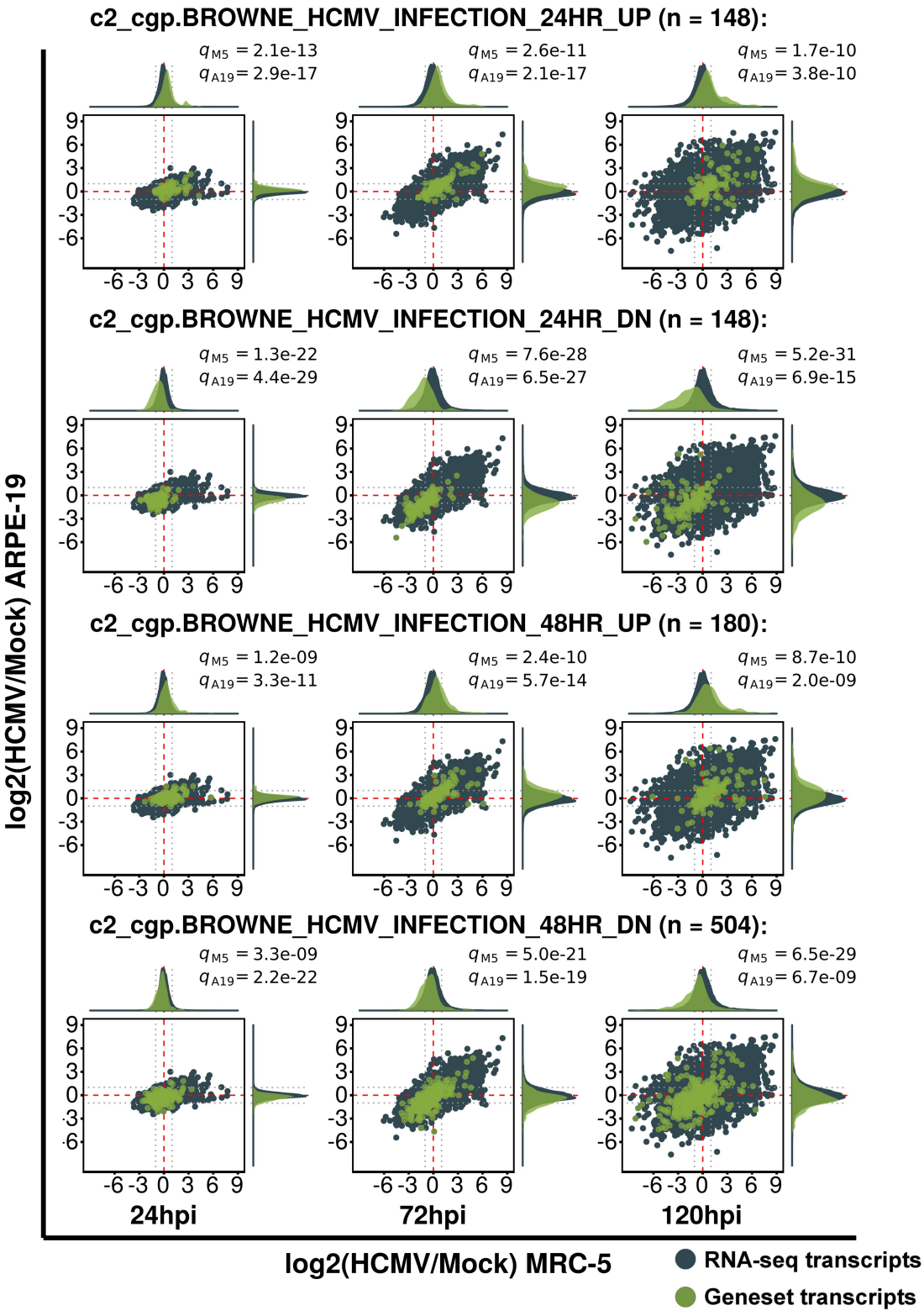

Figure S2. GSEA analysis of HCMV microarray gene sets from MSigDB module C2 (Related to Figure 1). RNA-seq ratio scatter plots comparing transcript fold-change ratios of HCMV microarray gene sets from MSigDB (green) with ratios determined using RNA-sequencing (blue). Distribution density plots used for significance testing are displayed above and to the right of each scatter plot. Dotted gray lines are 2-fold change reference guides.

