

SI Appendix

SI Materials and Methods

Cells. Primary CD34⁺ hematopoietic progenitor cells (HPCs) were isolated from de-identified bone marrow, which is not human subject research (1, 2). CD34⁺ HPCs were isolated via positive selection with the CD34 MicroBead Kit (MACS, Miltenyi Biotec, San Diego, CA). Pure populations of CD34⁺ HPCs were cultured in MyeloCult H5100 (Stem Cell Technologies) and maintained in long-term co-culture with M2-10B4 and SI/SI murine stromal cells lines (kind gift from Stem Cell Technologies on behalf of D. Hogge, Terry Fox Laboratory, University of British Columbia, Vancouver, BC) (2, 3). Human lung fibroblasts (MRC-5) were obtained from ATCC (Manassas, VA) and maintained as previously described (4). All cells were maintained at 37°C with 5% CO₂.

Viruses. The TB40/E bacterial artificial chromosome (BAC) was previously engineered from the TB40/E strain of HCMV, which is known to exhibit a broad cell tropism including tropism for hematopoietic cells (5). All viruses used in the present study were previously engineered to express the green fluorescent protein (GFP) as a marker of infection and have been previously described and characterized (1, 4, 6, 7). Briefly, recombinant viruses were engineered using a two-step, positive/negative selection approach that leaves no trace of recombination. Desired mutations were confirmed via Sanger sequencing and BAC integrity was confirmed by enzyme digest analysis. BAC genomes were maintained in SW102 *E. coli*. Virus stocks were produced BAC genome transfection (15-20 µg), along with co-transfection of 2 µg of an *UL82*- encoding plasmid into 5 x 10⁶ MRC-5 fibroblasts and incubating until 100% cytopathic effects (CPE) were observed. Virus stocks were purified, stored, and titered as previously described (1).

Infection of CD34⁺ HPCs was performed as previously described (1). Briefly, CD34⁺ HPCs were infected with 2 MOI. At 24 hpi, cells were sorted via FACS (FACSARIA, BD Biosciences Immunocytometry Systems, San Jose, CA) to obtain a pure population (>97%) of infected CD34⁺ HPCs by utilizing a phycoerythrin conjugated CD34 (PE-CD34) specific antibody (BD Biosciences) and sorting for CD34⁺ and GFP⁺ (infected) cells. Infected CD34⁺ cells were co-cultured in transwells above irradiated (3000 rads, ¹³⁷Cs gammacell-40 irradiator type B, Atomic Energy of Canada LTD, Ottawa, Canada) M2-10B4 and SI/SI stromal cells until harvesting at 2, 6, and 10 days after infection. For ganciclovir (GCV; InVivoGen, San Diego, CA) treatment, a final concentration of 27 µM GCV was added to the media follow cell sorting. Infected CD34⁺ HPCs were harvested, counted, lysed in RNA/DNA lysis buffer (Zymo Research, Irvine, CA) and stored at -80°C until nucleic acid was isolated.

Fibroblasts were infected with 1 MOI of WT, $\Delta UL135$, or $\Delta UL138$ TB40/E viruses. At 12, 24, 48, and 72 hours post infection (hpi), infected cells were washed twice with phosphate buffered saline (PBS) and lysed in the well with ZR-*duet* RNA/DNA Lysis Buffer (Zymo Research, Irvine, CA). Lysed samples were stored at -80°C until all samples were collected and nucleic acid was isolated.

Variability of $\Delta UL135$ and $\Delta UL138$ transcriptomes. The featured kernel density variation of CD34⁺ HPC infection across different donors was quantified using gene-wise dispersion estimates (the metadata column dispGeneEst) in DESeq2 (8). Genes expressed within the two

groups (wave 1 and wave 2) across all available Δ UL135_6dpi and Δ UL138_6dpi transcriptomes (from donor 1-NS/SS, donor 2-NS and donor 3-SS libraries) were selected. By comparing the dispersion difference between the two subsets of genes using the Wilcoxon rank sum test provides a biological-source of variation involved but normalized measurement of the variation between Δ UL135 and Δ UL138 transcriptomes. Candidate genes contributing to latency-like (Δ UL135) or replication (Δ UL138) were finally restrictively identified by a shared subset between kernel/dispersion and DE metrics. Similar process was applied to the dataset of fibroblasts infection and Wilcoxon rank sum test was used to compare the low (wave 1) and moderate (wave 2) expression of viral genes in fibroblasts.

Expression levels in FPKM. In addition to rlog normalized counts from DESeq2 (8), we quantified expression levels of genes by merging annotated isoforms' FPKM values from Cufflinks (v2.2.1)(9). Pearson correlation coefficient of WT samples between SS and NS libraries from one cell donor was calculated based on \log_2 FPKM values, and validated by distance metric based hierarchical clustering of multiple samples with DESeq2 (8). For the comparison to clinical latency, expression levels (FPKM) from in vitro WT infection and in vivo human samples (both from SS libraries) were used. Furthermore, to leverage the previously identified viral latency- and replication-associated genes together with wave 1 and wave 2 genes, expression of these genes was divided by the geometric mean, using *geometric.mean* function from psych R-package (v1.6.4), of a set of concordant genes across *in vitro* and clinical samples, defined by the absolute \log_2 FC<0.5 for the two comparisons, WT_2 or 6dpi vs. clinical latency.

Detection of viral genomes. PBMCs were thawed and lysed and DNA was isolated using the ZR-Duet DNA/RNA purification kit (Zymo Research). 5 replicates of 700 ng of DNA was analyzed by real time PCR for the presence of HCMV genomes using a highly sensitive primer to the β 2.7 region of the genome as described previously (10). Genomes were quantified relative to the RNaseP cellular housekeeping gene.

Detection of virus in donor plasma. Donor plasma (1mL) from each PBMC donor was diluted 1:2 in culture media and incubated with a monolayer of permissive fibroblasts (0.1mL per well of a 96 well dish) or a 0.02 MOI inoculum of TB40/E as a positive control. Ten days later, each well was examined by light microscopy for cytopathic effect and was immunostained for IE2 using a primary monoclonal antibody clone 3H9 (kind gift, Tom Shenk) and a secondary goat anti-mouse IgG antibody fluorescently conjugated to AlexaFluor 488. Immunofluorescence staining methods are described in (4).

Clinical vs. in vitro viral reads diversity analysis. Clinical samples together with in vitro WT samples of HCMV-infected CD34+ HPCs (donor 1 data) were used for this analysis. Starting with the Tophat alignment with two mismatches allowed per uniquely aligned read, Picard (v 2.10.6, <http://broadinstitute.github.io/picard>) was used to mark and remove the duplicate reads, and subsequently to extract a random subsample of 11,000 reads per sample. A same subset of the first 10,000 reads per sample was used. These HCMV reads were blasted against the database of reference TB40/E sequence (blastn, e-value <1e-5) and the best blast hit was stored for every read. Percent identity was then reported for each sample using boxplot. These reads were also pooled for clustering analysis. We used CD-HIT-EST (v4.6.8) (11) to cluster reads with $\geq 95\%$ sequence identity over 90% of the shorter sequence. A list of cluster size per

sample was then generated, analogous to classifying reads into OTUs based on sequence similarity. These lists were analyzed using entropy.empirical function of R to generate the Shannon entropy estimates.

