Functional single-cell genomics of human cytomegalovirus infection

Marco Y. Hein[1,2,3 ✉] and Jonathan S. Weissman [ID] [1,2,4 ✉]

Understanding how viral and host factors interact and how perturbations impact infection is the basis for designing antiviral interventions. Here we define the functional contribution of each viral and host factor involved in human cytomegalovirus infection in primary human fibroblasts through pooled CRISPR interference and nuclease screening. To determine how genetic perturbation of critical host and viral factors alters the timing, course and progression of infection, we applied Perturb-seq to record the transcriptomes of tens of thousands of CRISPR-modified single cells and found that, normally, most cells follow a stereotypical transcriptional trajectory. Perturbing critical host factors does not change the stereotypical transcriptional trajectory per se but can stall, delay or accelerate progression along the trajectory, allowing one to pinpoint the stage of infection at which host factors act. Conversely, perturbation of viral factors can create distinct, abortive trajectories. Our results reveal the roles of host and viral factors and provide a roadmap for the dissection of host–pathogen interactions.

The β-herpesvirus human cytomegalovirus (HCMV) is a pervasive pathogen that establishes lifelong infection in the majority of the human population. Activation of its lytic cycle triggers a characteristic cascade of events, starting with stereotypical waves of viral gene expression, continuing with the replication of its ~235-kb double-strand DNA genome and culminating in the budding of viral progeny[1]. Several systems-level studies have described these phenomena at the level of the transcriptome, the set of translated messages and the proteome in time and space[2–7], typically studying infection in fibroblasts. HCMV has a wide tropism in its human host, but fibroblasts are predominant in the pool of infected cells in vivo and have been broadly used for propagation in cell culture[8]. A core set of viral genes essential for replication was established by systematic mutagenesis[9,10]. These studies highlighted the complexity of hundreds of viral genes cooperating to manipulate the host and undermine its defense machinery, raising the question of what the best targets are for antiviral intervention.

CRISPR–Cas9 technology provides tools to systematically measure the functional contribution of each viral gene and host factor involved in productive infections[11]. However, it remains a challenge to translate a list of factors into a systematic understanding of their individual roles and organization in pathways. We address this challenge with Perturb-seq, a scalable approach that combines CRISPR-based genetic perturbations with rich phenotypic profiling by single-cell transcriptomics. First, we conducted systematic pooled CRISPR screens for both host and viral factors affecting survival of primary human fibroblasts upon HCMV infection. Survival screens are scalable to genome-wide libraries but capture only a compressed picture of the molecular events unfolding upon perturbation of host or viral factors. Moreover, lytic infection is inherently dynamic over time and heterogeneous from cell to cell[12,13]. We, therefore, recorded the transcriptomes of tens of thousands of single cells and monitored how perturbation of a set of critical factors—identified in the pooled screens—alters the timing, course and progression of infection.

Our data paint a high-resolution picture of the HCMV lytic cycle as a deterministic program that is distinctly vulnerable to host- and virus-directed interventions. We identify host restriction factors as well as dependency factors critical for viral entry and for progression from early to late stages of infection. Conversely, we show that targeting key viral factors derails the viral gene expression program in specific ways. Taken together, our findings reveal a dichotomy between the roles of host and viral factors, with the set of viral factors solely defining the trajectory of infection and host factors creating the environment permitting the execution of that program.

Results

High-resolution functional scanning of the HCMV genome. Cas9 is an effective tool for making targeted disruptions in the cytomegalovirus genome. Targeting individual essential herpesvirus genes by CRISPR–Cas9 was shown to disrupt their expression directly and their function through errors introduced by the host DNA repair machinery[14]. Cleavage of the viral DNA in non-essential regions has minimal effect on HCMV replication and host cell viability—likely because DNA repair is fast relative to replication—but can affect expression of genes proximal to the cut sites (see ref. [14] and our data below). To enable high-resolution scanning of viral elements for a comprehensive functional annotation of the HCMV genome, we designed a CRISPR library that targets every protospacer-adjacent motif (PAM) for *Streptococcus pyogenes* Cas9 (NGG PAM sequence present roughly every 8 bp) along the genome of the clinical strain Merlin (Fig. 1a and Supplementary Table 1). We delivered the library into primary human fibroblasts engineered to express Cas9, so that, upon infection, each cell executes a cut at a defined position along the viral genome, collectively tiling its entirety.

We mapped the functional landscape of the HCMV genome by quantifying the abundance of individual single guide RNA (sgRNA) cassettes in a population before/after infection (Extended Data Fig. 1). We found that cutting phenotypes are relatively constant within individual genes, indicating that the determining factor is which gene is targeted, rather than where Cas9 cuts within the gene

[1]Department of Cellular and Molecular Pharmacology, University of California, San Francisco, San Francisco, CA, USA. [2]Howard Hughes Medical Institute, University of California, San Francisco, San Francisco, CA, USA. [3]Chan Zuckerberg Biohub, San Francisco, CA, USA. [4]Whitehead Institute for Biomedical Research, Massachusetts Institute of Technology, Cambridge, MA, USA. ✉e-mail: marco.hein@czbiohub.org; weissman@wi.mit.edu
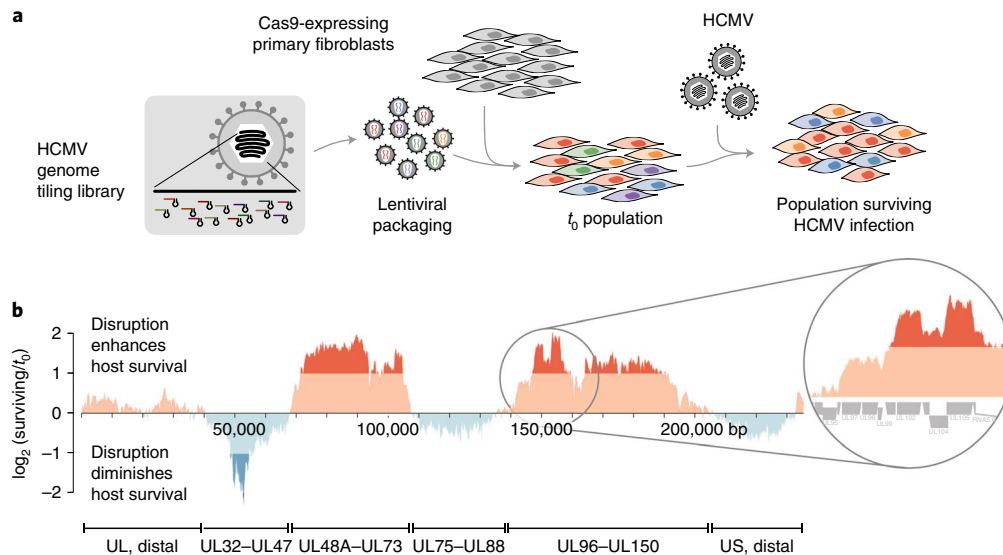
Fig. 1 | Virus-directed CRISPR nuclease screen maps the phenotypic landscape of the HCMV genome. a, Experimental design for pooled, virus-directed CRISPR screening. Our HCMV tiling library contains ~33,000 elements and was lentivirally delivered into primary human foreskin fibroblasts expressing the CRISPRn machinery, followed by infection with HCMV. sgRNA cassettes were quantified by deep sequencing in the initial (t_0) population and the surviving population. **b**, Phenotypic landscape of the HCMV genome obtained by locally averaging the phenotypes of individual sgRNAs along the genome and across two biological replicates. Strong changes in the magnitude of the phenotype coincide with gene–gene boundaries (inset).

body. Cutting adjacent sets of genes frequently showed similar phenotypes. However, some gene boundaries were marked by abrupt phenotype changes, arguing that direct consequences of Cas9 cuts are limited to their immediate vicinity (Fig. 1b, Extended Data Fig. 1 and Supplementary Table 1).

At a larger scale, changes in the direction and magnitude of the phenotypes defined six major genomic modules. Cuts in both distal regions of the genome, which lack genes essential for viral replication[9,10], had minimal effect on host cell survival. As expected, targeting the regions covering UL48A–UL73 and UL96–UL150, both of which contain essential genes involved in viral DNA replication, packaging and nuclear egress[10,15], strongly protected infected cells. In the two remaining regions, we found that disruption of genes required for viral replication did not necessarily protect the host. Cuts within the UL32–UL47 region, which contains essential genes, actually led to a strongly increased ability of the virus to kill cells. The most strongly sensitizing phenotypes mapped to the known viral apoptosis inhibitors UL36, UL37 and UL38 (ref. [16]). Although this behavior can be rationalized for virally encoded anti-apoptotic proteins, it extended to many other virus-essential genes without known anti-apoptotic roles, including the DNA polymerase processivity factor UL44. Finally, and counterintuitively, cuts in the central region spanning UL75–UL88 caused very mild phenotypes. Many genes in this region encode essential components of the viral envelope, tegument and capsid, yet the outcomes on host survival after Cas9 cutting were similar to targeting the US distal region, resulting in even mildly enhanced cell death upon infection.

Targeting essential viral genes, by definition, undermines the production of viral offspring. It does not necessarily concomitantly protect the infected cell, underscoring that successful viral replication and death of the infected cell are distinct phenomena. It appears that disrupting essential genes involved in viral DNA replication mostly protects the host. However, interfering with the later steps of assembling new virions might not only be ineffective in protecting the host but might even place an additional burden.

Genome-wide screens for host factors of HCMV. Next, we carried out a pooled screen for host factors of HCMV infection by systematically repressing host genes by CRISPR interference (CRISPRi)[17,18].

Phenotypes were defined by enrichment or depletion of sgRNA cassettes before/after infection and a mock-infected control population (Fig. 2a and Supplementary Table 2).

Our screen revealed a diverse set of host factors (Fig. 2b). Genes involved in the biosynthesis of heparan sulfate were among the strongest protective hits. Heparan sulfate proteoglycans on the cell surface enable viral cell adhesion[19,20]. Additionally, we found vesicle trafficking factors: *RAB6A* and its GEFs *RIC1/KIAA1432* and *RGP1*, the conserved oligomeric Golgi (COG) complex, members of TRAPP complex III and *UNC50*. These factors converge on the Golgi apparatus and mediate retrograde and anterograde transport. Some of them (COG, TRAPP and UNC50) were previously implicated in the internalization of diverse bacterial and plant toxins, suggesting that HCMV and toxins exploit similar pathways for cell entry[17,21–25].

Other protective hits included members of the LAMTOR/Ragulator complex, folliculin (*FLCN*) and the lyspersin (*C17orf59*) subunit of the BORC complex, all linked to lysosome positioning and nutrient sensing[26–28]. This supports the recent observation that HCMV infection changes lysosome dynamics[7]. Additionally, host cell death was reduced by knockdown of cullin-RING E3 ligases (*RBX1* and *CUL3*), their adaptor subunits (*DDA1* and *TCEB2/ELOB*), substrate receptors (*DCAF4*) and the associated neddylation (*NEDD8*, *NAE1* and *RBX1*) and deneddylation (COP9 signalosome) machineries. Many viruses hijack this pathway to degrade host restriction factors, which can be prevented by broadly acting Nedd8-activating enzyme inhibitors[29,30]. Finally, we identified genes involved in tail-anchored protein insertion into the endoplasmic reticulum (ER) as well as ER-associated degradation: *AMFR*, an E3 ligase, and the TRC40/GET pathway members *BAG6* and *ASNA1*, which were shown to be required for insertion of membrane proteins of herpes simplex virus (HSV-1)[31].