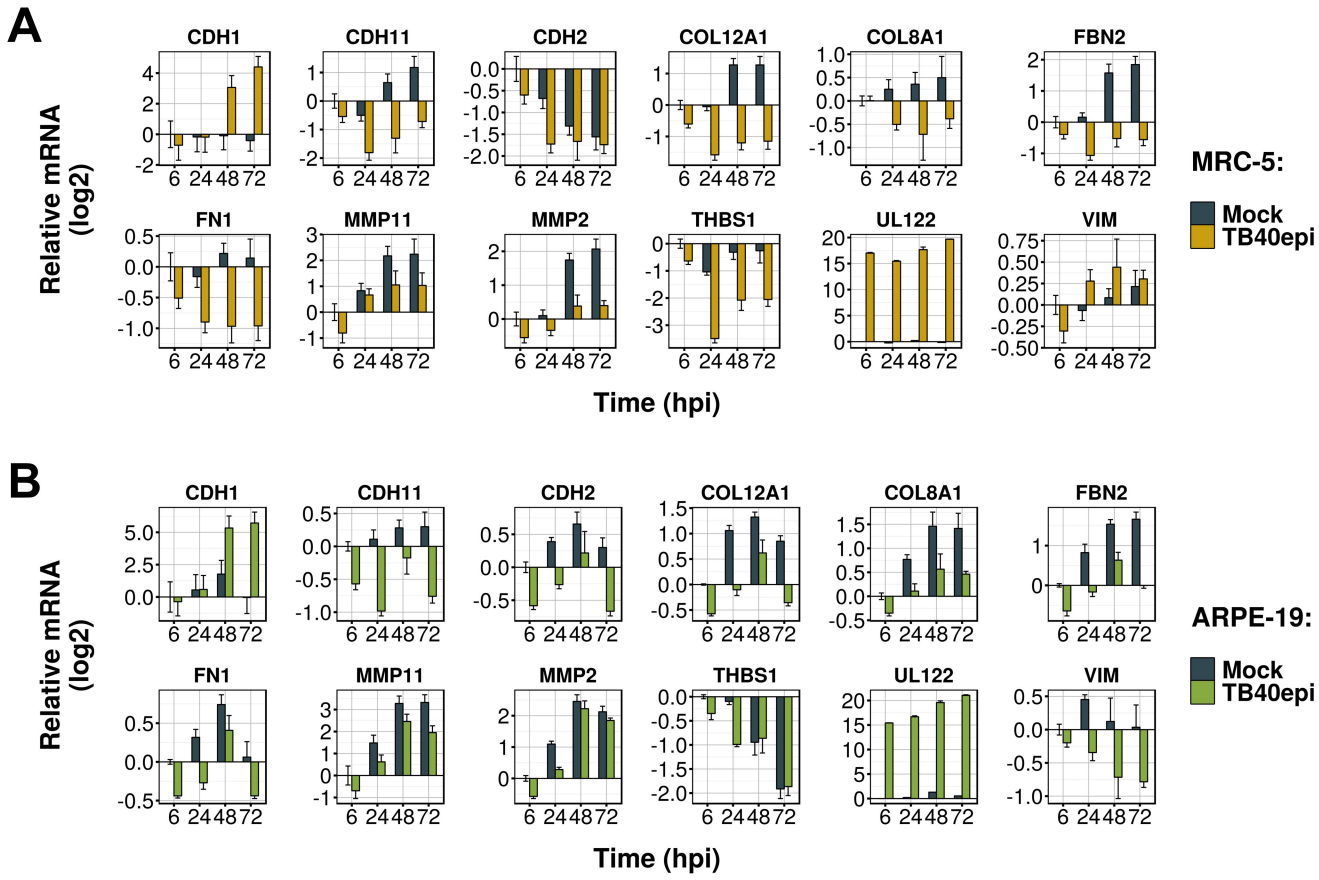

Figure S3. *q*RT-PCR validation of EMT inhibition and *CDH1* induction during HCMV infection (Related to Figure 6). (A) Relative mRNA levels over a 72 h infection time course in MRC-5 cells. (B) Same as panel *A*, except for ARPE-19 cells. UL122 cycle threshold (Ct) values were arbitrarily set to 37.0 for mock samples in order to calculate finite fold-change ratios. UL122 was used solely to mark infected samples.