SNP calling. SNPs were identified using the Best Practices workflow on RNA-seq data of GATK (Genome Analysis Toolkit) (v3.7) (12). Briefly, for the Tophat alignment file, add read groups, sort, mark duplicates, and create index; SplitN'Trim and reassign mapping qualities; variant calling; variant filtering for clusters of at least 3 SNPs that are within a window of 35 bases. In addition, SAMtools mpileup (v1.3) (13) was applied to the variant site identification. Consensus calls from both tools were further filtered using depth of read (DP) more than 10 covered in all in vitro samples from NS and SS libraries, but they were not detected in the clinical samples using either of the two independent tools.

References

1. Umashankar M, et al. (2011) A novel human cytomegalovirus locus modulates cell type-specific outcomes of infection. *PLoS pathogens* 7(12):e1002444.
2. Umashankar M, Goodrum F (2014) Hematopoietic long-term culture (hLTC) for human cytomegalovirus latency and reactivation. *Methods Mol Biol* 1119:99–112.
3. Miller CL, Eaves CJ (2002) Long-term culture-initiating cell assays for human and murine cells. *Hematopoietic Stem Cell Protocols*, eds Klug CA, Jordan CT (Humana Press, Totowa), pp 123–141.
4. Petrucelli A, Rak M, Grainger L, Goodrum F (2009) Characterization of a Novel Golgi-localized Latency Determinant Encoded by Human Cytomegalovirus. *Journal of virology* 83(11):5615–5629.
5. Sinzger C, et al. (2008) Cloning and sequencing of a highly productive, endotheliotropic virus strain derived from human cytomegalovirus TB40/E. *The Journal of general virology* 89(Pt 2):359–368.
6. Umashankar M, et al. (2014) Antagonistic determinants controlling replicative and latent states of human cytomegalovirus infection. *Journal of virology* 88(11):5987–6002.
7. Buehler J, et al. (2016) Opposing Regulation of the EGF Receptor: A Molecular Switch Controlling Cytomegalovirus Latency and Replication. *PLoS pathogens* 12(5):e1005655.
8. Love MI, Huber W, Anders S (2014) Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol* 15(12):550.
9. Trapnell C, et al. (2010) Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nature biotechnology* 28(5):511–515.

10. Gordon CL, et al. (2017) Tissue reservoirs of antiviral T cell immunity in persistent human CMV infection. *The Journal of experimental medicine* 214(3):651–667.
11. Li W, Godzik A (2006) Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. *Bioinformatics* 22(13):1658–1659.
12. DePristo MA, et al. (2011) A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nat Genet* 43(5):491–498.
13. Li H, et al. (2009) The Sequence Alignment/Map format and SAMtools. *Bioinformatics* 25(16):2078–2079.
14. Van Damme E, Van Loock M (2014) Functional annotation of human cytomegalovirus gene products: an update. *Front Microbiol* 5:218.
15. Mocarski ES, Shenk T, Pass RF (2007) Cytomegaloviruses. *Fields Virology*, eds Knipe DM, Howley PM (Lippincott, Williams & Wilkins, Philadelphia), pp 2701–2673. 5 Ed.
16. Weekes MP, et al. (2014) Quantitative temporal viromics: an approach to investigate host-pathogen interaction. *Cell* 157(6):1460–1472.

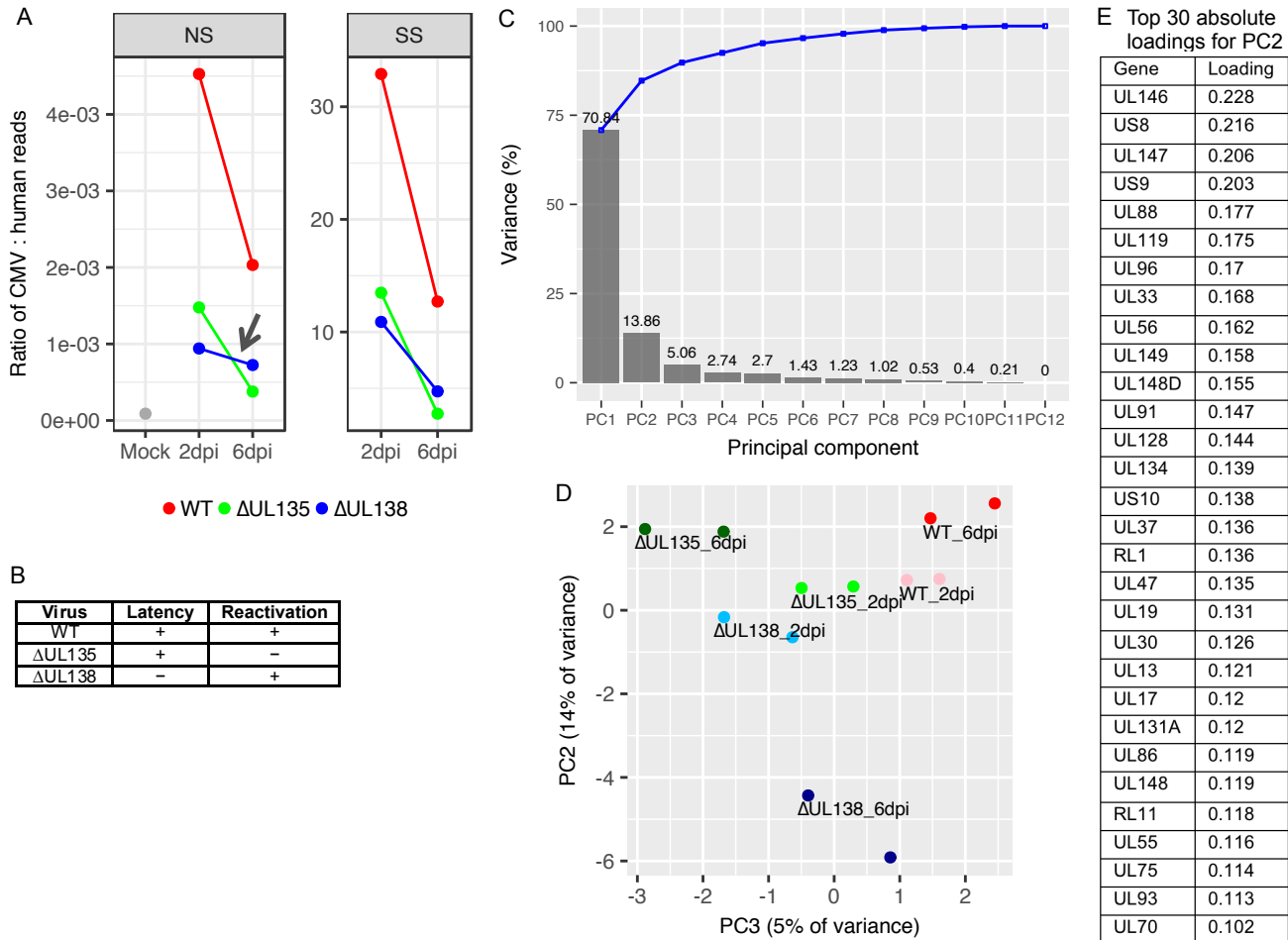


Fig. S1. Comparison of HCMV WT, Δ UL135 and Δ UL138 infected CD34⁺ HPCs. (A) The ratio of virus-to-human reads (V/H) in samples from NS and SS libraries. Mock-infected control is included in the left NS panel. The arrow indicates that two lines of two mutant viruses cross each other while two lines of Δ UL135 and WT are almost parallel over time of post infection. (B) Summary of infection patterns. (C) PCA scree plot for individual and cumulative (blue line) variance of PCs. (D) PC2 vs. PC3 score plot. (E) Given the separation between two mutant (Δ UL135 and Δ UL138) virus infections at 6dpi along the PC2 projection (Fig. 2A and D panel), 30 genes with the highest absolute loadings for PC2 are shown.