Our screens also identified genes whose knockdown sensitizes cells to death upon infection—for example, known restriction factors such as *PML* and *DAXX* and members of the interferon (IFN) type I pathway. Although restriction factors are strictly defined by their effect on preventing viral replication, our findings suggest that screening for genes that sensitize the host to cell death
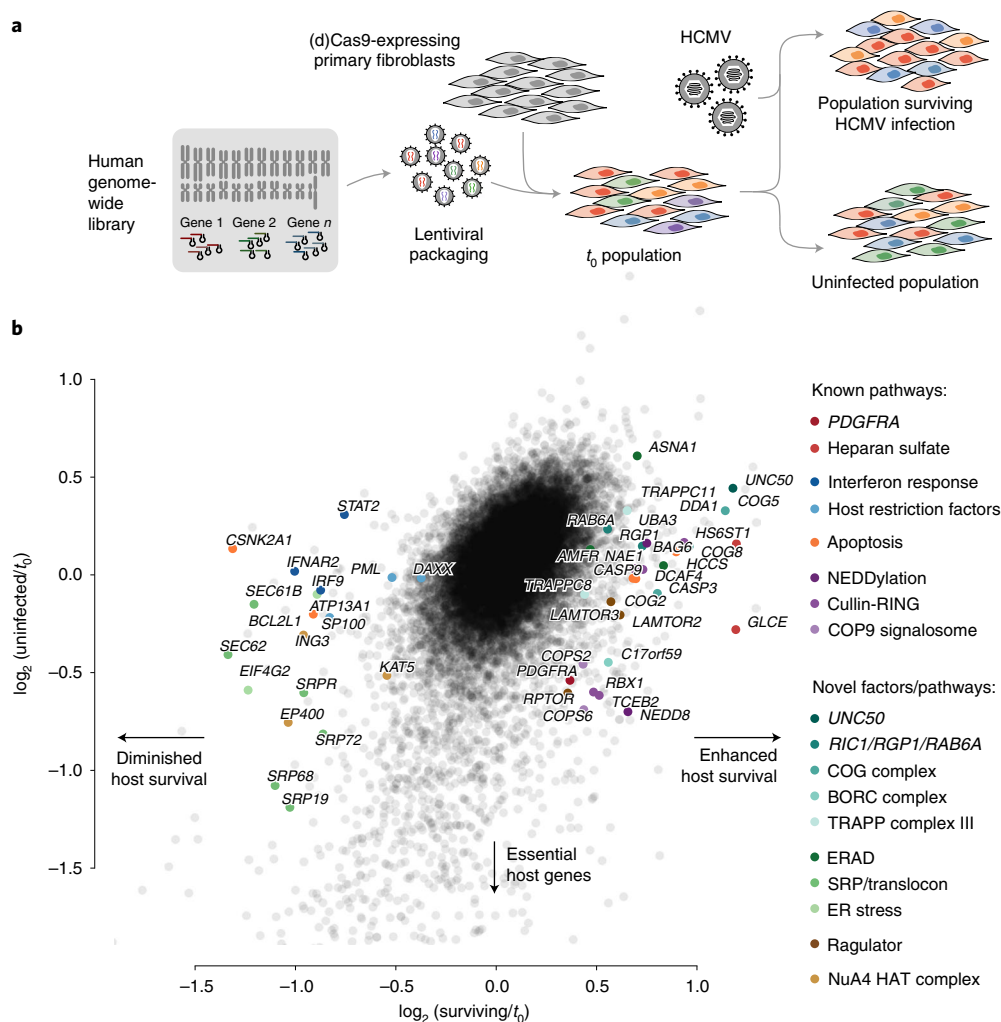
Fig. 2 | Host-directed CRISPR screens identify host dependency and restriction factors. a, Experimental design for pooled, host-directed CRISPRi and CRISPRn screening. Genome-wide sgRNA libraries targeting human genes with multiple sgRNAs each were lentivirally delivered into primary human foreskin fibroblasts expressing the CRISPRi or CRISPRn machineries, followed by infection with HCMV. sgRNA cassettes were quantified by deep sequencing in the initial (t_0) population, the surviving population and an uninfected control population to account for gene essentiality in the absence of infection. **b**, Results of the host-directed CRISPRi screens displayed as a scatter plot of average gene essentiality (that is, infection-independent phenotype; y axis) versus protection/sensitization to death upon HCMV infection (that is, infection-dependent phenotype; x axis) averaged across two biological replicates. Highlighted are hit genes representing selected pathways. HAT, histone acetyltransferase.

when knocked down is a viable approach for their identification. Sensitizing hits included subunits of the NuA4 histone acetyltransferase complex, which was shown to counteract hepatitis B virus replication by repressing transcription from chromatinized viral DNA[32] and was described as an interactor of the HIV-1 TAT protein[33]. Furthermore, we identified members of the signal recognition particle, the translocon and associated factors as well as genes involved in ER stress[34,35]. Finally, we found genes with anti-apoptotic function, including several caspases, whose knockdown likely increases the sensitivity to apoptosis triggered by HCMV infection.

A recent study reported *PDGFRA* as the dominant hit in a CRISPR nuclease (CRISPRn) knockout screen designed to identify host factors required for HCMV entry[36], underscoring its reported role as the receptor on the surface of fibroblasts exploited by HCMV strains expressing the trimeric virion glycoprotein complex[37–39]. In our CRISPRi screen, *PDGFRA* knockdown conferred mild protection from cell death upon infection (Fig. 2b).

To validate and extend the host factors identified by CRISPRi, we conducted a knockout screen under the same conditions, using

an established CRISPR cutting library[40] (Extended Data Fig. 2a). CRISPRi and CRISPRn screening identified similar numbers of protective hits, of which around 20% scored as significant in both screening modes (Supplementary Table 2). We found good agreement at the pathway level even though gene-level phenotypes varied (Extended Data Fig. 2b). Of note, protective hits in the knockout screen were dominated by genes involved in virus entry, such as *PDGFRA* and heparan sulfate biosynthesis genes. This underscores that genes with very strong protective phenotypes are more readily identified in knockout screens, likely because selection pressure can act more strongly on cells with true null alleles. Overall, our findings emphasize the benefits of combining orthogonal modes of genetic screening[41].

The lytic cascade resolved by single-cell transcriptomics. Our pooled screens provide a genome-scale picture of the factors involved in lytic HCMV infection, but placing them into biological pathways and linking them to a stage of the viral life cycle requires prior knowledge or dedicated follow-up experiments. To investigate

the roles of critical host and viral factors systematically in more depth, we used Perturb-seq, which combines CRISPR-based genetic perturbations with a rich single-cell transcriptional readout[34,42–44]. Measuring tens of thousands of single-cell transcriptomes from a population with a library of genetic perturbations provides a massively parallel way of assessing the outcome of those perturbations under uniform conditions. The single-cell nature of this approach makes it particularly well suited for studying viral infection, a process with great inherent variability from cell to cell[12,13,45–49].

As the groundwork for the Perturb-seq analysis, we first explored the progression of HCMV infection by recording single-cell transcriptomes of 12,919 genetically unperturbed cells, sampled from eight time points with two multiplicities of infection (MOIs) each (Fig. 3a,b). Instead of relying primarily on synchronizing cells experimentally, which has inherent limits due to the intrinsic heterogeneity in the timing of initial infection and its progression, we staged cells computationally by their transcriptional signatures. The largest sources of variability between cells were the extent of IFN signaling and the fraction of viral RNA per cell ('viral load'), which reached levels of around 75% (Fig. 3c,d and Supplementary Table 3). Cells with high viral loads showed a marked increase in their total observed mRNA molecules (that is, unique molecular identifier (UMI) counts) per cell. This indicates that the eponymous increase in cell size during infection (cyto megalo—large cell) is reflected in a higher cellular RNA content (Extended Data Fig. 3a,b). Together, these properties define three main subpopulations of cells: a naive population (uninfected, IFN-negative); a bystander population (not expressing viral genes, IFN-positive); and an internally heterogeneous infected population with varying amounts of viral transcripts, which we divided into multiple subclusters (Fig. 3e and Extended Data Fig. 3b). The distribution of cell cycle phases was altered between clusters (Extended Data Fig. 3c), in line with the well-established G1 arrest early during infection[50] and a subsequent shift toward a state resembling S-phase in cells with very high viral loads.

Each cluster contained cells from both the low- and high-MOI samples, and the gene expression patterns between those groups of cells were extremely highly correlated (Extended Data Fig. 3d). This highlights both the excellent technical reproducibility of our transcriptomics workflow and that the MOI determines the population-level response (that is, the fraction of cells at a given stage of infection) rather than the nature of the transcriptional responses in individual cells.

Interestingly, viral gene expression and the expression of interferon-stimulated genes (ISGs) were almost entirely mutually exclusive, a phenomenon that has been observed for HSV-1 (ref. [46]). Cells with high viral loads showed entirely suppressed IFN signaling. Only cells in cluster 'infected 1' (Extended Data Fig. 3b) expressed both ISGs and low amounts of viral RNA, mainly classic immediate-early genes such as UL123 (IE1) (Fig. 3g), indicating that these are cells in the earliest stage of infection. Together, this underscores the rapidity with which the virus effectively suppresses IFN signaling and highlights the power of single-cell analyses in resolving this phenomenon, which might appear as concurrent expression of viral genes with ISGs in bulk measurements.

We next tracked how the cell population is distributed to the different clusters as a function of time and MOI (Fig. 3f and Supplementary Table 3). All cells are initially in the naive cluster. At 6 hours post infection (h.p.i.), cells begin to transition to either the bystander or infected clusters, and, at 20 h.p.i., almost no naive cells remain. The ratio of infected to bystander cells depends on the initial MOI and stays relatively constant between 20 and 48 h.p.i., with infected cells progressing to clusters with higher viral load. After 48 h.p.i., we detected another marked increase in the number of infected cells and a concomitant decrease in bystander cells, corresponding to a second wave of infection, likely caused by virions

released from cells that were infected early. By 96 h.p.i., even the population with low initial MOI is almost completely infected.

Among the infected clusters, most cells follow a dominant trajectory with increasing viral load (clusters 'infected 1–6'), and viral marker genes for specific phases peak in defined clusters (Fig. 3g and Supplementary Table 3). To define the viral gene expression patterns along this trajectory at high resolution, we grouped cells from those clusters into narrow viral load bins and determined the profiles of all robustly quantified viral genes (Extended Data Fig. 4a). Many genes display stereotypic kinetics, such as immediate-early (UL123, US10 and US2) or true-late (UL99 and UL100, among others) genes. However, our high-resolution approach revealed that the (pseudo)temporal patterns of many genes were subtly distinct from one another, and many did not align with the canonical temporal patterns. For instance, US6, UL78, US26, UL42 and US34 showed bimodal kinetics, and expression of UL4 and UL48A peaked mid-infection.

A small but prominent set of host transcripts were upregulated with increasing viral load, resembling 'leaky-late' or 'true-late' viral gene kinetics (Fig. 3d): *APOE*, *CD55*, *TSPAN13*, *HES6*, *SPINT2* and *PCSK5*. Upregulation of CD55 has also been observed at the protein level[5], and it was shown to be incorporated into budding virions to counteract the complement system[51].

A small subpopulation, ~1% of infected cells, did not follow the dominant trajectory but diverted from infected cluster 2, following an off-trajectory where cells reach high viral loads in a distinct region of gene expression space (Fig. 3c). Cells in that cluster were defined by lower UMI counts, suggesting no increase in cell size (Extended Data Fig. 3b). Their pattern of viral gene expression was markedly different (Extended Data Fig. 4b and Supplementary Table 3). Immediate-early and delayed-early genes were strongly overexpressed, whereas true-late genes and all long noncoding RNAs were strongly depleted. As true-late gene expression depends on genome replication[52], we conclude that this trajectory is abortive. Looking at host transcripts, cells on the abortive trajectory were characterized by lack of suppression of the IFN response and were the only ones expressing interferon-β (*IFNB1*) itself, along with other cytokines and many stress response genes, prominently from the NF-κB pathway (*NFKBIA* and *RELB*) as well as *JUN*. This pattern was reminiscent of the finding that HSV-1 initiates pathways upstream of IFN production in a rare subset of abortively infected cells[46]. Together, our findings suggest that, similarly, a small subpopulation of cells on an abortive viral trajectory actively secrete interferon-β while escaping suppression of the downstream response to interferon, possibly involving autocrine feedback loops in addition to paracrine signaling.

Host perturbations can block, slow or accelerate infection. We next conducted a series of Perturb-seq experiments exploring the effect of targeting host and viral factors on the viral life cycle. In contrast to the pooled screen, where phenotypes emerge by enrichment or depletion of cells over multiple days, Perturb-seq provides a high-resolution view of the effect of targeting a critical gene over the first 72 h.p.i., covering roughly one viral replication cycle.