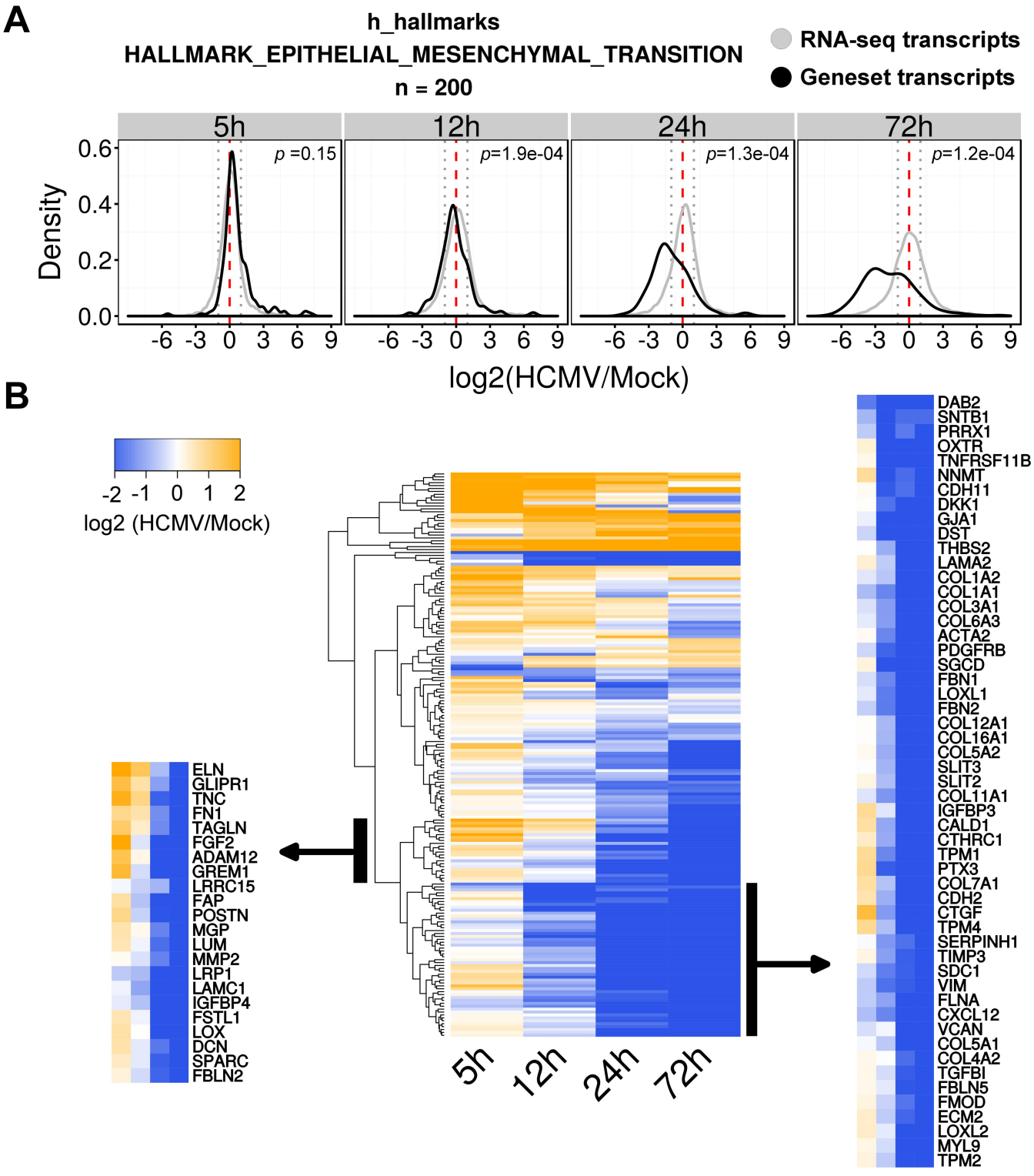

Figure S4. Validation of EMT inhibition using RNA-seq data from Tirosh et al., 2015 (Related to Figure 6). (A) Comparison of RNA-seq (*grey*) and MSigDB "HALLMARK EPITHELIAL MESENCHYMAL TRANSITION" gene-set (*black*) kernel densities showing downregulation of EMT genes during HCMV infection. (B and C) Heatmaps showing identities of genes from the EMT gene set tested in panel *A*.