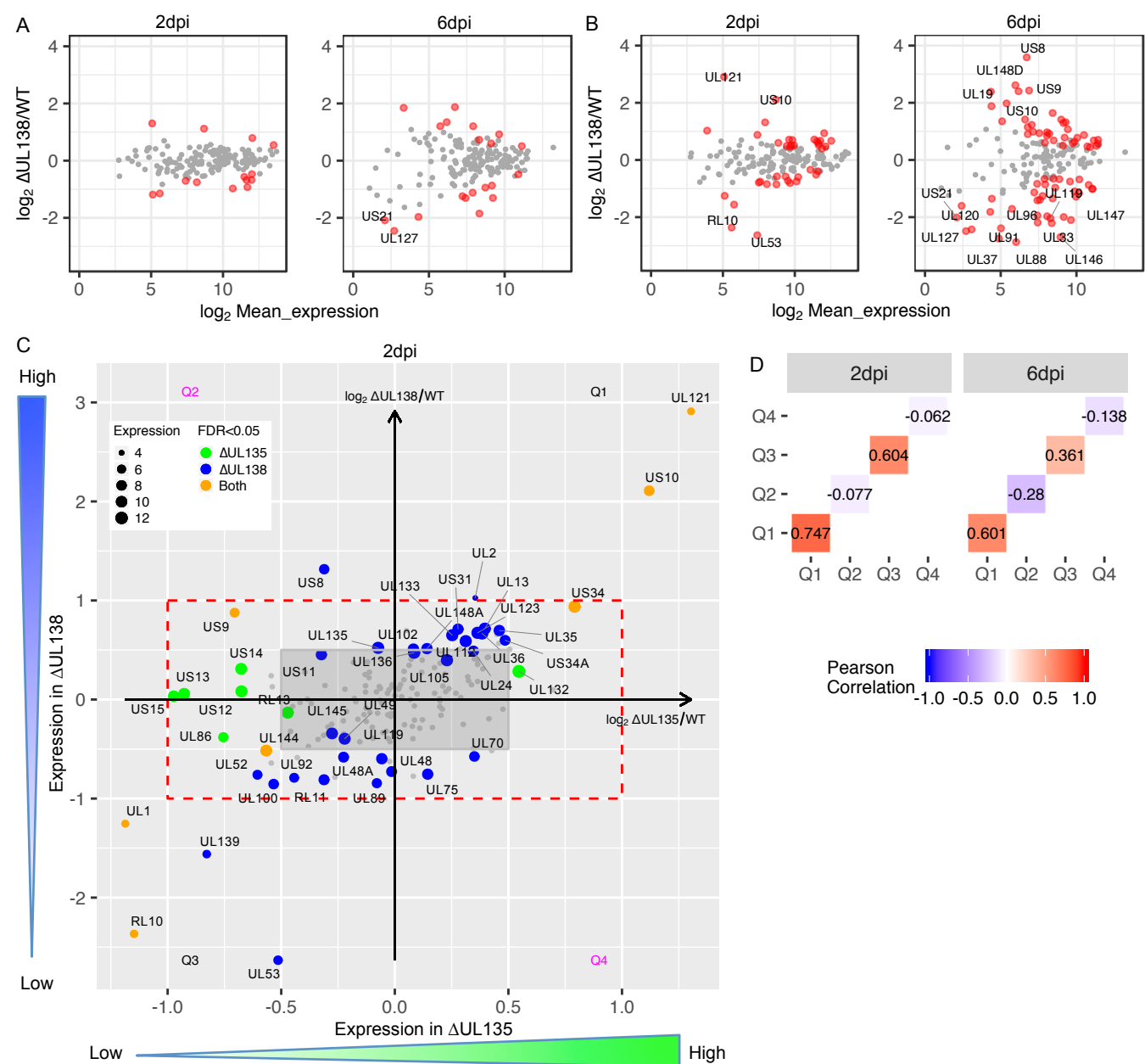


Fig. S2. Differential viral gene expression between Δ UL135 vs. WT and Δ UL138 vs. WT in HCMV-infected CD34⁺ HPCs. MA plot of Δ UL135 vs. WT (A) and Δ UL138 vs. WT (B) at 2 and 6dpi. Genes with more than fourfold change are indicated. (C) Two-dimensional differential expression combining A (x-axis) and B (y-axis) both at 2dpi. Red dashed rectangle highlights the twofold change. Green and blue dot size is proportional to the mean expression of individual genes in all Δ UL135 and Δ UL138 infections, respectively. Orange dot size is proportional to the mean expression of individual genes in all Δ UL135 and Δ UL138 infections. See Fig.2E for 6dpi. (D) Correlation or anti-correlation of \log_2 FC between the two-dimensional differential expression profiles at 2 and 6dpi using all genes falling into Q1, Q2, Q3 or Q4.