We first selected 52 host genes identified as hits in the pooled screens, cloned them into a targeted library, along with non-targeting control sgRNAs, and delivered the library into a population of fibroblasts expressing the CRISPRi machinery (Fig. 4a,b and Supplementary Table 4). We challenged that population with an MOI of HCMV of 0.5 for 1 h, followed by a medium exchange, and monitored the effects of the genetic perturbations in an average of 165 ± 50 cells per target per time point (Extended Data Fig. 5a). CRISPRi reduced expression of the host targets by a median of 87% (Extended Data Fig. 5b) and triggered target-specific transcriptional responses (Extended Data Fig. 5c). In uninfected cells, we observed the strongest transcriptional responses after knockdown of IFN pathway members, LAMTOR/Ragulator subunits and
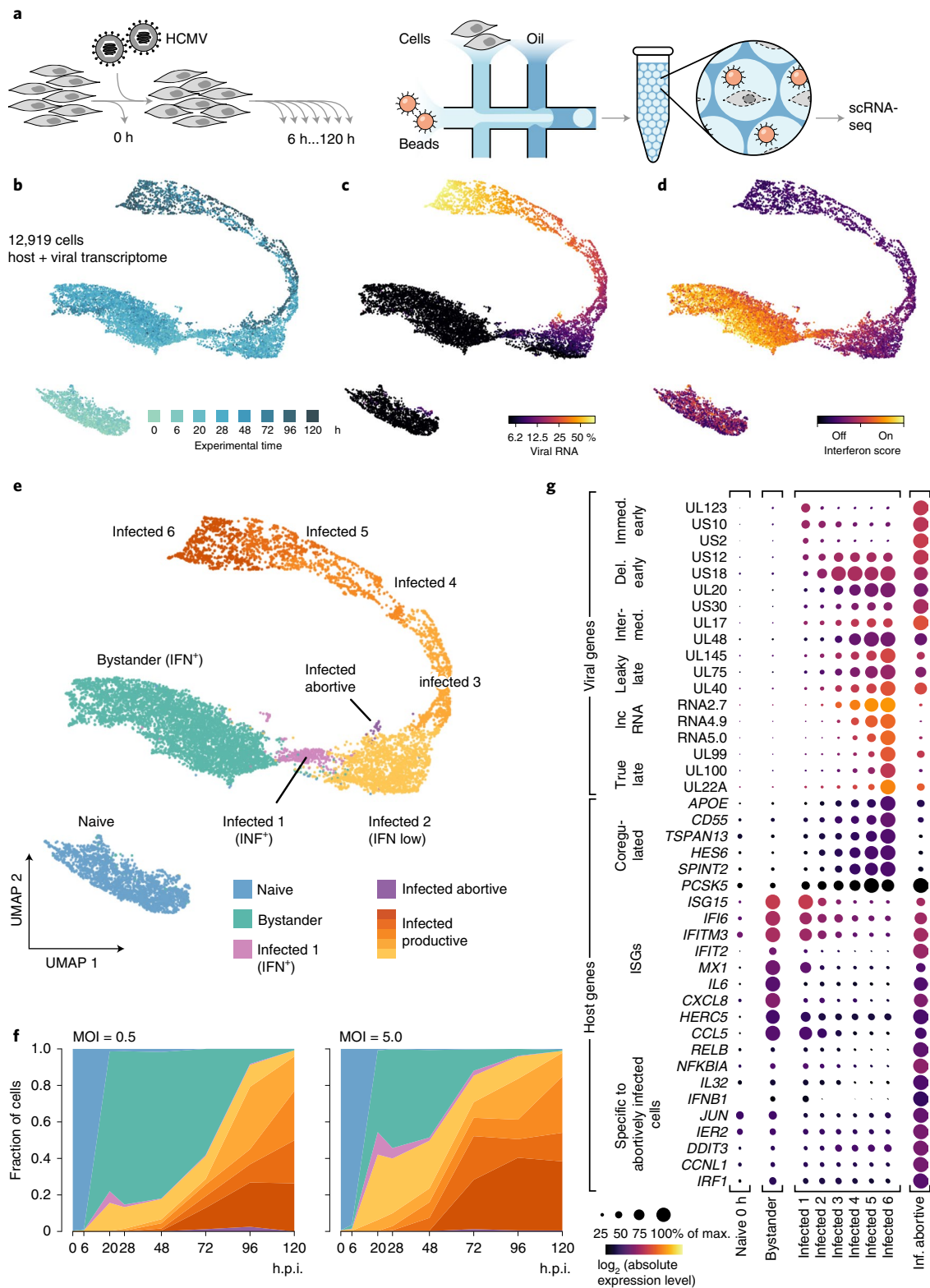
Fig. 3 | Single-cell infection time course defines the lytic cascade of expression events as a trajectory in gene expression space. a, Cells were infected with a low (0.5) or high (5.0) MOI of HCMV, harvested at times ranging from 6 to 120 h.p.i., pooled and subjected to emulsion-based single-cell RNA sequencing (scRNA-seq). **b–e**, UMAP projections of the transcriptomes of 12,919 single cells (both host and viral transcripts), color-coded by experimental time after infection (**b**), percentage of viral transcripts per cell (**c**), IFN score, calculated from the normalized expression of ISGs (**d**), and cluster membership (**e**). **f**, Fraction of cells in each cluster as a function of time after infection and MOI. **g**, Expression dot plot of select viral and host genes in the different clusters. Sizes of dots represent the expression normalized to the cluster with the highest expression for each gene. Colors represent absolute expression levels (scaled UMIs per gene per cell, averaged for all cells of a cluster). lncRNA, long noncoding RNA.
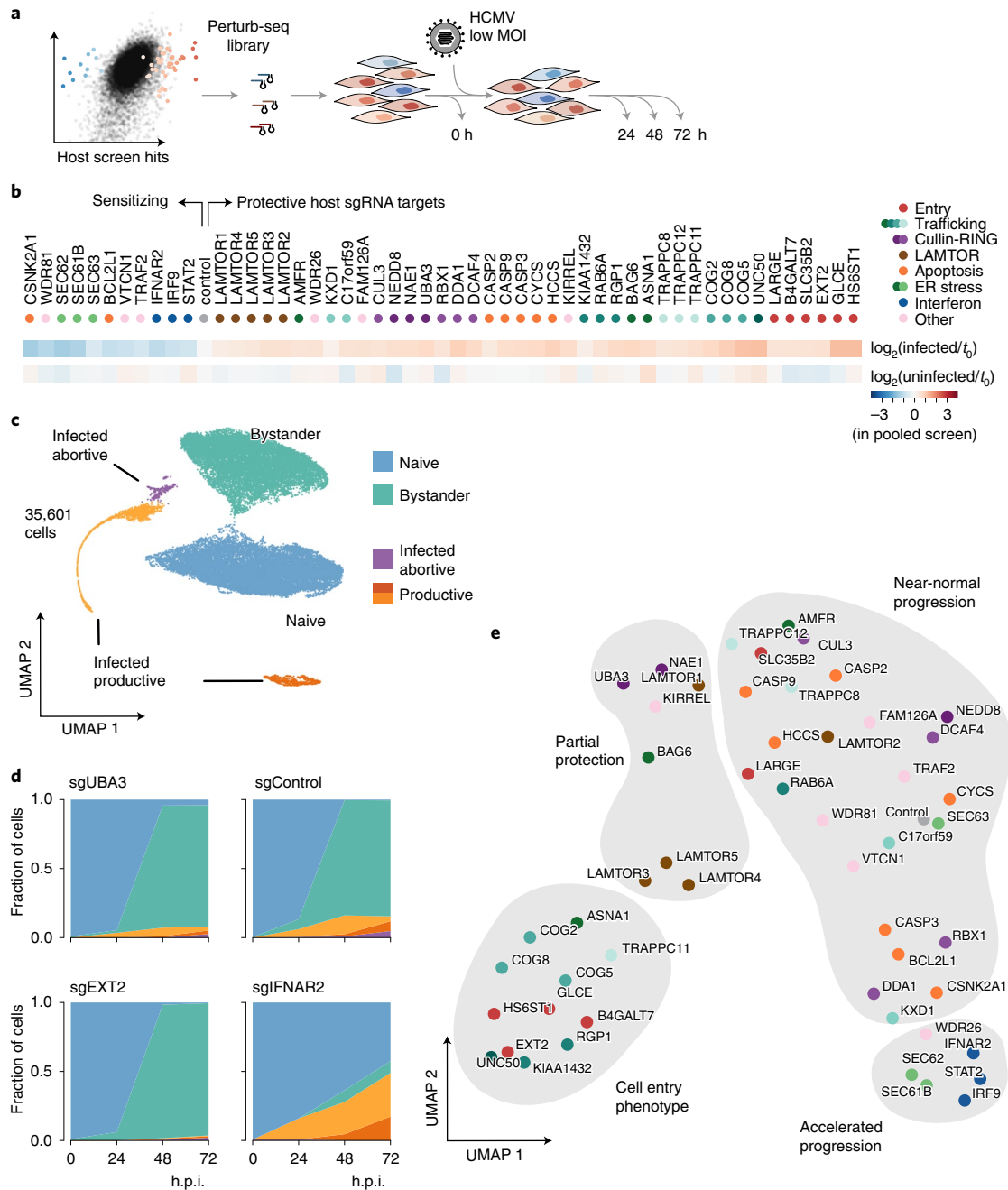
Fig. 4 | Perturbing host factors can alter the propensity of a cell to be infected. a, Host dependency and restriction factors were selected from the pooled screen, cloned into a Perturb-seq library and delivered into dCas9-expressing fibroblasts, which were challenged with an MOI of 0.5 of HCMV for 24–72 h. **b**, Selected host factors have a wide range of sensitizing to protective phenotypes, have varying degrees of essentiality and cover different pathways. **c**, UMAP projection of the transcriptomes of 35,601 cells with confidently identified sgRNAs shows the same naive, bystander and productively and abortively infected clusters found in the unperturbed infection time course (Fig. 3). **d**, Cluster membership as a function of time after infection for cells expressing non-targeting control sgRNAs (sgControl), as well as sgRNAs targeting *UBA3*, *EXT2* and *IFNAR2*, as representatives for the different types of responses. For a complete set of cluster membership graphs, see Extended Data Fig. 5h. **e**, UMAP representation of the cluster membership data (Extended Data Fig. 5h) organizes host factors by their phenotypes of altered progression of infection in single cells, spanning cell entry phenotypes, partial protection from infection, near-normal progression and accelerated progression of infection.

the cullin-RING/neddylation machinery, as well as mild responses to the knockdown of vesicle trafficking factors. The patterns of the transcriptional responses to knockdown organized host factors by pathway in a principled fashion (Extended Data Fig. 5c), providing a layer of information that remains unresolved by pooled screening.

We observed the same split of the population into a naive, a bystander and an infected cluster, branching into a productive and an

abortive trajectory (Fig. 4c and Extended Data Fig. 5d–f). Cells from the uninfected (0 h) sample were overwhelmingly in the naive cluster and transitioned into the bystander and infected clusters, starting at 24 h.p.i. (Fig. 4d, Extended Data Fig. 5h and Supplementary Table 4). The kinetics of transition of cells between the clusters was markedly different in cells with certain host factor knockdowns compared to control cells. Targeting members of the heparan sulfate biosynthesis
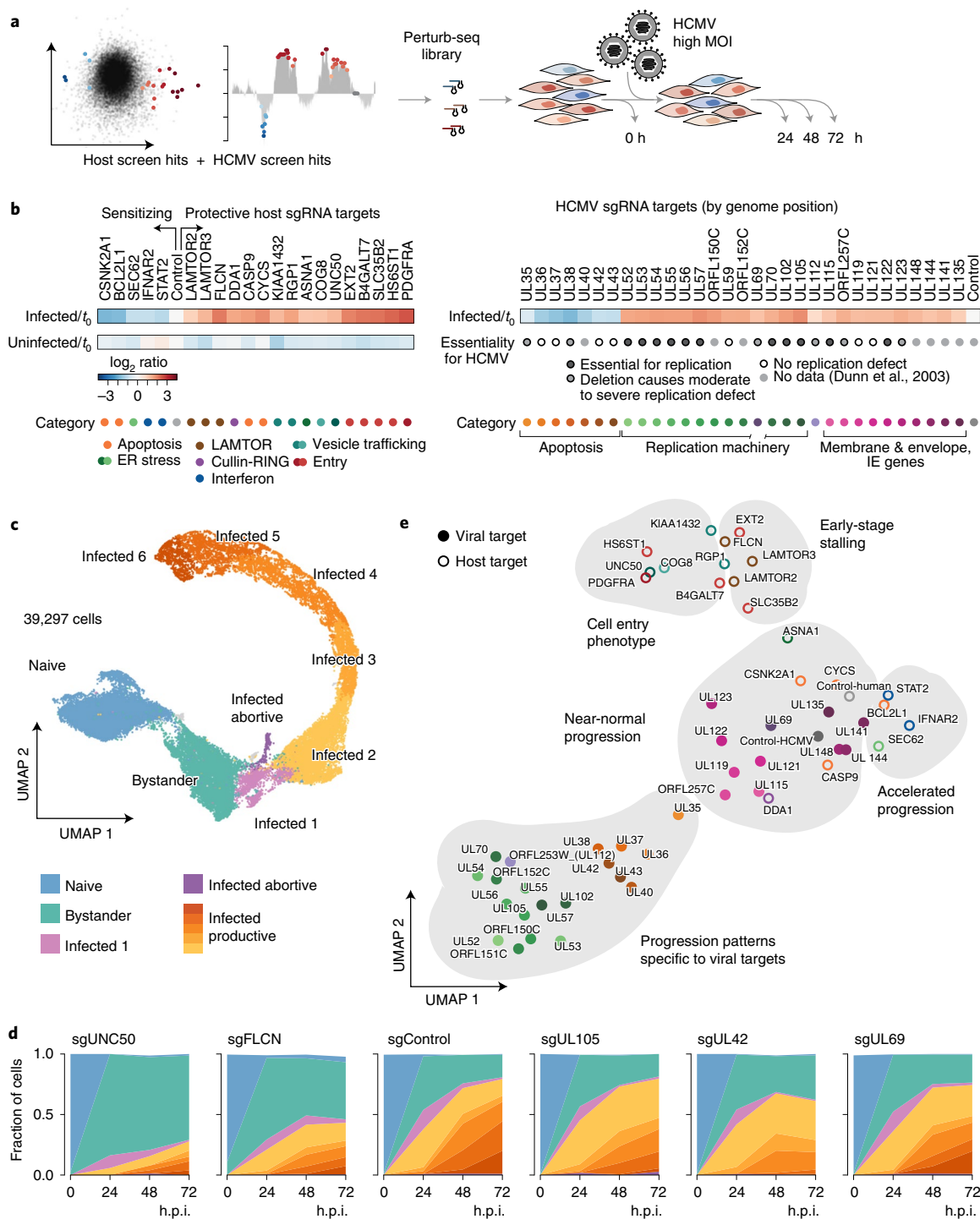
Fig. 5 | Host- and virus-directed perturbations stall or accelerate progression or shift the patterns of viral gene expression. a, Host and viral factors were selected from the pooled screens, cloned into a Perturb-seq library and delivered into Cas9-expressing fibroblasts, which were challenged with an MOI of 5.0 of HCMV for 24–72 h. **b**, Selected factors organized by their respective phenotypes in the pooled screens, essentiality for the host (determined by the uninfected arm of our pooled screen) and the virus[10], and pathway membership. IE, immediate early. **c**, UMAP projection of the transcriptomes of 39,297 cells with confidently identified sgRNAs shows the same naive, bystander and productively and abortively infected clusters found in the unperturbed infection time course (Fig. 3) and the host-directed Perturb-seq dataset (Fig. 4). **d**, Cluster membership as a function of time after infection for cells expressing sgRNAs targeting two host factors (*UNC50* and *FLCN*), a safe-target region of the viral genome (sgControl) and three viral factors (UL105, UL42 and UL69), as representatives for the different types of responses. For a complete set of cluster membership graphs, see Extended Data Fig. 6c,d. **e**, UMAP representation of the cluster membership data (Extended Data Fig. 6c,d) organizes host and viral factors by their phenotypes of altered progression of infection in single cells, spanning cell entry phenotypes, partial protection from infection, near-normal progression and accelerated progression of infection as well as patterns specific to viral targets.