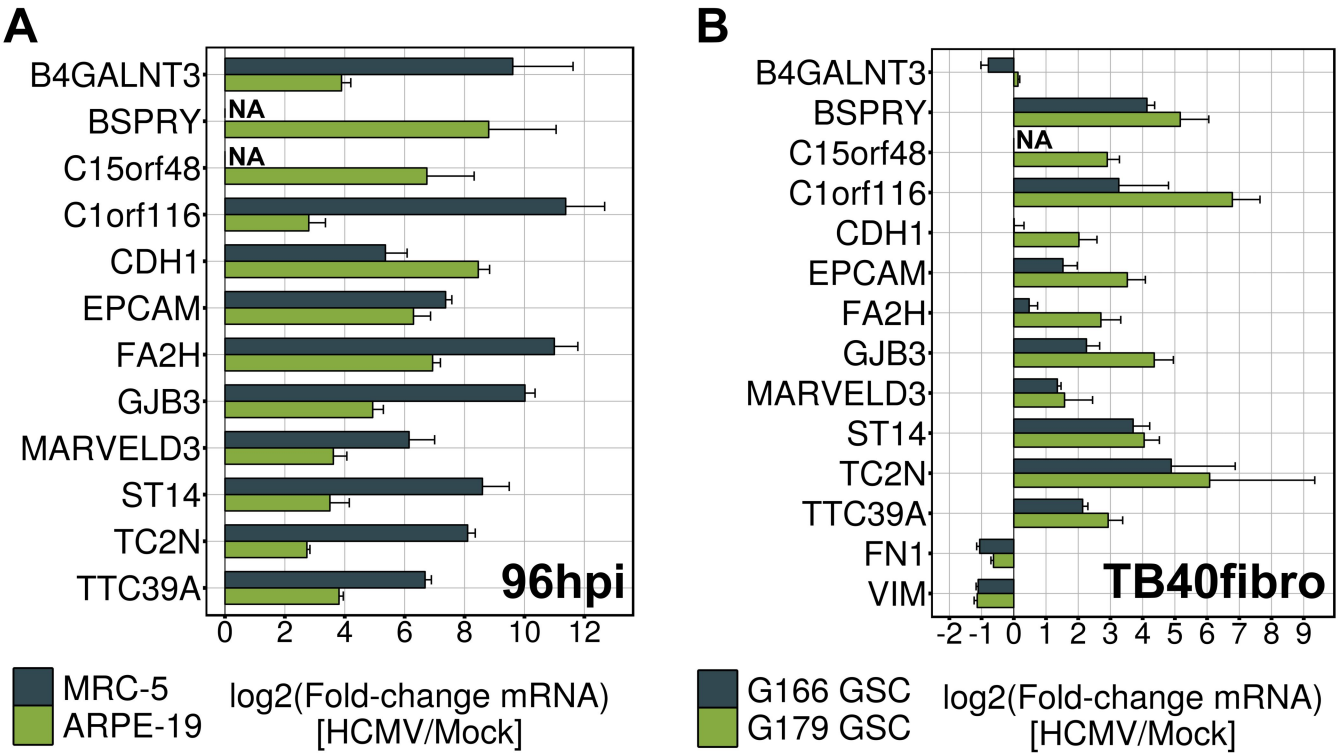

Figure S5. Analysis of HCMV-MET signature genes by *q*RT-PCR (Related to Figure 7). (A) Relative RNA levels of MET-signature genes in MRC-5 and ARPE-19 cells infected at a multiplicity of 3 infectious units/cell with TB40*epi.* RNA samples were collected at 96 hpi. Data are mean ± S.D. from three separate infections, each assayed in technical triplicate. (B) Relative mRNA levels of MET-signature genes in the glioma stem cells lines G166 and G179. Infections were performed at multiplicity of 3 infectious units/cell with TB40*fibro* and RNA samples were collected at 6 dpi. Data are mean ± S.D. from one representative infection assayed in technical quadruplicate.

Table S1. Validation of gene set testing procedure (Related to Figure S2). Ranks of HCMV microarray gene sets from MSigDB module c2.cgp. Gene set distributions were compared to RNA-seq distributions and ranked by the inverse product of their MRC-5 and ARPE-19 *q*-values. Ranks within top 5% (rank ≤ 116) of all gene set tests are appear in bold.

	24hpi			72hpi			120hpi		
Geneset	**q-value MRC-5**	**q-value ARPE-19**	**Rank***	**q-value MRC-5**	**q-value ARPE-19**	**Rank***	**q-value MRC-5**	**q-value ARPE-19**	**Rank***
BROWNE HCMV INFECTION 24HR DN	1.13E−22	1.41E−29	**3**	1.87E−28	1.48E−27	**12**	1.73E−31	6.67E−15	**18**
BROWNE HCMV INFECTION 14HR DN	9.82E−17	2.12E−27	**5**	2.31E−18	1.00E−15	**43**	1.63E−18	2.62E−07	**57**
BROWNE HCMV INFECTION 16HR UP	3.01E−21	5.13E−17	**10**	8.41E−14	3.76E−16	**54**	1.29E−06	3.47E−12	**106**
BROWNE HCMV INFECTION 20HR DN	4.87E−16	1.99E−21	**11**	1.32E−15	2.61E−17	**47**	3.07E−19	5.22E−06	**60**
BROWNE HCMV INFECTION 20HR UP	5.22E−19	1.53E−17	**13**	2.34E−09	5.92E−13	**101**	1.08E−06	4.96E−10	133
BROWNE HCMV INFECTION 48HR DN	2.05E−09	4.92E−23	**15**	1.65E−21	4.31E−20	**32**	1.93E−29	6.58E−09	**30**
BROWNE HCMV INFECTION 24HR UP	1.32E−13	7.34E−18	**19**	9.14E−12	6.54E−18	**56**	6.81E−11	4.08E−10	**83**
BROWNE HCMV INFECTION 14HR UP	4.34E−18	4.49E−12	**23**	1.67E−14	1.54E−13	**64**	1.26E−07	4.85E−12	**96**
BROWNE HCMV INFECTION 18HR DN	1.43E−11	5.23E−18	**25**	8.86E−20	1.64E−15	**41**	9.92E−19	1.64E−07	**56**
BROWNE HCMV INFECTION 16HR DN	4.18E−09	1.64E−15	**43**	2.92E−14	9.34E−17	**50**	3.83E−16	1.00E−07	**69**
BROWNE HCMV INFECTION 12HR UP	4.12E−15	1.36E−07	**52**	8.49E−08	3.56E−06	217	6.66E−04	5.12E−06	304
BROWNE HCMV INFECTION 18HR UP	4.71E−12	6.23E−10	**55**	3.08E−09	3.78E−13	**100**	5.41E−07	1.64E−08	150
BROWNE HCMV INFECTION 48HR UP	7.02E−10	1.01E−11	**59**	8.65E−11	2.02E−14	**78**	3.48E−10	2.02E−09	**97**
BROWNE HCMV INFECTION 12HR DN	8.18E−07	6.95E−14	**61**	3.64E−08	6.05E−11	129	2.38E−10	1.19E−03	176
BROWNE HCMV INFECTION 8HR UP	4.17E−12	1.77E−05	**76**	2.06E−04	5.94E−03	476	5.50E−02	2.12E−04	580
BROWNE HCMV INFECTION 10HR UP	6.93E−10	4.37E−04	122	2.31E−06	4.24E−04	307	6.76E−03	1.72E−04	480
BROWNE HCMV INFECTION 2HR DN	2.02E−06	7.57E−05	188	4.60E−02	1.45E−02	897	3.88E−02	1.21E−01	1252
BROWNE HCMV INFECTION 8HR DN	3.42E−03	5.33E−06	271	8.57E−04	4.81E−05	376	3.59E−04	2.01E−02	560
BROWNE HCMV INFECTION 6HR DN	2.31E−02	3.56E−06	314	9.70E−04	1.22E−05	347	7.37E−04	3.07E−03	503
BROWNE HCMV INFECTION 10HR DN	4.65E−03	4.44E−05	348	1.55E−02	1.86E−03	616	7.60E−04	2.36E−02	604
BROWNE HCMV INFECTION 4HR UP	8.50E−06	2.12E−01	423	4.35E−02	1.17E−01	1196	2.98E−01	3.25E−01	2227
BROWNE HCMV INFECTION 6HR UP	6.11E−03	4.11E−02	680	1.20E−01	2.48E−02	1103	2.06E−02	1.52E−01	1177
BROWNE HCMV INFECTION 2HR UP	2.49E−03	3.23E−01	809	1.07E−01	1.13E−01	1409	4.78E−02	2.44E−01	1487
BROWNE HCMV INFECTION 30MIN UP	2.33E−03	3.88E−01	825	2.63E−01	3.63E−01	2216	2.22E−01	2.42E−02	1279
BROWNE HCMV INFECTION 1HR DN	3.16E−03	3.76E−01	859	4.37E−02	2.56E−02	955	2.67E−01	2.82E−02	1361
BROWNE HCMV INFECTION 4HR DN	1.20E−01	8.60E−02	1281	5.80E−02	2.07E−01	1405	3.09E−02	4.06E−01	1509
BROWNE HCMV INFECTION 30MIN DN	8.30E−02	1.52E−01	1329	2.53E−01	2.49E−01	2002	1.45E−01	4.96E−02	1348
BROWNE HCMV INFECTION 1HR UP	3.83E−01	2.71E−01	2125	1.84E−01	1.73E−01	1693	2.83E−01	3.76E−01	2253