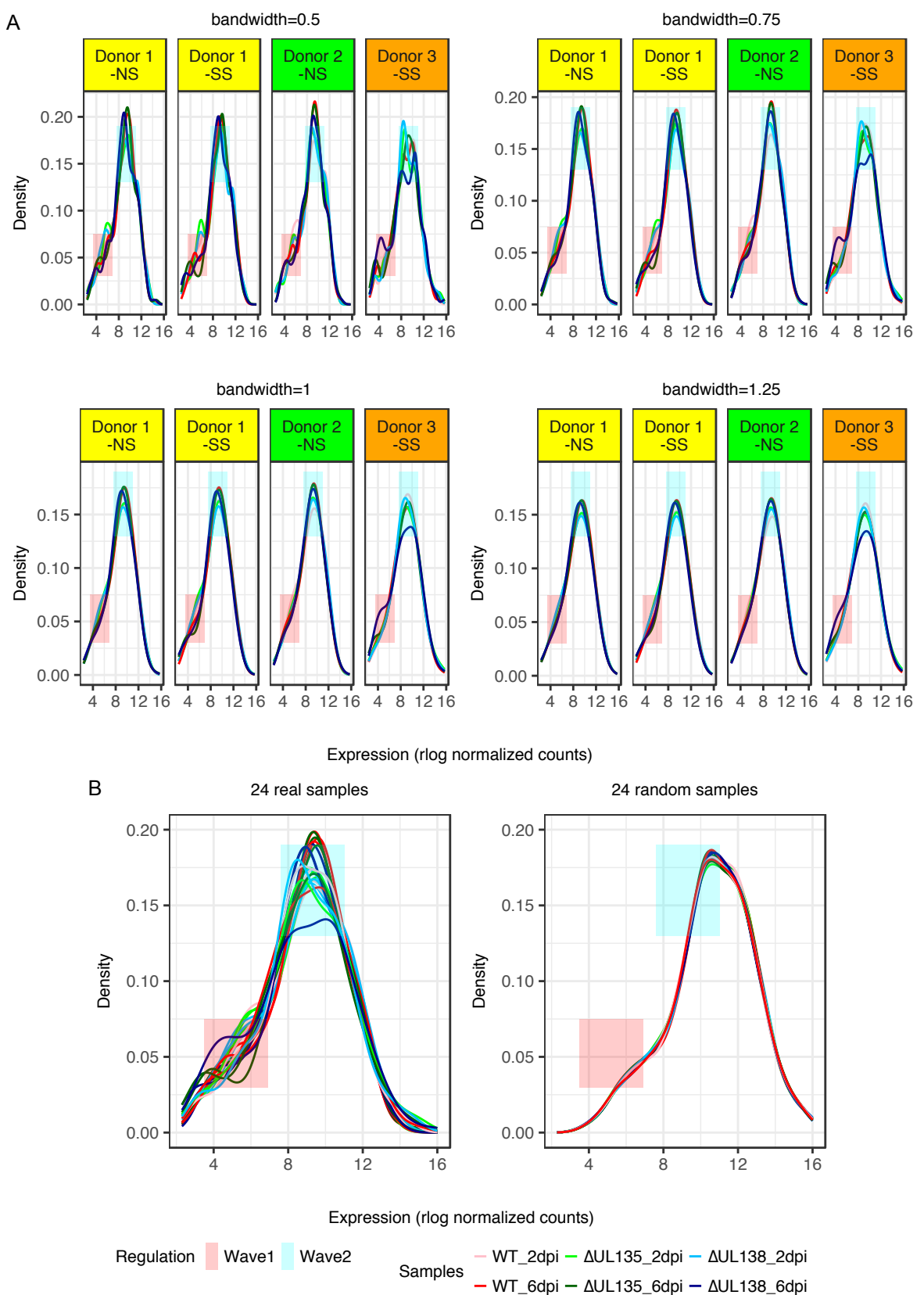


Fig. S3. Kernel density estimates of viral gene expression of HCMV-infected CD34⁺ HPCs across three cell donors. Two regions of low and moderate expression (termed wave 1 and wave 2) exhibit high variation. (A) Using four different bandwidth settings (0.5, 0.75, 1, 1.25). Every panel is composed of six biological samples. (B) Comparison between 24 real and random samples using optimal kernel density estimates.

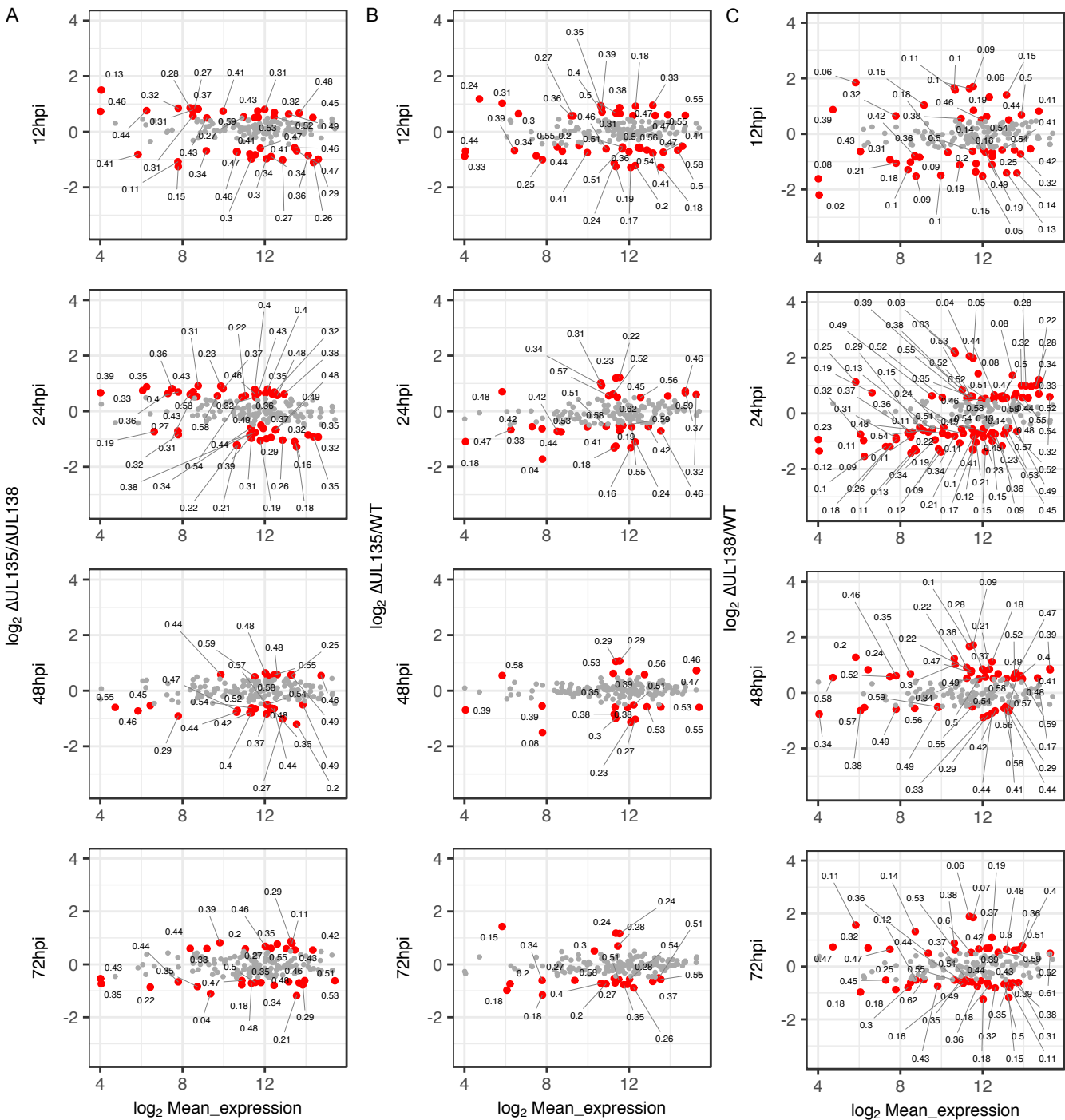


Fig. S6. MA plots of differential viral gene expression between Δ UL135 and Δ UL138 (A), between Δ UL135 and WT (B) and between Δ UL138 and WT (C) in HCMV-infected fibroblasts. Four time points, 12, 24, 48 and 72hpi, are shown for each comparison. Genes with absolute log₂FC > 0.5 are colored in red, and P values from DESeq2 results are labeled.