pathway, such as *EXT2*, COG complex members, the KIAA1432/RIC1–RGP1 complex and other trafficking factors such as *UNC50* efficiently prevented infection. Cells lacking the NEDD8-activating enzyme subunits *UBA3* and *NAE1*, as well as LAMTOR complex members, became infected but in decreased numbers. Conversely, targeting *SEC61B*, a non-essential subunit of the translocon, increased the numbers of infected cells at 24 h. Similarly, targeting the interferon receptor subunit *IFNAR2*, or its downstream effectors *STAT2* and *IRF9*, increased infection rates early. Additionally, cells with those perturbations failed to mount the interferon-driven transcriptional response characteristic of bystander cells and remained transcriptionally naive as long as they stayed uninfected. Those cells kept getting infected at increased rates, especially at later time points, when most other cells showed a robust interferon response. For a systematic classification of host targets by their progression phenotypes, we performed dimensionality reduction of the temporal cluster membership data (Extended Data Fig. 5h), organizing the different host factors by phenotype on a spectrum ranging from cell entry defects to accelerated progression (Fig. 4e).

Next, we extended our Perturb-seq analysis to viral factors, using CRISPRn as the mode of genetic perturbation. We reasoned that, when targeting a viral factor, only infected cells would carry meaningful information. We, therefore, challenged the cells with a high MOI of 5.0 of HCMV, without removing the inoculum, to maximize the proportion of infected cells (Fig. 5a). We selected 31 viral gene targets based on their strong phenotypes in the pooled, virus-directed screens. Moreover, we added knockout guides targeting a representative set of 21 host factors as well as safe-targeting guides targeting non-essential regions of the human and HCMV genomes (Fig. 5b and Supplementary Table 5). We recovered 188 ± 77 cells per target per time point (Extended Data Fig. 6a). Our experimental conditions resulted in more than 50% infected cells at 24 h.p.i. and good representation of cells in the different infected subclusters (Fig. 5c).

The progression of infection again varied widely depending on the targeted gene, visualized by the distributions of viral loads for each target (Extended Data Fig. 6b) and cluster membership of cells with a given target (Fig. 5c,d, Extended Data Fig. 6c,d and Supplementary Table 5). Host-directed knockouts confirmed our findings from the host-directed CRISPRi experiment. The high percentages of infected cells improved the resolution of some protective phenotypes, distinguishing two scenarios: a reduced propensity of a cell to be infected versus delays in progression from the early to later stages of infection (Extended Data Fig. 6c). For instance, knockout of *PDGFRA* (the proposed viral receptor on fibroblasts[38]) or *HS6ST1* (involved in heparan sulfate biosynthesis) almost entirely prevented infection, even with high MOI. Similar levels of protection were observed in cells where *COG8* or *UNC50* was knocked out, implicating these factors in viral entry. Conversely, perturbation of *FLCN*, *LAMTOR2/3*, *KIAA1432/RIC1* and *RGP1* permitted infection (albeit at reduced levels) but substantially slowed the progression of infection to the late stage, indicating that these factors are essential in early-stage infection, acting downstream of viral uncoating but before genome replication. Representative examples for each progression phenotype are shown in Fig. 5d.

Virus-directed perturbations alter the infection trajectory. Compared to host genes, targeting viral genes led to qualitatively different outcomes (Fig. 5d,e and Extended Data Fig. 6d). Expectedly, cells with virus-targeting sgRNAs generally had equal propensities to become infected (Extended Data Fig. 6d), with the exception of cells targeting UL122 and UL123, which appeared to have slightly reduced propensities. This finding confirms the known roles of those two genes in initiating immediate-early viral gene expression, which, when suppressed, can make an infected cell present as uninfected in gene expression space.

Once infected, cells with virus-targeting sgRNAs progressed in ways specific to the target gene, evident by more complex patterns of viral load distributions (Extended Data Fig. 6b) and, consequently, of their progression through the different clusters (Extended Data Fig. 6d and Fig. 5d). This observation revealed that targeting a viral factor does not just stall infection at the stage where that factor becomes essential, as is the case with host factors. Rather, cells with virus-targeting guides assume distinct areas in transcriptional space but only in infected cells (Extended Data Fig. 6e–g). Together, these findings prompted us to examine in more detail the patterns of viral gene expression in infected cells.

On a dimensionality-reduced projection of the viral transcriptomes in infected cells, the course of infection can be visualized as a trajectory by a rolling average of the positions of cells with increasing viral load (Fig. 6a). Cells with host-targeting sgRNAs all followed trajectories that are nearly congruent with the default trajectory (defined by cells with control sgRNAs). However, some host factor trajectories were necessarily shorter because some perturbations preclude cells from reaching late-stage infection (Fig. 6b,c). In marked contrast, cells with virus-targeting sgRNAs followed trajectories that diverged from the default (Fig. 6d,e). These differences were driven by viral expression patterns and not by concomitant variations in host gene expression.

To study the nature of these alternate trajectories, we quantified the expression of all viral genes along each trajectory relative to the default trajectory. Extended Data Fig. 7a outlines this analytical workflow. Cells with sgRNAs targeting non-essential regions of the viral genome followed a trajectory virtually unchanged from those with host-targeting controls, showing only mild transcriptional effects on genes in the immediate vicinity of the cut sites, within ~10 kb (Extended Data Fig. 7b and Supplementary Table 5). Reduced expression not just of the target gene itself but also of genes located near the cut sites was a common feature for all virus-targeting sgRNAs. In addition, for all virus-targeting sgRNAs other than our safe-targeting controls, we observed widespread upregulation or downregulation of genes in *trans*—that is, genes encoded far from the sgRNA target—indicating indirect effects on viral gene regulation caused by disruption of the target gene.

When comparing the similarities of these trajectories, we noticed that they came in 'bundles' (Fig. 6e) and, specifically, that targeting genes in the same region of the viral genome tended to result in similar patterns of deregulated viral gene expression (Extended Data Fig. 7b,c). This applies both to the effects on genes in *cis*, which was expected, and for the expression changes of genes in *trans*, which indicates that genes are arranged in the viral genome in func-

Fig. 6 | Virus-directed perturbations create alternative trajectories in viral gene expression space. a, UMAP projection of the viral parts of the transcriptomes of 17,511 cells with >2.5% viral RNA, color-coded by the fraction of viral RNA per cell. **b**, Subsets of cells with host-directed sgRNAs, color-coded by guide identity. **c**, Trajectories of infection for host-directed perturbations, determined by averaging the geometric position of cells with a given sgRNA target, ranked by viral load. **d**, Subsets of cells with virus-directed sgRNAs, color-coded by guide identity. **e**, Trajectories of infection for virus-directed perturbations. **f**, UMAP representation of the different trajectories for each host- and virus-directed sgRNA target, calculated by dimensionality reduction of the viral gene expression matrices for all virus- and host-directed perturbations shown in Extended Data Fig. 7a. All host-directed perturbations (shaded area) result in trajectories that are most similar to the control trajectories defined from both the host and virus safe-targeting controls. **g**, Pairwise correlation matrix of the relative viral gene expression matrices for the different trajectories highlights the three main bundles of trajectories generated by virus-directed perturbations.

tional modules. To understand the relationship of the trajectories caused by targeting different genes, we projected the deviations in gene expression into two dimensions by uniform manifold approxi-

mation and projection (UMAP) (Fig. 6f) and quantified how correlated viral gene expression changes were for the different trajectories (Fig. 6g). Although all trajectories generated by targeting
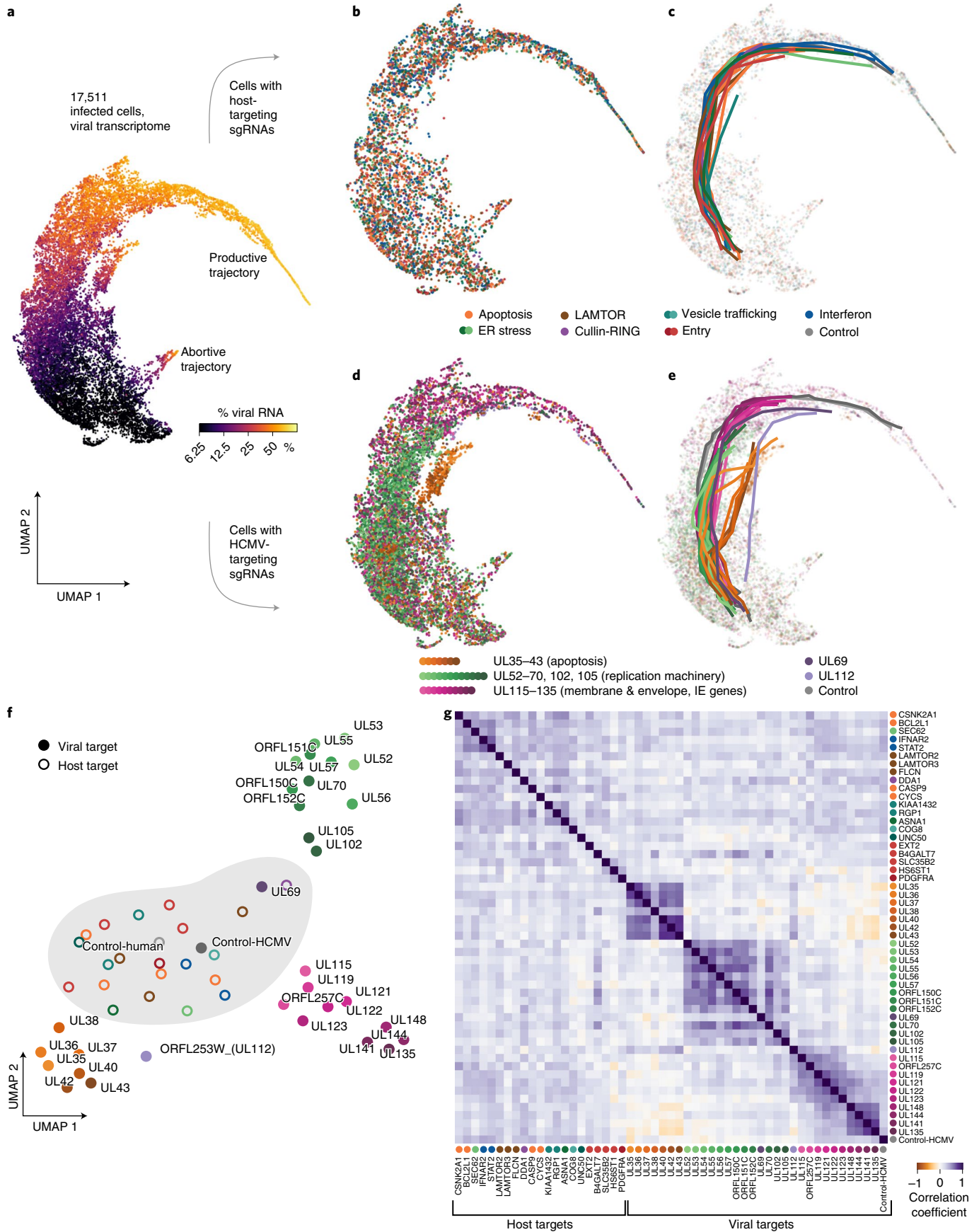
host factors were quantitatively similar to the default trajectory, viral trajectories came in three main classes. The first class of trajectories resulted from targeting genes in the sensitizing region (UL35–UL43). These perturbations caused reduced expression of RNA1.2 and RNA2.7, combined with overexpression to varying degrees of almost all viral genes encoded downstream of UL48, with US3 and UL54 being among the most strongly overexpressed. Cells following these trajectories rarely reached very high viral loads, and representation of those targets dropped substantially in the later time points (Extended Data Fig. 6a), underscoring that these viral trajectories coincide with the cells undergoing apoptosis.