*Rank of 2312 gene sets in MSigDB c2.cgp passing size filtering (see methods).

Table S2. Gene-set network community statistics at 24 hpi (Related to Figure 2). Statistics for concordantly regulated gene-set communities larger than 0.5% of total network nodes at 24 hpi. Leading gene-sets are unique identifiers used to color communities during network analysis.

Community	Nodes	% of total	Trend	Up nodes	Down nodes	RGB color*	Leading gene-set
Translation	228	20.25%	down	0	228	[106, 61, 154]	CYTOPLASMIC RIBOSOMAL PROTEINS HOMO SAPIENS
Adhesion#1	115	10.21%	down	3	112	[177, 89, 40]	HALLMARK EPITHELIAL MESENCHYMAL TRANSITION
Ion and small molecule transport	100	8.88%	up	100	0	[166, 206, 227]	VOLTAGE GATED POTASSIUM CHANNELS
MYC/mTOR/ESR signaling	91	8.08%	down	31	60	[51, 160, 44]	HALLMARK MYC TARGETS V2
Adhesion#2	28	2.49%	down	4	24	[207, 119, 70]	INTEGRIN LINKED KINASE SIGNALING
Sphingolipid metabolism	24	2.13%	up	22	2	[31, 120, 180]	GLYCEROPHOSPHOLIPID METABOLISM
Nucelosome assembly	20	1.78%	down	0	20	[244, 164, 96]	NUCLEOSOME ASSEMBLY
Mitochondrial translation	12	1.07%	up**	5	7	[0, 0, 0]	MITOCHONDRIAL TRANSLATION
Intraflagellar transport	11	0.98%	down	0	11	[150, 150, 150]	INTRAFLAGELLAR TRANSPORT
T/B-cell receptor pathway	11	0.98%	up	8	3	[170, 170, 170]	B CELL RECEPTOR SIGNALING PATHWAY HOMO SAPIENS
DNA replication #1	9	0.80%	up	5	4	[227, 26, 28]	DNA REPLICATION PRE INITIATION
PRC2 mediated H3K27Me3	8	0.71%	up	8	0	[255, 127, 0]	HESC H3K27ME3 20682450
Stress response	7	0.62%	up	7	0	[178, 223, 138]	HSF1 DEPENDENT TRANSACTIVATION
RNA helicase activity	7	0.62%	up	7	0	[74, 118, 110]	ATP DEPENDENT RNA HELICASE ACTIVITY

*As shown is Figure 2
** Manually assigned

Table S3. Gene-set network community statistics at 72 hpi (Related to Figure 2). Statistics for concordantly regulated gene-set communities larger than 0.5% of total network nodes at 72 hpi. Leading gene-sets are unique identifiers used to color communities during network analysis.