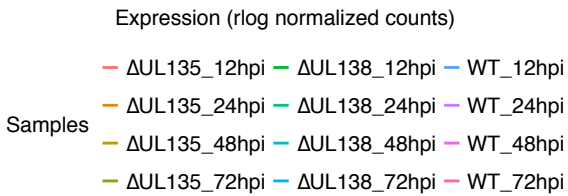
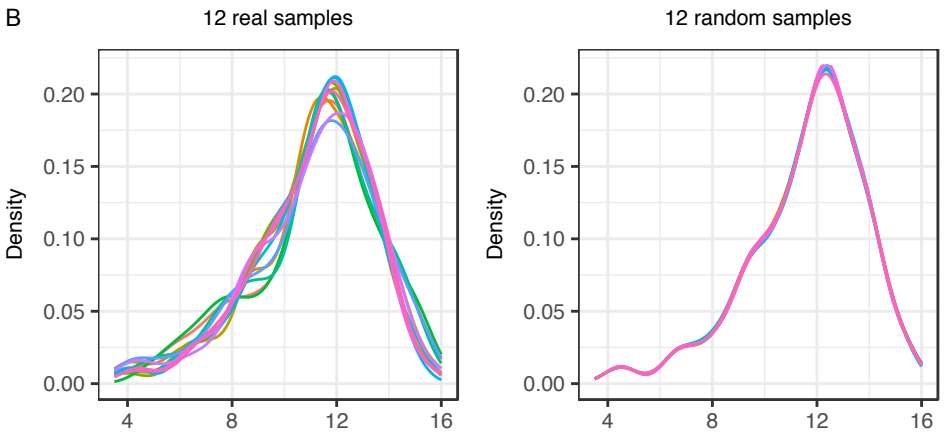
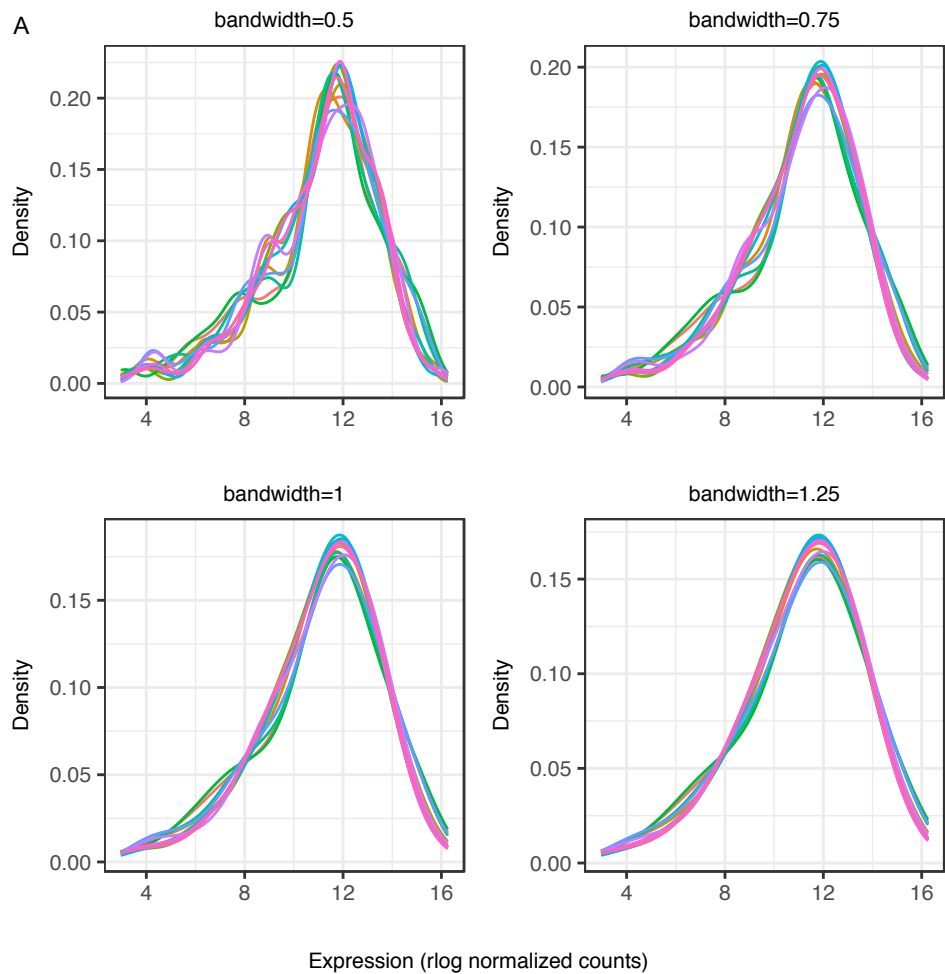


Fig. S7. Kernel density estimates of viral gene expression of HCMV-infected fibroblasts. (A) Using different bandwidth settings (0.5, 0.75, 1, 1.25). (B) Comparison between 12 real and random samples using optimal kernel density estimates.

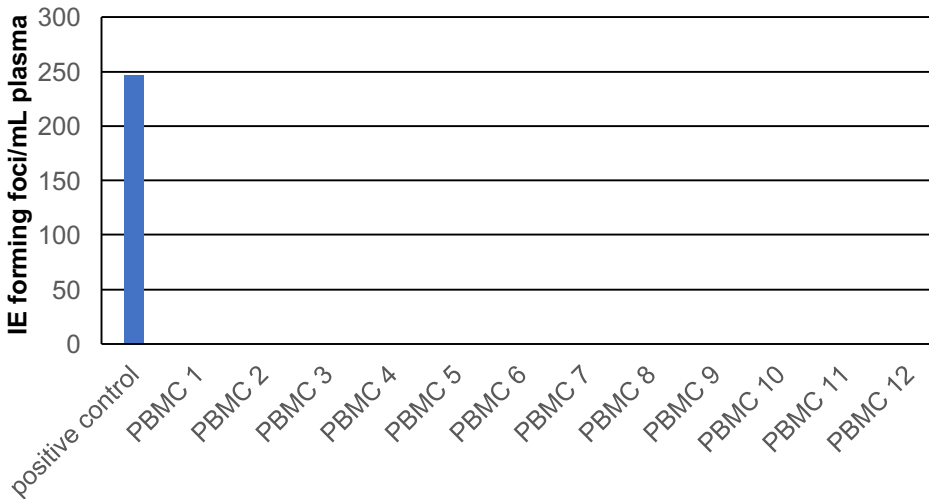
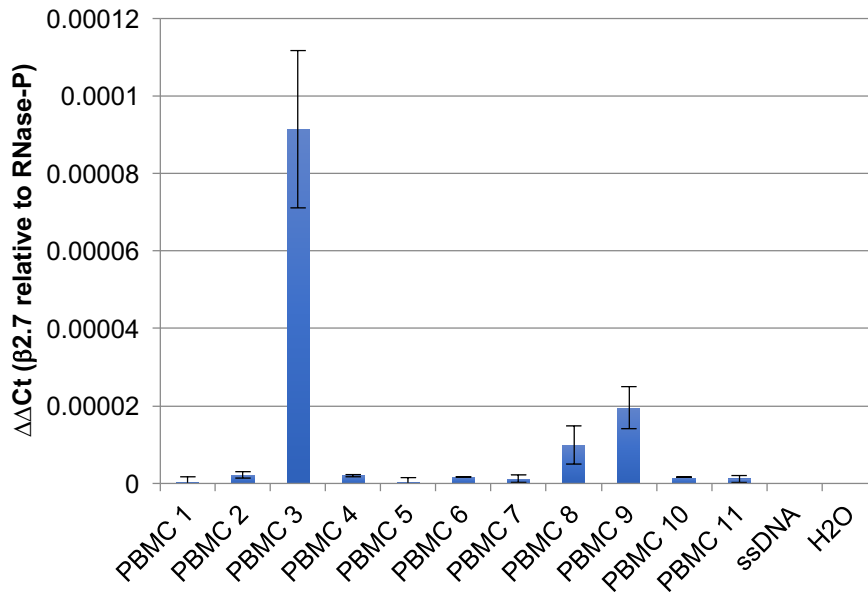
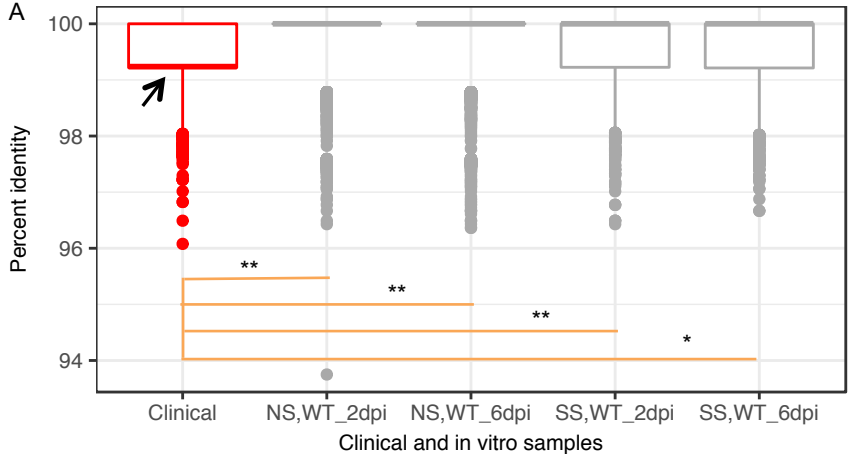
A**B**