A second trajectory class was linked to perturbations of genes in one of the protective genomic modules (UL52–ORFL152C) as well as to UL102 and UL105, which are located around 50 kb downstream. These genes are all involved in the viral replication machinery. Consequently, cells following these trajectories also failed to reach high viral loads and underexpressed late viral genes, indicating that viral genome replication fails upon targeting those factors.

A third trajectory class corresponded to perturbations of genes located within the UL115–UL148 region. Here, viral gene expression patterns were most similar to the unperturbed trajectory. Of note, the kinetics of progression varied among the targeted genes in this group (Fig. 5c, right panel), with perturbation of the major immediate-early transactivator genes UL122 and UL123 causing the strongest delays.

Notably, two viral gene targets led to trajectories of infection that were distinct from one another and from viral targets in their immediate genomic vicinity: UL69 and UL112 (ORFL253W) (Extended Data Fig. 7b). Both genes showed relatively weak protective phenotypes when targeted in the pooled tiling screen (Extended Data Fig. 1b). Furthermore, both the UL112 and UL69 loci encode microRNAs, which are, however, thought to target host rather than viral genes[53].

UL69 has been described as a transactivator of gene expression[54] and as involved in promoting nuclear export of unspliced RNA[55]. Targeting UL69 caused a distinct viral gene expression pattern, including downregulation of RL12, RL13, UL144 and UL155 and slight overexpression of the noncoding RNA1.2.

The UL112 locus gives rise to multiple gene products by alternative splicing, all of which are likely disrupted with Cas9. Some splice variants are involved in recruitment of the viral polymerase activator UL44 to nuclear replication sites[56]. Targeting the UL112 locus caused a viral gene expression pattern that includes overexpression of US3 as well as upregulation of genes in the 5′ region of the genome, such as of UL22A and UL38. This pattern bears some resemblance to the pattern caused by targeting genes in the UL35–UL43 module, which sensitizes cells to cell death, whereas targeting UL112 is protective.

Taken together, our results show that the trajectory of infection can be derailed in defined ways by targeting viral genes, whereas targeting host factors determines the rate of progression along the default trajectory. This implies that host factors create a permissive environment, whereas viral genes solely orchestrate and control the viral gene expression program.

Discussion

The waves of viral gene expression during lytic infection are a key signature of herpesvirus biology[1]. Our study redefines the lytic cascade at the single-cell level as a continuum of cellular states. We found that the large majority of cells follow this stereotypical trajectory in gene expression space, whereas a small but prominent subpopulation take an alternative, abortive trajectory.

Technologically, our study breaks ground on several levels, including the use of saturating functional screens of a large viral genome, the comprehensive discovery of critical sets of genes on both sides of a host–pathogen system and the use of single-cell

analyses to understand the functional consequences of targeting these factors. Our work establishes Perturb-seq as a powerful method for functional genomics in a highly dynamic virus–host system. The single-cell approach is paramount to both capturing the inherent cell-to-cell heterogeneity of infection and enabling a highly multiplexed, functional analysis of genetic perturbations.

Based on our high-dimensional dataset, we organized host factors both by the transcriptional responses in cells where these factors are targeted and by how infection progresses in those cells. This provides a systematic classification of host factors by functional category that reveals a range of factors acting in viral entry, in early-stage infection and as restriction factors. Reading out genetic perturbation phenotypes as rich transcriptional signatures also revealed that, by targeting viral factors, infected cells progress along trajectories in expression space that are both non-productive and not encountered normally.

HCMV is entirely dependent on the transcriptional and translational machinery of its host. At the same time, our findings indicate that, in fibroblasts, the lytic cascade, once in motion, is a deterministic program that is hard-wired into the viral genome rather than reactive to the state of the cell. HCMV features a wide cell tropism in its human host, and it harnesses different receptors and entry mechanisms[1,39], underscoring that the set of host factors and possibly critical viral factors will vary across permissive cells. It will be interesting to investigate whether the dichotomy of host-directed versus virus-directed perturbations, which we observed, is a general feature across cell types in the case of HCMV and of other virus–host systems more generally.

Our work provides a roadmap for the design of antiviral combination therapies by selecting sets of targets that drive the virus into distinct non-productive pathways. Similarly, our data can inform the design of attenuated viral strains for vaccine development purposes. More generally, we envision that our approach of single-cell functional genomics can serve as a blueprint for studying other viruses and can define their vulnerabilities to genetic or pharmacological interventions.

Online content

Any methods, additional references, Nature Research reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at https://doi.org/10.1038/s41587-021-01059-3.

References

1. Mocarski Jr., E. S., Schenk, T., Griffiths, P. D. & Pass, R. F. Cytomegaloviruses. in *Fields Virology* 1960–2014 (Lippincott Williams & Wilkins, 2013).
2. Gatherer, D. et al. High-resolution human cytomegalovirus transcriptome. *Proc. Natl Acad. Sci. USA* **108**, 19755–19760 (2011).
3. Marcinowski, L. et al. Real-time transcriptional profiling of cellular and viral gene expression during lytic cytomegalovirus infection. *PLoS Pathog.* **8**, e1002908 (2012).
4. Stern-Ginossar, N. et al. Decoding human cytomegalovirus. *Science* **338**, 1088–1093 (2012).
5. Weekes, M. P. et al. Quantitative temporal viromics: an approach to investigate host-pathogen interaction. *Cell* **157**, 1460–1472 (2014).
6. Tirosh, O. et al. The transcription and translation landscapes during human cytomegalovirus infection reveal novel host-pathogen interactions. *PLoS Pathog.* **11**, e1005288 (2015).
7. Jean Beltran, P. M., Mathias, R. A. & Cristea, I. M. A portrait of the human organelle proteome in space and time during cytomegalovirus infection. *Cell Syst.* **3**, 361–373 (2016).
8. Sinzger, C. et al. Fibroblasts, epithelial cells, endothelial cells and smooth muscle cells are major targets of human cytomegalovirus infection in lung and gastrointestinal tissues. *J. Gen. Virol.* **76**, 741–750 (1995).