Community	Nodes	% of total	Trend	Up nodes	Down nodes	RGB color*	Leading gene-set
Adhesion#1	218	10.61%	down	3	215	[177, 89, 40]	HALLMARK EPITHELIAL MESENCHYMAL TRANSITION
Translation	143	6.96%	down	3	140	[106, 61, 154]	CYTOPLASMIC RIBOSOMAL PROTEINS HOMO SAPIENS
DNA replication #1	143	6.96%	up	140	3	[227, 26, 28]	HALLMARK E2F TARGETS
DNA replication #2	126	6.13%	up	119	7	[251, 154, 153]	DNA REPLICATION HOMO SAPIENS
Ion and small molecule transport	99	4.82%	up	69	0	[166, 206, 227]	VOLTAGE GATED POTASSIUM CHANNELS
TGF beta signaling	74	3.60%	down	1	73	[0, 255, 255]	TGF BETA SIGNALING PATHWAY HOMO SAPIENS
Hypoxia and interferon signaling	74	3.60%	down	0	74	[202, 178, 214]	ELVIDGE HYPOXIA UP
MYC/mTOR/ESR signaling	62	3.02%	up	58	4	[51, 160, 44]	HALLMARK MYC TARGETS V2
Adhesion#2	42	2.04%	down	17	25	[207, 119, 70]	ANCHORING JUNCTION
Stress response	41	2.00%	up	32	9	[178, 223, 138]	UNFOLDED PROTEIN BINDING
Monoamine GPCR signaling	27	1.31%	up	25	2	[255, 0, 255]	MONOAMINE GPCRS HOMO SAPIENS
Protein glycosylation	26	1.27%	up	26	0	[255, 255, 0]	PROTEIN GLYCOSYLATION
PRC2 mediated H3K27Me3	25	1.22%	up	25	0	[255, 127, 0]	HESC H3K27ME3 20682450
ATF/CREB-family TF targets	24	1.17%	up	24	0	[255, 99, 71]	VdollarATF4 Q2
Apoptosis	21	1.02%	down	0	21	[178, 34, 34]	HALLMARK APOPTOSIS
E2F-family TF targets	20	0.97%	up	20	0	[0, 255, 0]	VdollarE2F Q4 01
Nuclear protein import	16	0.78%	down	1	15	[218, 165, 32]	REGULATION OF PROTEIN LOCALIZATION TO NUCLEUS
Unknown#1	13	0.63%	up	12	1	[255, 255, 153]	MORF FLT1

*As shown is Figure 2

Table S4. Gene-set network community statistics at 120 hpi (Related to Figure 2). Statistics for concordantly regulated gene-set communities larger than 0.5% of total network nodes at 120 hpi. Leading gene-sets are unique identifiers used to color communities during network analysis.

Community	Nodes	% of total	Trend	Up nodes	Down nodes	RGB color*	Leading gene-set
Translation	227	11.06%	down	4	223	[106, 61, 154]	CYTOPLASMIC RIBOSOMAL PROTEINS HOMO SAPIENS
DNA replication #1	169	8.24%	up	166	3	[227, 26, 28]	HALLMARK E2F TARGETS
DNA replication #2	139	6.77%	up	139	0	[251, 154, 153]	CELL CYCLE HOMO SAPIENS
Adhesion#1	137	6.68%	down	2	135	[177, 89, 40]	HALLMARK EPITHELIAL MESENCHYMAL TRANSITION
TGF beta signaling	124	6.04%	down	59	65	[0, 255, 255]	TGF BETA SIGNALING PATHWAY HOMO SAPIENS
Ion and small molecule transport	104	5.07%	up	102	2	[166, 206, 227]	VOLTAGE GATED POTASSIUM CHANNELS
Monoamine GPCR signaling	68	3.31%	up	68	0	[255, 0, 255]	MONOAMINE GPCRS HOMO SAPIENS
Adhesion#2	65	3.17%	down	28	37	[207, 119, 70]	CELL SUBSTRATE ADHERENS JUNCTION
MYC/mTOR/ESR signaling	47	2.29%	up	43	4	[51, 160, 44]	HALLMARK MYC TARGETS V2
PRC2 mediated H3K27Me3	39	1.90%	up	39	0	[255, 127, 0]	HESC H3K27ME3 20682450
Cholesterol biosynthesis	38	1.85%	up	32	6	[189, 183, 107]	METABOLISM OF LIPIDS AND LIPOPROTEINS
Unknown#1	33	1.61%	up	33	0	[255, 255, 153]	MORF FLT1
E2F-family TF targets	24	1.17%	up	24	0	[0, 255, 0]	VdollarE2F Q4 01
Nucelosome assembly	21	1.02%	up	21	0	[244, 164, 96]	NUCLEOSOME ASSEMBLY
Microtubule motor activity	20	0.97%	up	20	0	[0, 0, 128]	KINESINS
Unknown#2	13	0.63%	up	11	2	[255, 255, 153]	GNF2 RTN1

*As shown is Figure 2

Table S5. RNA-seq ratios of selected epithelial genes (Related to Figure 6). RNA-seq fold-change ratios from DESeq2. Ratios are on a log2 scale and represent normalized fragment counts from HCMV-infected / Mock-infected samples, at each time point. An asterisk (*) indicates an adjusted ratio p-value ≤ 0.05, and concordantly regulated ratios (i.e. ratios with p-values ≤ 0.05 for both MRC-5 and ARPE-19 cells at each time point) are shown in bold.