Fig. S8. Analysis of viremia and genome load in clinical samples. (A) Fibroblast monolayers in 96 well dishes were incubated with 100 μ L of seropositive, healthy donor plasma or a low MOI inoculum (0.02) of TB40/E as a positive control for 10 days. Each well was examined for CPE and IE2 immunostaining. The number of IE-positive foci per mL of inoculum is shown. (B) cDNA derived from PBMCs derived from seropositive, healthy donors was analyzed for the presence of HCMV genomes using a primer recognizing the $\beta 2.7$ region of the genome relative to a host gene, RNaseP. Relative quantitation of $\beta 2.7$ is shown using $\Delta\Delta Ct$. Salmon sperm DNA (ssDNA) and water (H₂O) serve as negative controls. Standard deviation of five replicates is shown.



B

Sample	#Clusters	Shannon entropy
Clinical	4966	8.181
NS,WT_2dpi	3245	7.574
NS,WT_6dpi	2801	7.199
SS,WT_2dpi	4662	8.127
SS,WT_6dpi	4681	8.057

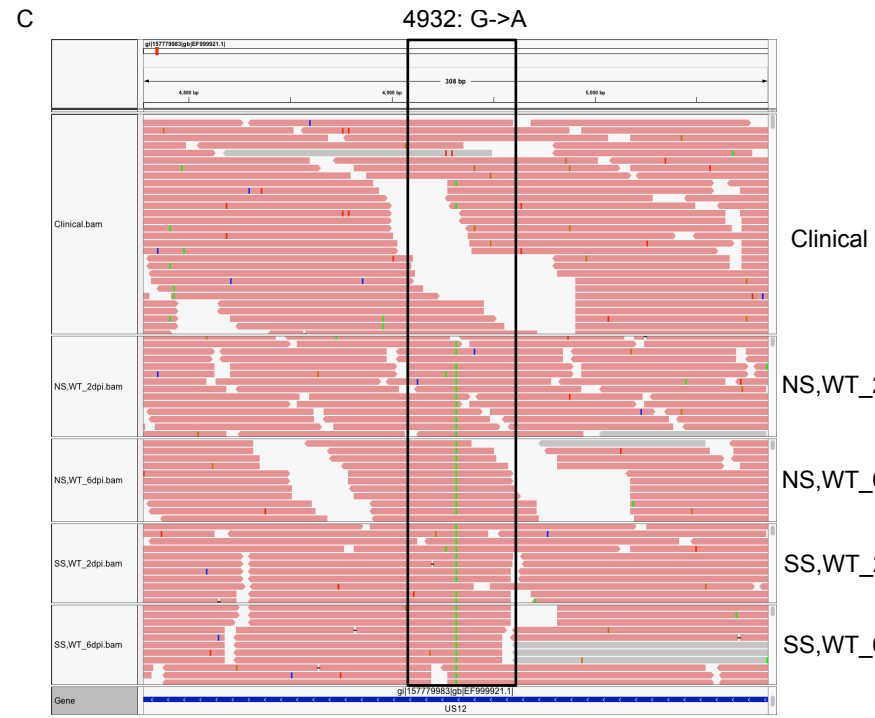


Fig. S9. Reads diversity comparison between clinical sample and WT HCMV-infected CD34+ HPCs at 2 and 6dpi (Donor 1 data). (A) 10,000 uniquely mapped HCMV reads were randomly generated from each sample. Boxplot of percent identity between those reads and HCMV reference sequence TB40/E. Median (thick line), first and third quartiles are shown. Whiskers extend to 1.5 times the interquartile range (IQR). The arrow indicates almost overlapped median and first quartile are below 100% in the clinical sample. * $P < 0.01$, ** $P < 0.001$; Wilcoxon rank sum test. (B) CD-hit clusters and Shannon entropy metrics using pooled reads in A. (C) A reliable SNP call using GATK and SAMtools. Alignment displaying a variant found in all *in vitro* samples in A but not detected in the clinical sample by either of the tools. DP indicates the number of filtered reads that support the reported SNP (the values are from GATK).

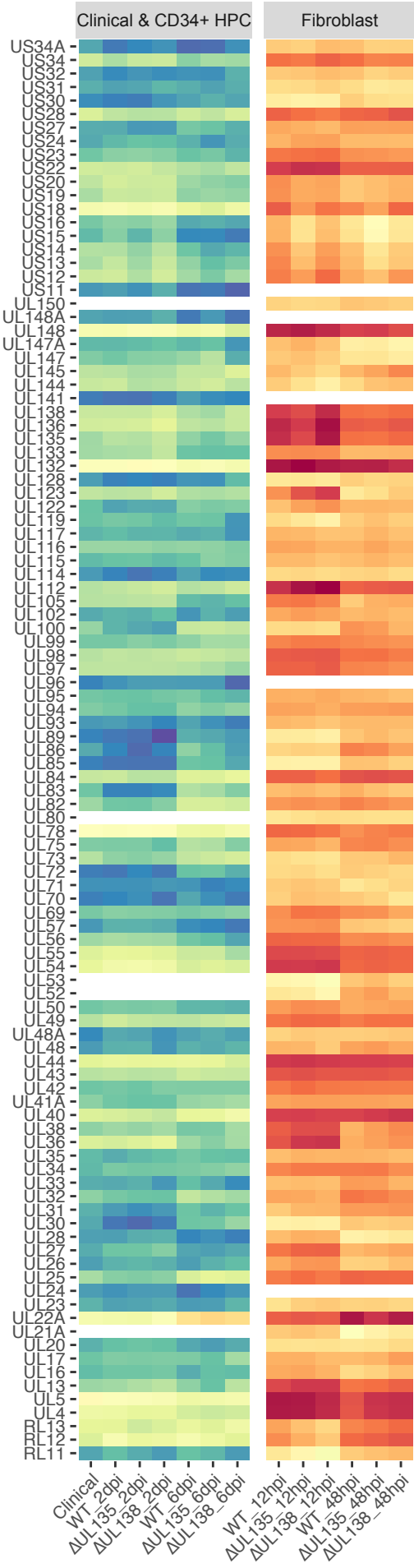


Fig. S10. Expression heatmap. 100 most highly expressed viral genes either in clinical latency and HCMV-infected CD34+ HPCs (NS, left) or fibroblasts (right) infected *in vitro* are shown.