9. Yu, D., Silva, M. C. & Shenk, T. Functional map of human cytomegalovirus AD169 defined by global mutational analysis. *Proc. Natl Acad. Sci. USA* **100**, 12396–12401 (2003).
10. Dunn, W. et al. Functional profiling of a human cytomegalovirus genome. *Proc. Natl Acad. Sci. USA* **100**, 14223–14228 (2003).
11. Puschnik, A. S., Majzoub, K., Ooi, Y. S. & Carette, J. E. A CRISPR toolbox to study virus-host interactions. *Nat. Rev. Microbiol.* **15**, 351–364 (2017).
12. Russell, A. B., Trapnell, C. & Bloom, J. D. Extreme heterogeneity of influenza virus infection in single cells. *eLife* **7**, e32303 (2018).
13. Erhard, F. et al. scSLAM-seq reveals core features of transcription dynamics in single cells. *Nature* **571**, 419–423 (2019).
14. van Diemen, F. R. et al. CRISPR/Cas9-mediated genome editing of herpesviruses limits productive and latent infections. *PLoS Pathog.* **12**, e1005701 (2016).
15. Van Damme, E. & Van Loock, M. Functional annotation of human cytomegalovirus gene products: an update. *Front. Microbiol.* **5**, 218 (2014).
16. McCormick, A. L., Skaletskaya, A., Barry, P. A., Mocarski, E. S. & Goldmacher, V. S. Differential function and expression of the viral inhibitor of caspase 8-induced apoptosis (vICA) and the viral mitochondria-localized inhibitor of apoptosis (vMIA) cell death suppressors conserved in primate and rodent cytomegaloviruses. *Virology* **316**, 221–233 (2003).
17. Gilbert, L. A. et al. Genome-scale CRISPR-mediated control of gene repression and activation. *Cell* **159**, 647–661 (2014).
18. Horlbeck, M. A. et al. Compact and highly active next-generation libraries for CRISPR-mediated gene repression and activation. *eLife* **5**, e19760 (2016).
19. Compton, T., Nowlin, D. M. & Cooper, N. R. Initiation of human cytomegalovirus infection requires initial interaction with cell surface heparan sulfate. *Virology* **193**, 834–841 (1993).
20. Hetzenecker, S., Helenius, A. & Krzyzaniak, M. A. HCMV induces macropinocytosis for host cell entry in fibroblasts. *Traffic* **17**, 351–368 (2016).
21. Pusapati, G. V., Luchetti, G. & Pfeffer, S. R. Ric1-Rgp1 complex is a guanine nucleotide exchange factor for the late Golgi Rab6A GTPase and an effector of the medial Golgi Rab33B GTPase. *J. Biol. Chem.* **287**, 42129–42137 (2012).
22. Smith, R. D. et al. The COG complex, Rab6 and COPI define a novel Golgi retrograde trafficking pathway that is exploited by SubAB toxin. *Traffic* **10**, 1502–1517 (2009).
23. Liu, S., Dominska-Ngowe, M. & Dykxhoorn, D. M. Target silencing of components of the conserved oligomeric Golgi complex impairs HIV-1 replication. *Virus Res.* **192**, 92–102 (2014).
24. Bassik, M. C. et al. A systematic mammalian genetic interaction map reveals pathways underlying ricin susceptibility. *Cell* **152**, 909–922 (2013).
25. Selyunin, A. S., Iles, L. R., Bartholomeusz, G. & Mukhopadhyay, S. Genome-wide siRNA screen identifies UNC50 as a regulator of Shiga toxin 2 trafficking. *J. Cell Biol.* **216**, 3249–3262 (2017).
26. Starling, G. P. et al. Folliculin directs the formation of a Rab34-RILP complex to control the nutrient-dependent dynamic distribution of lysosomes. *EMBO Rep.* **17**, 823–841 (2016).
27. Filipek, P. A. et al. LAMTOR/Ragulator is a negative regulator of Arl8b- and BORC-dependent late endosomal positioning. *J. Cell Biol.* **216**, 4199–4215 (2017).
28. Pu, J., Keren-Kaplan, T. & Bonifacino, J. S. A Ragulator-BORC interaction controls lysosome positioning in response to amino acid availability. *J. Cell Biol.* **216**, 4183–4197 (2017).
29. Becker, T., Le-Trilling, V. T. K. & Trilling, M. Cellular cullin RING ubiquitin ligases: druggable host dependency factors of cytomegaloviruses. *Int. J. Mol. Sci.* **20**, 163 (2019).
30. Le-Trilling, V. T. K. et al. Broad and potent antiviral activity of the NAE inhibitor MLN4924. *Sci. Rep.* **6**, 19977 (2016).
31. Ott, M., Marques, D., Funk, C. & Bailer, S. M. Asna1/TRC40 that mediates membrane insertion of tail-anchored proteins is required for efficient release of herpes simplex virus 1 virions. *Virol. J.* **13**, 175 (2016).
32. Nishitsuji, H., Ujino, S., Harada, K. & Shimotohno, K. TIP60 complex inhibits hepatitis B virus transcription. *J. Virol.* **92**, e01788-17 (2018).
33. Kamine, J., Elangovan, B., Subramanian, T., Coleman, D. & Chinnadurai, G. Identification of a cellular protein that specifically interacts with the essential cysteine region of the HIV-1 Tat transactivator. *Virology* **216**, 357–366 (1996).
34. Adamson, B. et al. A multiplexed single-cell CRISPR screening platform enables systematic dissection of the unfolded protein response. *Cell* **167**, 1867–1882 (2016).
35. Cohen, Y. et al. The yeast p5 type ATPase, spf1, regulates manganese transport into the endoplasmic reticulum. *PLoS ONE* **8**, e85519 (2013).
36. Wu, K., Oberstein, A., Wang, W. & Shenk, T. Role of PDGF receptor-α during human cytomegalovirus entry into fibroblasts. *Proc. Natl Acad. Sci. USA* **115**, E9889–E9898 (2018).
37. Soroceanu, L., Akhavan, A. & Cobbs, C. S. Platelet-derived growth factor-α receptor activation is required for human cytomegalovirus infection. *Nature* **455**, 391–395 (2008).
38. Kabanova, A. et al. Platelet-derived growth factor-α receptor is the cellular receptor for human cytomegalovirus gHgLgO trimer. *Nat. Microbiol.* **1**, 16082 (2016).
39. Martinez-Martin, N. et al. An unbiased screen for human cytomegalovirus identifies neuropilin-2 as a central viral receptor. *Cell* **174**, 1158–1171 (2018).
40. Tzelepis, K. et al. A CRISPR dropout screen identifies genetic vulnerabilities and therapeutic targets in acute myeloid leukemia. *Cell Rep.* **17**, 1193–1205 (2016).
41. Morgens, D. W., Deans, R. M., Li, A. & Bassik, M. C. Systematic comparison of CRISPR/Cas9 and RNAi screens for essential genes. *Nat. Biotechnol.* **34**, 634–636 (2016).
42. Datlinger, P. et al. Pooled CRISPR screening with single-cell transcriptome readout. *Nat. Methods* **14**, 297–301 (2017).
43. Dixit, A. et al. Perturb-seq: dissecting molecular circuits with scalable single-cell RNA profiling of pooled genetic screens. *Cell* **167**, 1853–1866 (2016).
44. Jaitin, D. A. et al. Dissecting immune circuits by linking CRISPR-pooled screens with single-cell RNA-seq. *Cell* **167**, 1883–1896 (2016).
45. Zanini, F., Pu, S.-Y., Bekerman, E., Einav, S. & Quake, S. R. Single-cell transcriptional dynamics of flavivirus infection. *eLife* **7**, e32942 (2018).
46. Drayman, N., Patel, P., Vistain, L. & Tay, S. HSV-1 single-cell analysis reveals the activation of anti-viral and developmental programs in distinct sub-populations. *eLife* **8**, e46339 (2019).
47. Wyler, E. et al. Single-cell RNA-sequencing of herpes simplex virus 1-infected cells connects NRF2 activation to an antiviral program. *Nat. Commun.* **10**, 4878 (2019).
48. Kotliar, D. et al. Single-cell profiling of Ebola virus disease in vivo reveals viral and host dynamics. *Cell* **183**, 1383–1401 (2020).
49. An, P. et al. Single cell transcriptomics reveals a heterogeneous cellular response to BKV infection. *J. Virol.* **95**, e02237 (2020).
50. Bogdanow, B. et al. Human cytomegalovirus tegument protein pp150 acts as a cyclin A2-CDK-dependent sensor of the host cell cycle and differentiation state. *Proc. Natl Acad. Sci. USA* **110**, 17510–17515 (2013).
51. Spear, G. T. et al. Host cell-derived complement control proteins CD55 and CD59 are incorporated into the virions of two unrelated enveloped viruses. Human T cell leukemia/lymphoma virus type I (HTLV-I) and human cytomegalovirus (HCMV). *J. Immunol.* **155**, 4376–4381 (1995).
52. Isomura, H. & Stinski, M. F. Coordination of late gene transcription of human cytomegalovirus with viral DNA synthesis: recombinant viruses as potential therapeutic vaccine candidates. *Expert Opin. Ther. Targets* **17**, 157–166 (2013).
53. Zhang, L., Yu, J. & Liu, Z. MicroRNAs expressed by human cytomegalovirus. *Virol. J.* **17**, 34 (2020).
54. Winkler, M., Rice, S. A. & Stamminger, T. UL69 of human cytomegalovirus, an open reading frame with homology to ICP27 of herpes simplex virus, encodes a transactivator of gene expression. *J. Virol.* **68**, 3943–3954 (1994).
55. Lischka, P., Toth, Z., Thomas, M., Mueller, R. & Stamminger, T. The UL69 transactivator protein of human cytomegalovirus interacts with DEXD/H-Box RNA helicase UAP56 to promote cytoplasmic accumulation of unspliced RNA. *Mol. Cell. Biol.* **26**, 1631–1643 (2006).
56. Schommartz, T., Tang, J., Brost, R. & Brune, W. Differential requirement of human cytomegalovirus UL112–113 protein isoforms for viral replication. *J. Virol.* **91**, e00254-17 (2017).

Methods

Cell and virus culture. Human foreskin fibroblasts (HFFs; CRL-1634) and HCMV (strain Merlin; VR-1590) were purchased from the American Type Culture Collection. HFFs were cultured in DMEM, supplemented with 10% FBS and penicillin/streptomycin. HCMV stocks were expanded by two rounds of propagation on HFFs and titered by serial dilution.

For stable expression of the CRISPRi/n machineries in HFFs, we modified established lentiviral (d)Cas9 expression vectors[17] by inserting a minimal ubiquitous chromatin opening element (UCOE)[57] upstream of the SFFV promoter, resulting in pMH0001 (UCOE-SFFV-dCas9-BFP-KRAB; Addgene no. 85969) and pMH0004 (UCOE-SFFV-Cas9-BFP; Addgene no. 174087). The UCOE prevented epigenetic silencing that affected the original constructs.

Pooled CRISPR screening. The HCMV tiling library was designed to contain sgRNAs targeting every single one of the 33,465 PAMs in the HCMV Merlin genome (NC_006273.2) as well as 533 non-targeting controls (Supplementary Table 1). It was synthesized and cloned into a lentiviral vector (Addgene no. 84832) as previously described[17,18]. For targeting host genes, we used the human CRISPRi v2 library (Addgene no. 83969)[18] and the Yusa et al. human knockout CRISPR v1 library (Addgene no. 67989)[40].

Libraries were packaged into lentiviruses and delivered into (d)Cas9-expressing HFFs at an MOI of 0.3–0.5, followed by puromycin selection. Pooled screens were carried out at 500–1,000× coverage—that is, ~500–1,000 cells per library element per sample taken.

A t_0 sample was harvested, and the remaining cells either were passaged normally or were infected with HCMV at an MOI of 0.5–1.0 (for the HCMV tiling screens) or 0.1 (for the host-directed screens). Infected flasks were washed with PBS and given fresh medium at days 3, 5 and 7 after infection to remove dead cells and harvested at days 7–10. Genomic DNA was extracted and digested with MfeI (pCRISPRia v2-based libraries) or HindIII (Yusa et al. library) to release a fragment containing the sgRNA cassette, followed by gel-based extraction, PCR amplification and deep sequencing as described[18].

Raw data were normalized for read depth, and a small constant was added to account for missing values. Phenotypes of individual sgRNAs were expressed as $\log_2$-transformed ratios of adjusted read counts between samples (Supplementary Table 2). We calculated the mean of all sgRNAs specific to each host gene and computed MAGeCK scores[58]. For the HCMV tiling screen, we calculated a rolling average in a 250-bp window, with the average of all non-targeting sgRNAs defining the baseline.

Single-cell RNA sequencing. For the single-cell infection time course, wild-type HFFs were lentivirally transduced with barcoded Perturb-seq vectors to encode the experimental condition (pBA571, Addgene no. 85968; Supplementary Table 3), followed by puromycin selection. Cells were seeded at a density of 250,000 per well of a 12-well plate and infected with an MOI of 0.5 or 5.0 with no additional medium change before harvest. Infection times were staggered so that all time points for a given MOI were harvested in parallel and pooled, aiming for roughly equal cell numbers for each time point, with a slight over-representation of the 20- and 28-h time points (Extended Data Fig. 3a). For each MOI, pools of ~10,000 cells were prepared for single-cell transcriptomics using one lane each of the Chromium Single Cell 3′ Gene Expression Solution v2 according to the manufacturer's instructions (10x Genomics) and sequenced on a NovaSeq platform (Illumina) at ~100,000 reads per cell. Barcodes encoding the experimental condition were PCR amplified from the final library and sequenced as a 5% spike-in as previously described[34].

Perturb-seq. For the host-directed CRISPRi Perturb-seq experiment, we initially selected 53 candidate genes by their strong protective or sensitizing phenotypes in the pooled screen (one gene was later removed during analysis; see below). We manually picked the two best-performing sgRNAs for each candidate. Additionally, we added six control constructs targeting green fluorescent protein (which is not present in our HFFs).

For the host- and virus-directed CRISPRn Perturb-seq experiment, we selected a set of 21 host factors, of which 19 were already among the targets of the CRISPRi Perturb-seq experiment and had no strong essentiality knockout phenotypes and similar protective or sensitizing phenotypes in both the pooled host-directed CRISPRi and CRISPRn screens (Extended Data Fig. 2). We further added PDGFRA and FLCN, both of which were strong hits in the pooled CRISPRn screen. For each host target, we manually picked the two best-performing sgRNAs from the pooled screen.

In addition, we selected 31 viral targets with strong protective or sensitizing phenotypes, corresponding to the three strongest modules identified in the HCMV tiling screen (Fig. 1b and Extended Data Fig. 1b). From the tiling screen, we selected the two highest-ranking sgRNAs for each target gene based on the following scoring system. From the pool of unique sgRNAs falling within the gene boundaries and having a Doench score[59] of more than 0.5, we calculated the absolute average phenotype across replicates and subtracted a penalty defined as the difference between replicates plus the average absolute essentiality phenotypes on a $\log_2$ scale.

We designed several safe-targeting control sgRNAs targeting intergenic DNA in the US2–US12 region. This region was selected based on its near-neutral phenotypes in the tiling screen (Extended Data Fig. 1b), its lack of essential genes[9,10] and its comparatively large spaces between consensus genes. Furthermore, in some bacterial artificial chromosome (BAC) constructs harboring HCMV genomes, this region was replaced by the BAC backbone, underlining its non-essential nature during infection in tissue culture[60]. We picked five sgRNAs based on their Doench scores from a pool of unique sgRNAs targeting the intergenic regions and having survival and essentiality phenotypes of <0.5 ($\log_2$ scale) in all replicates. In addition, we included four control sgRNAs directed against safe-harbor loci in the host genome, which we repurposed from gene knock-in applications.

All sgRNAs were synthesized as individual oligonucleotide pairs (IDT) and cloned into a barcode library-containing plasmid pool (pBA571, Addgene no. 85968), thereby linking each sgRNA to a unique guide barcode contained within the 3′ untranslated region of the puromycin resistance gene[34]. Barcodes were validated to not contain homo-oligomers or sequences resembling transcription termination signals. All sgRNA and barcode sequences are listed in Supplementary Tables 4 and 5.

sgRNA vectors were individually packaged into lentiviruses, titered separately and pooled to ensure equal representation. This workflow prevents scrambling of guide sequences and associated barcodes by recombination, which is a concern in pooled lentivirus preparations[61]. We delivered the pooled library into (d)Cas9-expressing HFFs at an MOI of 0.3 followed by puromycin selection. Cells were seeded at 250,000 per well of a 12-well plate and infected with HCMV at an MOI of 0.5 for 1 h, followed by medium exchange (for the CRISPRi host-directed experiment), or an MOI of 5.0, leaving the inoculum on the cells, with the goal of maximizing the numbers of infected cells (for the CRISPRn host- and virus-directed experiment). Cells were harvested in the uninfected state (designated as 0 h) and at 24, 48 and 72 h.p.i. We aimed at a representation of each library element by around 100 cells per time point (for actual cell numbers, see Extended Data Figs. 5a and 6a).

Cells were collected and prepared for scRNA-seq using the 10x Chromium platform as described above for the single-cell infection time course. Libraries were sequenced on a HiSeq 4000 (Illumina) at ~40,000 reads per cell.