Gene	log2FC MRC-5 24hpi	log2FC ARPE-19 24hpi	log2FC MRC-5 72hpi	log2FC ARPE-19 72hpi	log2FC MRC-5 120hpi	log2FC ARPE-19 120hpi
CLDN6	0.5	0.5	**4.0***	**3.1***	**3.9***	**4.1***
EPCAM	2.7*	-0.2	**5.9***	**4.8***	**6.6***	**5.5***
PKP1	1.4	1.0	**2.8***	**2.3***	**4.4***	**2.0***
CRB3	-0.9	0.1	0.1	0.7	**4.6***	**2.1***
GJB2	1.3	0.6	**4.9***	**2.7***	**3.6***	**3.8***
GJB3	0.5	0.4	**3.7***	**2.9***	**5.3***	**2.0***
KRT8	-1.3	0.0	-2.4	-0.2	**-1.2***	**-2.3***
KRT18	-1.2	-0.1	-1.4	-0.3	1.4*	-2.2*
TJP1	-0.2	-0.5*	**-0.8***	**-0.9***	-0.6*	-0.5
PARD6A	-0.6	0.4	-1.1	0.6	1.9*	0.5
CTNNB1	0.0	-0.2	-0.2	0.0	-0.3	0.2

Table S6. Recombineering primers and GFP-P2A geneblock sequence (Related to Figures 6 & 7). Recombineering primers used to construct TB40-GFP:P2A:UL83 and Merlin-GFP:P2A:UL83. Sequences homologous to the HCMV genome are shown in uppercase and sequences homologous to the plasmid pGalK or the geneblock GFPA206K-P2A are shown in lowercase. The restored UL83 start codon is highlight in bold font. The full sequence of the synthetic GFPA206K-P2A geneblock is also listed.

Primer name	Sequence
UL83galk-F	GCCGCAGAGGGCGCGCCGCTCAGTCGCCTACACCCGTACGCGCAGGCA GCcctgttgacaattaatcatcggca
UL83galk-R	ATGGGACCCAGTACGGATATCATTTCGGGACAACGGCGACCGCGCGACTC tcagcactgtcctgctccttgt
UL83gfp-F	GCCGCAGAGGGCGCGCCGCTCAGTCGCCTACACCCGTACGCGCAGGCA GCatggtgagcaagggcgagg
UL83gfp-R	ATGGGACCCAGTACGGATATCATTTCGGGACAACGGCGACCGCGCGACTC **CAT**aggtccagggttctcctccac
GFPA206K-P2A	ATGGTGAGCAAGGGCGAGGAGCTGTTCACCGGGGTGGTGCCCATCCTGG TCGAGCTGGACGGCGACGTAAACGGCCACAAGTTCAGCGTGTCCGGCGA GGGCGAGGGCGATGCCACCTACGGCAAGCTGACCCTGAAGTTCATCTGC ACCACCGGCAAGCTGCCCGTGCCCTGGCCCACCCTCGTGACCACCCTGA CCTACGGCGTGCAGTGCTTCAGCCGCTACCCCGACCACATGAAGCAGCA CGACTTCTTCAAGTCCGCCATGCCCGAAGGCTACGTCCAGGAGCGCACC ATCTTCTTCAAGGACGACGGCAACTACAAGACCCGCGCCGAGGTGAAGTT CGAGGGCGACACCCTGGTGAACCGCATCGAGCTGAAGGGCATCGACTTC AAGGAGGACGGCAACATCCTGGGGCACAAGCTGGAGTACAACTACAACA GCCACAACGTCTATATCATGGCCGACAAGCAGAAGAACGGCATCAAGGTG AACTTCAAGATCCGCCACAACATCGAGGACGGCAGCGTGCAGCTCGCCG ACCACTACCAGCAGAACACCCCCATCGGCGACGGCCCCGTGCTGCTGCC CGACAACCACTACCTGAGCACCCAGTCCAAGCTGAGCAAAGACCCCAAC GAGAAGCGCGATCACATGGTCCTGCTGGAGTTCGTGACCGCCGCCGGGA TCACTCTCGGCATGGACGAGCTGTACAAGggaagcggagctactaacttcagcctgctg aagcaggctggagacgtggaggagaaccctggacct

Table S7. *q*RT-PCR primers used in this study (Related to Figures 6, 7, S3, and S5).