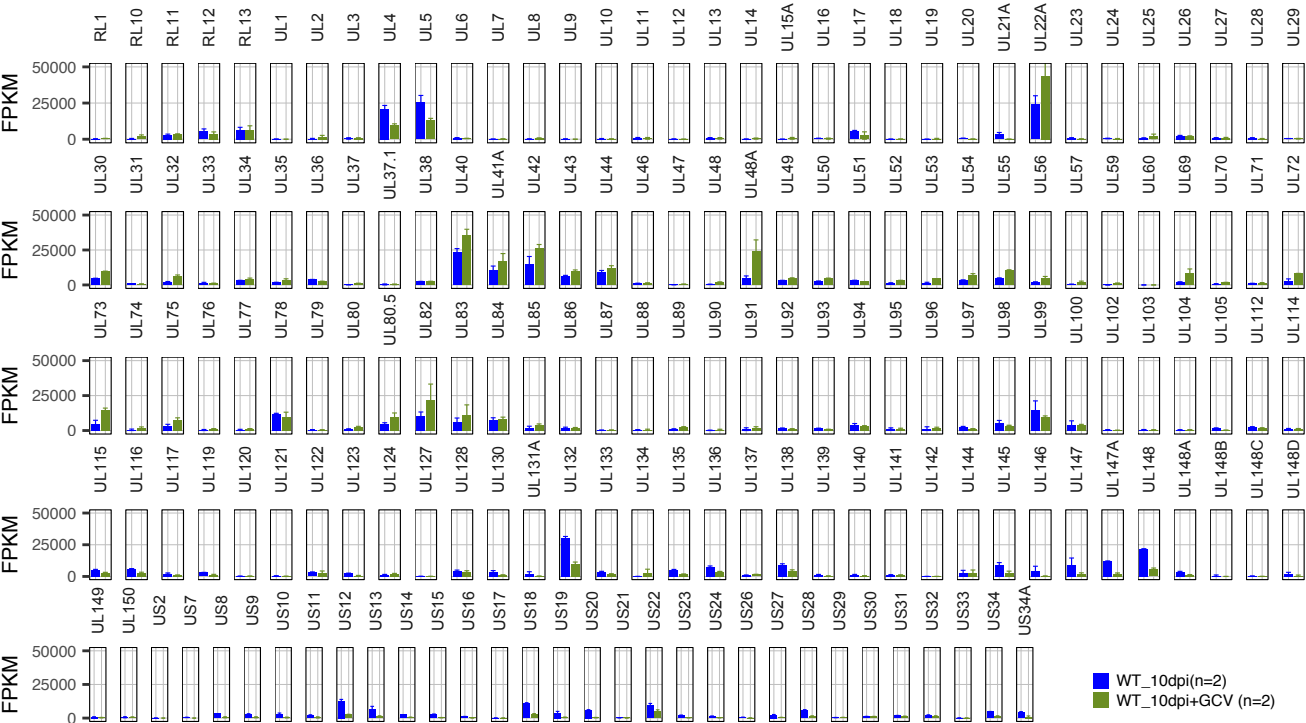


Fig. S11. HCMV gene abundance (FPKM) across genome at 10dpi and 10dpi+GCV. All error bars are SEM.

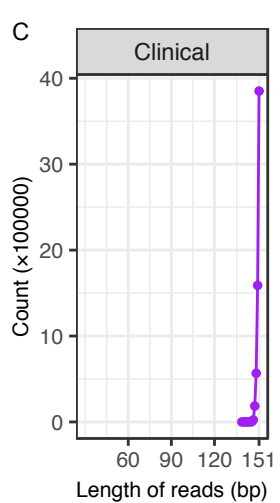
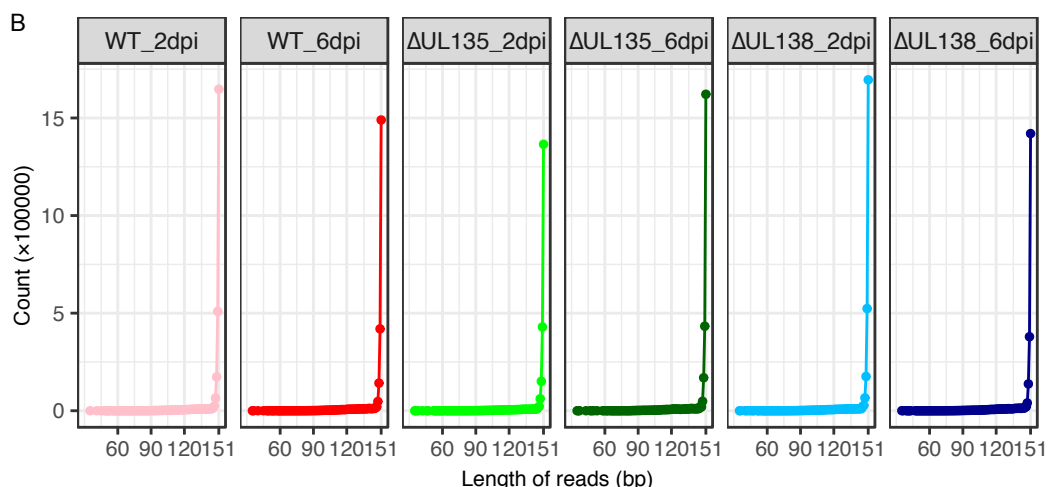
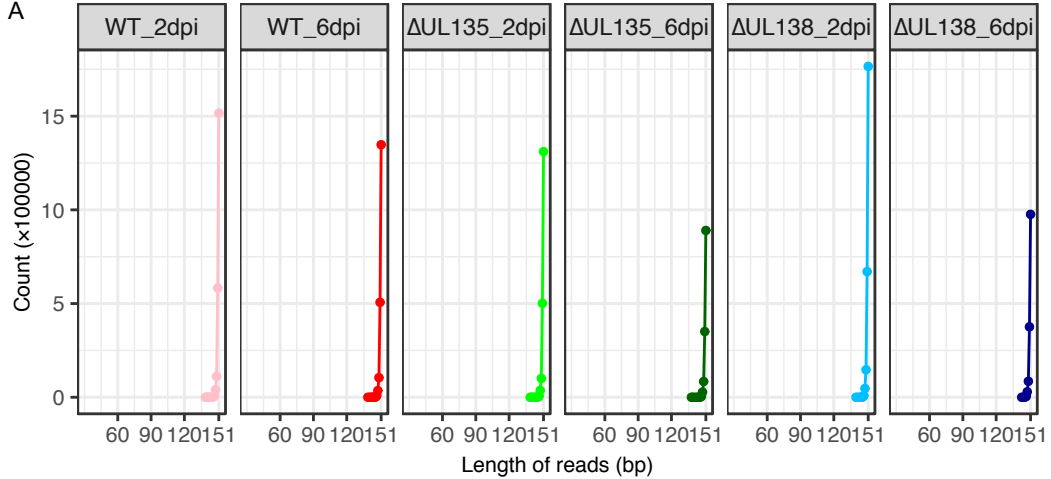


Fig. S12. Distribution of read length in SS enrichment samples of HCMV-infected CD34+ HPCs using Illumina MiSeq. (A) From Donor 1. (B) From Donor 3. (C) From three clinical samples.