Single-cell data analysis. Raw sequencing data were submitted to cellranger v2.0.1 (10x Genomics) according to the manufacturer's instructions. We compiled a reference transcriptome from the hg19 human genome and a custom assembly of HCMV coding transcripts based on our previous ribosome profiling dataset[4] as distributed as part of the 'Plastid' Python library demo dataset[62]. We manually added four well-established long noncoding RNA transcripts (RNA1.2, 2.7, 4.9 and 5.0). Internal open reading frames (iORFs) were removed as they would create ambiguous mappings, as were ORFs overlapping with the aforementioned long noncoding RNAs.

Cells retained in the final dataset had to cross the default cellranger quality thresholds as well as have one unique lentiviral barcode assigned with high confidence[34]. During data analysis of the Perturb-seq experiments, three CRISPRn sgRNAs targeting host genes were removed computationally because they were found to be inactive, as seen by lack of transcriptional responses and viral load patterns similar to those in cells with control sgRNAs. One host gene, RBBP5, was similarly excluded from both the CRISPRi and CRISPRn datasets as it became apparent that its knockdown/knockout causes differentiation of cells and a strong transcriptional response rather than true protection against infection (Supplementary Tables 4 and 5).

Percentages of viral RNA (viral loads) were calculated as the fraction of total UMIs per cell mapping to viral genes.

Gene expression was normalized in each cell by a factor scaling the total UMIs mapping to human transcripts to its average number across all cells in a population. This accounts for the fact that infected cells have much higher total UMI counts, indicating that viral transcripts go 'on top' of human transcripts (Extended Data Fig. 3b).

From the unperturbed time course experiment, we defined a set of robustly detected genes as those with more than 10,000 UMIs total across all cells in that population (3,588 genes in total, of which 106 are viral genes). This set of expressed genes was used consistently for the analysis of all whole-transcriptome single-cell datasets in this study.

For heat map representations of gene expression as a function of viral load, cells were binned by viral load, and gene-level expression values were averaged in each bin. Bin widths of 2% or 10% were selected depending on the available number of cells. We visualized a slightly larger set of viral genes, namely those expressed in more than 95% of cells in at least one of these 2% viral load bins (114 genes in total). Viral transcriptome-centric trajectory analyses (Fig. 6 and Extended Data Fig. 7) were also based on this set of viral genes.

Cell cycle phases were scored based on marker genes as previously described[34]. Using a similar approach, we calculated an IFN score by summing (in each cell) and subsequently z-scoring (across cells) the normalized expression values of the following set of robustly quantified ISGs: PSMB8, PSMB9, PSME1, PSME2, ISG15,

ISG20, IRF7, MX1, MX2, GBP1, GBP2, GBP3, IFI6, IFI44, IFI35, IFI16, IFI27, IFIH1, IFI44L, IFIT1, IFIT2, IFIT3, IFIT5, IFITM1, IFITM2, IFITM3, EIF2AK2, OAS1, OAS2, OAS3, CNP, PLSCR1, BST2, BTN3A2, XAF1, CASP1, CASP4, CASP7 and *GSDMD*.

To visualize single-cell datasets, we performed dimensionality reduction by UMAP[63] (v0.1.1) based on the matrix of scaled expression values of the set of robustly detected genes (host + viral genes in Figs. 3 and 4 and Extended Data Figs. 3 and 4; viral genes in Fig. 6 and Extended Data Fig. 6). Clusters of cells were defined by Leiden clustering[64] (leidenalg v0.8.0) or HDBSCAN[65] (v0.8.26). To determine trajectories, selected cells were ranked by viral load, and the geometric position of cells was averaged in a sliding window that was shifted in increments of 0.2 window sizes. Window sizes were selected based on the total number of available cells: 100 cells for each sgRNA target; 500 cells for cells with control sgRNAs.

UMAP was also used for a dimensionality-reduced visualization of the similarities of the cluster membership data as a function of time and sgRNA target (Figs. 4e and 5e; underlying data in Supplementary Tables 4 and 5) and of the viral gene expression data along the trajectories defined by cells with individual sgRNA targets (Fig. 6f; underlying data in Supplementary Table 5).

Reporting Summary. Further information on research design is available in the Nature Research Reporting Summary linked to this article.

Data availability
Raw and processed sequencing data from all pooled screens and single-cell experiments were uploaded to the Gene Expression Omnibus (GSE165291).

Code availability
We used published software for pooled screen data processing (https://github.com/mhorlbeck/ScreenProcessing) and for the analysis of Perturb-seq data (https://github.com/thomasmaxwellnorman/perturbseq_demo) with modifications.

References
57. Müller-Kuller, U. et al. A minimal ubiquitous chromatin opening element (UCOE) effectively prevents silencing of juxtaposed heterologous promoters by epigenetic remodeling in multipotent and pluripotent stem cells. *Nucleic Acids Res.* **43**, 1577–1592 (2015).
58. Li, W. et al. MAGeCK enables robust identification of essential genes from genome-scale CRISPR/Cas9 knockout screens. *Genome Biol.* **15**, 554 (2014).
59. Doench, J. G. et al. Optimized sgRNA design to maximize activity and minimize off-target effects of CRISPR–Cas9. *Nat. Biotechnol.* **34**, 184–191 (2016).
60. Borst, E. M., Hahn, G., Koszinowski, U. H. & Messerle, M. Cloning of the human cytomegalovirus (HCMV) genome as an infectious bacterial artificial chromosome in *Escherichia coli*: a new approach for construction of HCMV mutants. *J. Virol.* **73**, 8320–8329 (1999).
61. Adamson, B., Norman, T. M., Jost, M. & Weissman, J. S. Approaches to maximize sgRNA-barcode coupling in Perturb-seq screens. Preprint at *bioRxiv* https://doi.org/10.1101/298349 (2018).
62. Dunn, J. G. & Weissman, J. S. Plastid: nucleotide-resolution analysis of next-generation sequencing and genomics data. *BMC Genomics* **17**, 958 (2016).
63. Becht, E. et al. Dimensionality reduction for visualizing single-cell data using UMAP. *Nat. Biotechnol.* https://doi.org/10.1038/nbt.4314 (2018).
64. Traag, V. A., Waltman, L. & van Eck, N. J. From Louvain to Leiden: guaranteeing well-connected communities. *Sci. Rep.* **9**, 5233 (2019).
65. Ding, J., Shah, S. & Condon, A. densityCut: an efficient and versatile topological approach for automatic clustering of biological data. *Bioinformatics* **32**, 2567–2576 (2016).
66. Heer, J., Kong, N. & Agrawala, M. Sizing the horizon: the effects of chart size and layering on the graphical perception of time series visualizations. in *Proc. 27th Int. Conf. Human Factors in Computing Systems* (ACM Press, 2009).
67. Murphy, E., Rigoutsos, I., Shibuya, T. & Shenk, T. E. Reevaluation of human cytomegalovirus coding potential. *Proc. Natl Acad. Sci. USA* **100**, 13585–13590 (2003).

Author contributions
M.Y.H. and J.S.W. conceptualized the study, interpreted the experiments and wrote the manuscript. M.Y.H. designed and carried out the experiments and analyzed the data.

Competing interests
J.S.W. has filed patent applications related to CRISPRi screening and Perturb-seq. J.S.W. consults for and holds equity in KSQ Therapeutics and Maze Therapeutics and consults for 5AM Ventures.
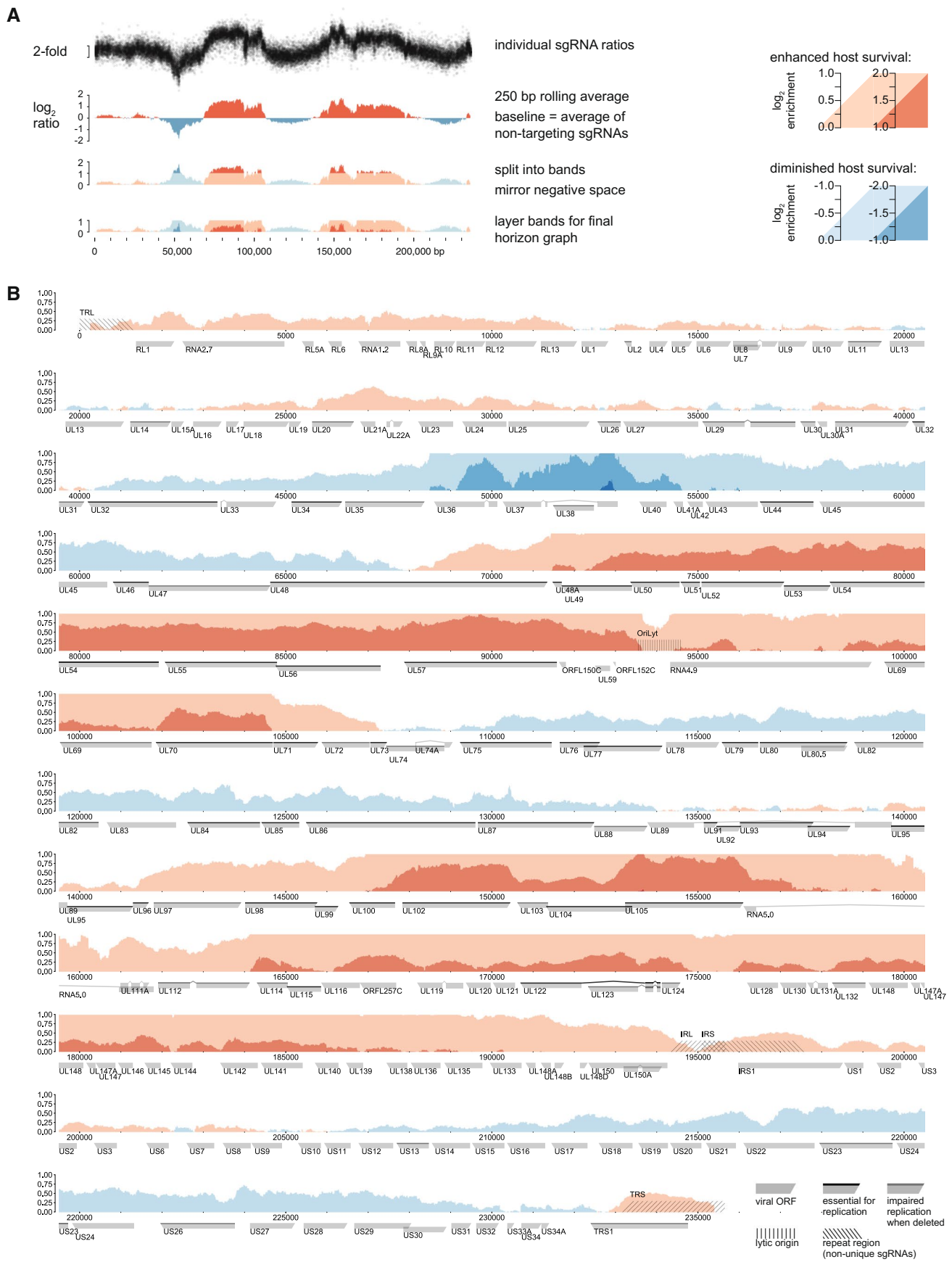
Additional information
Extended data is available for this paper at https://doi.org/10.1038/s41587-021-01059-3.

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41587-021-01059-3.

Correspondence and requests for materials should be addressed to Marco Y. Hein or Jonathan S. Weissman.