Primer name	Sequence	Source
VIM-fwd	GAGAACTTTGCCGTTGAAGC	(Mani et al., 2008)
VIM-rev	GCTTCCTGTAGGTGGCAATC	(Mani et al., 2008)
CDH1-fwd	TGTTCACCATTAACAGGAACAC	This study
CDH1-rev	GGGTATACGTAGGGAAACTCTC	This study
CDH2-fwd	TCAGTGAAGGAGTCAGCAG	This study
CDH2-rev	CTTCTGCCTTTGTAGGTGG	This study
CDH11-fwd	GCTGACTTGTGAATGGGAC	This study
CDH11-rev	TTGAGCTCATCACGTCAGG	This study
COL8A1-fwd	AGAGTATCCACACCTACCC	This study
COL12A1-fwd	ACTTGAAGAACGTGGTTCTC	This study
COL12A1-rev	GCAGATGTCCAATAAGCCC	This study
COL8A1-rev	GGCTAATGGTATTTCTTTGCC	This study
FBN2-fwd	TGTGGACAACCTGTCTGTG	This study
FBN2-rev	TGAACCCATAAACACAAGCAC	This study
FN1-fwd	AAATGGCCAGATGATGAGC	This study
FN1-rev	TAACACGTTGCCTCATGAG	This study
MMP11-fwd	CACATACACATATGGAGACCC	This study
MMP11-rev	CTTTGCATGATGACGGAGG	This study
MMP2-fwd	ATACCATCGAGACCATGCG	This study
MMP2-rev	CCAATGATCCTGTATGTGATCTG	This study
PPIA-fwd	AGCCAGGTACTTGGTGCTACAGTC	This study
PPIA-rev	TGCAGGTAGTCTGCGCCTTAAC	This study
THBS1-fwd	CGATAGCCTCAACAACCGA	This study
THBS1-rev	GCCACCATCCTGTTTAAATCTC	This study
UL122-fwd	GCGTGGAGCCTCAAAGAATTGC	This study
UL122-rev	CAGCGTGGATGATCATGTTGCG	This study
UL99-fwd	ACATCACCACCATCAGCAAAGC	This study
UL99-rev	GGTGTCCTCTTCGTCTAAGTCGTC	This study
B4GALNT3-fwd	TTTGAAGACTGGTGTGGCA	This study
B4GALNT3-rev	CAGGGTTGTGCGAATATGG	This study
BSPRY-fwd	GCAGAAACTTGACACCATCC	This study
BSPRY-rev	TGCTCCTTCTGAATCAGCT	This study
C15orf48-fwd	AGGAAGGAACTCATTCCCT	This study
C15orf48-rev	CCAAAGAGAATACACAGCGA	This study
C1orf116-fwd	CTCTGTCTCCATCTCTGCC	This study
C1orf116-rev	GCTATCACTCTCCACTGGG	This study

Primer name	Sequence	Source
EPCAM-fwd	CGAGTGAGAACCTACTGGA	This study
EPCAM-rev	TGATCTCCTTCTGAAGTGCA	This study
FA2H-fwd	CTCCTACCTGTACAGCCTG	This study
FA2H-rev	TTAGTGCTGATACCAAATCCTG	This study
GJB3-fwd	CTCATCATTGAGTTCCTCTTCC	This study
GJB3-rev	GCATATTGAAGCCATGCCA	This study
MARVELD3-fwd	AGAGATATCTGCCCTCGAC	This study
MARVELD3-rev	TCTGACTGGTAATATTCCACCTC	This study
ST14-fwd	CATGGAACATTGAGGTGCC	This study
ST14-rev	ATCTCCACGTAGTCCTTGG	This study
TC2N-fwd	GGATGTTTCTATGGTGAAACAG	This study
TC2N-rev	ACTATTTGGGACTGCTGCT	This study
TTC39A-fwd	GGAAGTCTTCTGTAACAGATTCC	This study
TTC39A-rev	CGTGGATTTCCTCTTCAGTG	This study

Table S8. Antibodies used in this study.

ANTIBODY	SOURCE	IDENTIFIER
Mouse monoclonal anti-E-Cadherin	BD Biosciences	Cat#610181; RRID: AB_397580
Goat polyclonal anti-E-Cadherin	R & D Sytems	Cat#AF648; RRID: AB_355504
Goat polyclonal anti-EPCAM	R & D Sytems	Cat#AF960; RRID: AB_355745
Mouse monoclonal anti-β-tubulin	The Developmental Studies Hybridoma Bank	Cat#E7; RRID: AB_528499
Mouse monoclonal anti-Vimentin	The Developmental Studies Hybridoma Bank	Cat#AMF-17b; RRID: AB_528505
Rabbit polyclonal anti-Fibronectin	GeneTex	Cat# GTX112794; RRID: AB_1950298
Mouse monoclonal anti-IE1 (clone 1B12)	(5)	N/A
Mouse monoclonal anti-UL99 (clone 10B4-29)	(31)	N/A
Goat anti-mouse Alexa Fluor 647	ThermoFisher	Cat#A-21235; RRID: AB_2535804