Table S1. Summary of, concordant Q1/Q3 or antagonistic Q2/Q4, regulation of 8 low expressed genes associated with the switch between Δ UL135 (latency-like) and Δ UL138 (replication) in CD34+ HPCs at 6dpi

Type	Gene ^a	Log ₂ FC (5) ^b	Log ₂ FC (8) ^c	FDR (5) ^b	FDR (8) ^c	Mean ^d	Dsp. ^d	Function ^e	Process/protein/gene family ^e	mRNA class ^f
Q1	UL134	0.75	1.97	0.41	0.00	4.94	2.69	Unknown		
Q3	UL18	-1.96	-1.81	0.00	0.00	3.96	2.04	Immunomodulation	MHC-I homologue; Modulation NK cell signaling/function; Putative membrane glycoprotein; UL18 family	
	UL59	-0.47	-1.35	0.58	0.03	4.29	0.38	Unknown		
	UL120	-0.91	-2.42	0.39	7e-04	4.08	1.69	Unknown	Putative membrane glycoprotein; UL120 family	
Q2	UL19	-0.85	2.38	0.39	2e-04	4.09	1.43	Unknown		Tr5
	UL131A	-1.53	1.88	0.07	0.00	3.99	2.27	Cell tropism/Cell type-specific replication; Virion protein	Endothelial and epithelial cell tropism; Putative secreted protein	
Q4	UL37	0.61	-2.75	0.43	0	4.95	0.94	Unknown		
	UL91	0.53	-2.39	0.48	0	5.28	0.59	Gene expression/regulation	Essential for transcription of viral true late (γ 2) genes	

^a Herpesvirus-common genes

^b log₂FC(Δ UL135/WT) and FDR using the data from donor 1

^c log₂FC(Δ UL138/WT) and FDR using the data from donor 1

^d Mean expression and dispersion of Δ UL135_6dpi and Δ UL138_6dpi across three different donors

^e This information was based on references reviewing gene function (14, 15).

^f This information was based on temporal profiles of productive infection, defined in (16). Three mRNA classes, Tr1, Tr2-4 and Tr5, represent the mRNA expression that peaked at 0 to 24hpi, 24hpi to 72hpi and 72hpi, respectively.

Table S2. Summary of, concordant Q1/Q3 or antagonistic Q2/Q4, regulation of 22 moderately expressed genes associated with the switch between Δ UL135 (latency-like) and Δ UL138 (replication) in CD34+ HPCs at 6dpi

Type	Gene ^a	Log ₂ FC (5) ^b	Log ₂ FC (8) ^c	FDR (5) ^b	FDR (8) ^c	Mean ^d	Dsp. ^d	Function ^e	Process/protein/gene family ^e	mRNA class
Q1	UL16	0.36	0.68	0.41	0.03	8.7	0.54	Immunomodulation; Viral growth	Modulation NK cell signaling/function; Membrane glycoprotein; Temperance in fibroblasts	Tr2-4
	UL26	0.39	0.84	0.41	0.01	8.26	6e-02	Gene expression/regulation; Virion protein; Virion stability	Activator of MIEP; Tegument protein; US22 family	Tr2-4
	UL30	0.34	1.07	0.38	0	9.76	9e-02	Unknown	Expressed in latency	Tr5
	UL38	0.35	1.41	0.41	0	8.9	0.24	Apoptosis; Gene expression/regulation; Virion protein	Inhibitor of apoptosis; US3 regulator (repressor); Glycoprotein	Tr2-4
	UL102 ^a	1.20	0.60	1e-04	0.03	8.67	0.29	(DNA) Replication	Component of DNA helicase-primase	
	UL124	0.45	0.98	0.40	0.00	8.51	1.01	Latency	Membrane glycoprotein	Tr5
	UL128	0.23	1.64	0.57	0	8.96	0.22	Cell tropism/Cell type-specific replication; Immunomodulation	Involved in cell tropism endothelial cells; Putative secreted protein; Modulation chemo- and/or cytokines; Modulation of monocyte migration; Expressed in latency	
	US30	0.73	0.11	0.01	0.67	8.82	0.23	Viral growth	May be involved in temperance in fibroblasts; Putative membrane glycoprotein	Tr5
	US32	0.60	0.91	0.18	0.00	8.82	0.47	Unknown	US1 family	
	Q3	RL11	-0.11	-1.34	0.76	0	8.31	0.29	Immunomodulation	IgG Fc-binding capacity; Glycoprotein; RL11 family
UL33		-0.38	-2.20	0.41	0	8.59	0.44	Immunomodulation; Modulation of host cell cycle/protein synthesis; Virion protein	Modulation of chemo- and/or cytokine receptor through binding (CCR5/CXCR4); Envelope protein; GPCR-7TM family	Tr5
UL70 ^a		-0.29	-1.25	0.48	0	8.21	0.78	DNA Replication	DNA helicase-primase; Expressed in latency	
UL89 ^a		-0.14	-0.65	0.68	0.00	8.42	0.24	Viral growth	Formation of infectious particles	Tr5
UL93 ^a		-0.12	-1.39	0.80	0	8.19	0.3	Assembly/Maturation/Egress; Cellular trafficking; Virion protein	Possible role in DNA packaging; Tegument protein	
UL95 ^a		-0.04	-0.96	0.92	5e-	9.04	0.18	Assembly/	Required for accumulation	

					04			Maturation/ Egress; Gene expression/ regulation	late transcripts; expressed in latency	
	US13	-1.84	-0.63	1e- 04	0.14	7.8	0.16	Unknown	Putative multiple transmembrane protein; US12 family	
	US14	-1.12	-0.66	3e- 04	0.01	7.68	0.16	Unknown	Putative multiple transmembrane protein; US12 family	Tr2-4
Q2	UL23	-0.52	0.96	0.23	0.00	8.25	0.16	Viral growth; Virion protein	Temperance in fibroblasts; Tegument protein; US22 family	Tr5
	UL24	-0.03	0.93	0.97	0.03	7.76	0.11	Cell tropism/Cell type- specific replication; Virion protein	Involved in cell tropism in endothelial cells; Tegument protein; US22 family	
	UL135	-0.94	0.15	0.00	0.61	8.71	0.23	Latency/Re activation	Reactivation, membrane organization/maturation in endothelial; Putative secreted protein	Tr1
	US12	-1.30	0.30	0	0.23	9.17	0.18	Unknown	Putative multiple transmembrane protein; US12 family	Tr2-4
Q4	UL85 ^a	0.01	-0.67	0.97	0.00	8.93	0.28	Assembly/ Maturation/ Egress	Minor capsid protein	Tr5

^a Herpesvirus-common genes

^b log₂FC(ΔUL135/WT) and FDR using the data from donor 1

^c log₂FC(ΔUL138/WT) and FDR using the data from donor 1

^d Mean expression and dispersion of ΔUL135_6dpi and ΔUL138_6dpi across three different donors

^e This information was based on references reviewing gene function (14, 15)

^f This information was based on temporal profiles of productive infection, defined in (16). Three mRNA classes, Tr1, Tr2-4 and Tr5, represent the mRNA expression that peaked at 0 to 24hpi, 24hpi to 72hpi and 72hpi, respectively.