Peer review information *Nature Biotechnology* thanks Antoine-Emmanuel Saliba and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

**Reprints and permissions information** is available at www.nature.com/reprints.

A

2-fold

log₂ ratio

individual sgRNA ratios

250 bp rolling average

baseline = average of non-targeting sgRNAs

split into bands

mirror negative space

layer bands for final horizon graph

enhanced host survival:

diminished host survival:

B

Extended Data Fig. 1 | See next page for caption.

viral ORF

essential for replication

impaired replication when deleted

lytic origin

repeat region (non-unique sgRNAs)

Extended Data Fig. 1 | High-resolution HCMV tiling screen. a, Data processing for the HCMV tiling screen. We calculated $\log_2$ ratios of each individual sgRNA in the surviving over the t_0 populations, averaged across two biological replicates. Ratios were averaged in a sliding 250 bp window. The average of the ratios of the non-targeting sgRNA population was set as the baseline. The plot was then colored based on the sign of the average phenotype and layered in bands of decreasing lightness, one $\log_2$ unit wide. The negative space was mirrored on the baseline, and bands were stacked for the final horizon plot representation[66]. **b**, High-resolution horizon graph of the phenotypic landscape of the HCMV genome. Shades of blue denote sensitization to host cell death, shades of red denote protection from cell death upon HCMV genome cleavage. Major features of the HCMV genome are annotated. sgRNAs targeting internal and terminal repeat regions (hashed) typically have multiple target sites and likely result in higher-order fragmentation of the HCMV genome, exacerbating their respective phenotypes. Viral ORFs are classified by their essentiality for viral replication based on ref. [10]. ORFL150C, ORFL151C (originally named UL59, but thought to not be expressed as a protein[67], causing it to be dropped from the consensus annotation), and ORFL152C were the only short ORFs with strong phenotypes in areas of the genome devoid of consensus genes. UL48 was the only gene that showed a substantial phenotype gradient within its gene body: Cutting the N-terminal region caused mild sensitization to death upon infection, whereas cutting the C-terminus had the opposite effect.
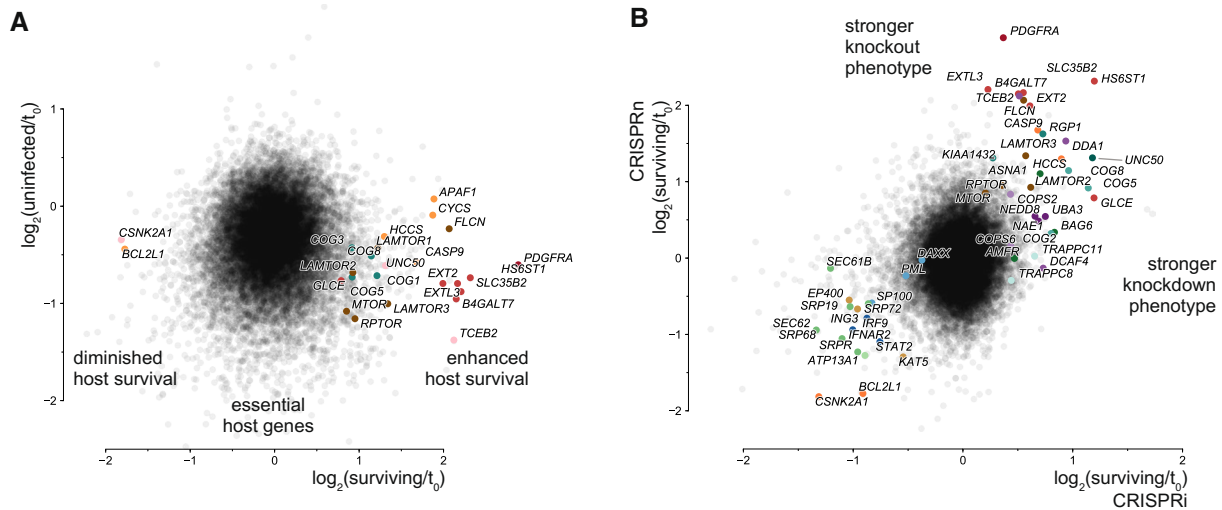
Extended Data Fig. 2 | Host-directed CRISPRi and CRISPRn screen. a, Results of the host-directed CRISPRn screen displayed as a scatter plot of average gene essentiality (that is infection-independent phenotype; y-axis) vs. protection/sensitization to death upon HCMV infection (that is infection-dependent phenotypes; x-axis), highlighting a manually selected set of hit genes. Note that due to the experimental design of the screen, the apparent gene essentiality phenotypes are underestimating the real essentiality because t_0 refers to the beginning of HCMV infection, not lentiviral delivery of the sgRNA library. **b**, Direct comparison of CRISPRi and CRISPRn phenotypes, highlighting select host targets represented in both libraries. Hits involved in viral adhesion and entry are more pronounced in the CRISPRn screen. Cullin/RING pathway members and some vesicle trafficking factors were only resolved in the CRISPRi screen. See Supplemental Table 2 for a systematic comparison.
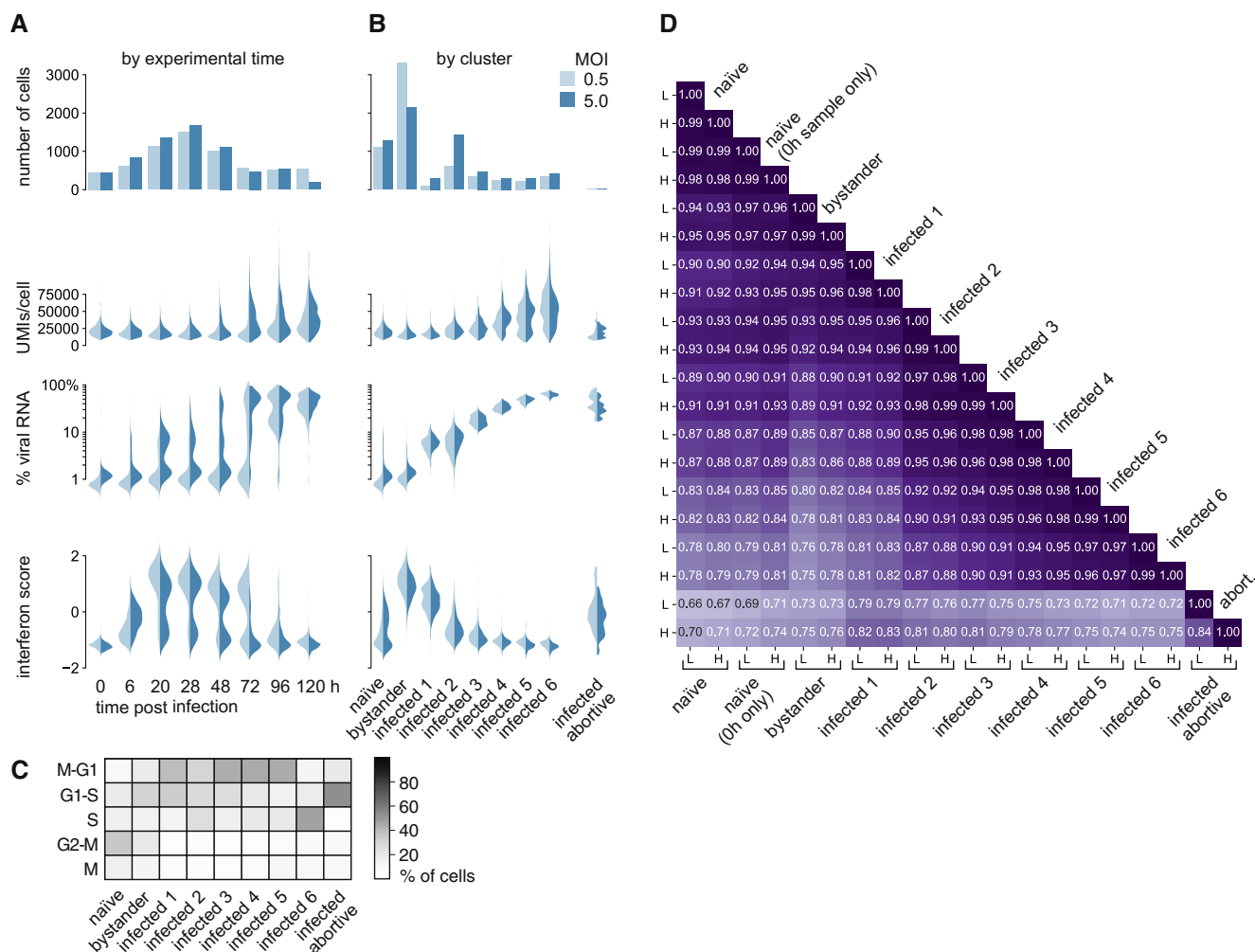
Extended Data Fig. 3 | Single-cell infection time-course. a, b, Numbers of cells, as well as distributions of UMIs per cell, percentage of viral transcripts per cell, and interferon score, broken down by cells for each MOI and each experimental time point (**a**), and for cells for each MOI and cluster membership (**b**). **c**, Distribution of cells across cell cycle phases in each of the clusters. **d**, Pearson's correlation matrix of gene expression values (average logarithmized, scaled UMIs per gene per cell) for all clusters, broken down by low (L) and high (H) MOI conditions.
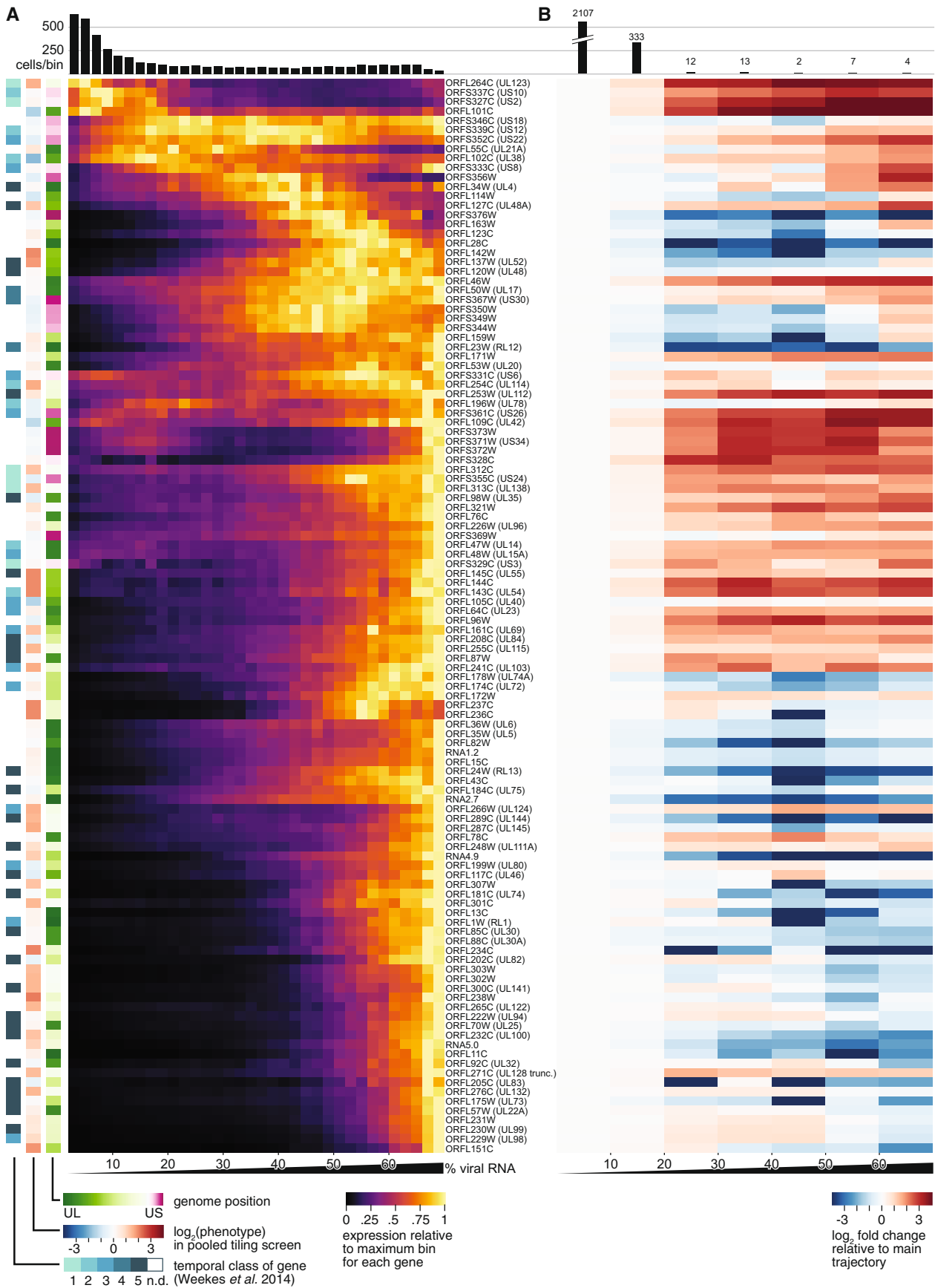
Extended Data Fig. 4 | See next page for caption.

Extended Data Fig. 4 | Viral gene expression profiles along the productive and abortive trajectory of infection. a, Gene expression profiles for robustly detected viral genes along the dominant trajectory (clusters 'infected 1–6'). Cells were grouped in bins spanning 2% of viral RNA and the gene expression (scaled UMIs per gene per cell) averaged for all cells in each bin. The heatmap shows the expression relative to the highest bin. Individual viral genes are ordered by similarity of the profiles, and annotated by genome position, phenotype of cutting within the body of a gene in the pooled virus-directed CRISPR screen (see Fig. 1 and Supplementary Fig. 1), and by the temporal profile as determined in a bulk proteomics study[5]. Note the relationship between a gene's temporal class and its phenotype in the pooled screen: True-late and leaky-late genes predominantly showed protective phenotypes, whereas earlier classes also contained sensitizing genes. **b**, Gene expression profiles of viral genes along the abortive trajectory (clusters 'infected 1–2' and 'infected abortive'). Cells were grouped in bins spanning 10% of viral RNA and the gene expression averaged for all cells in each bin. The heatmap shows the expression relative to the expression in an equivalent bin of the dominant trajectory.
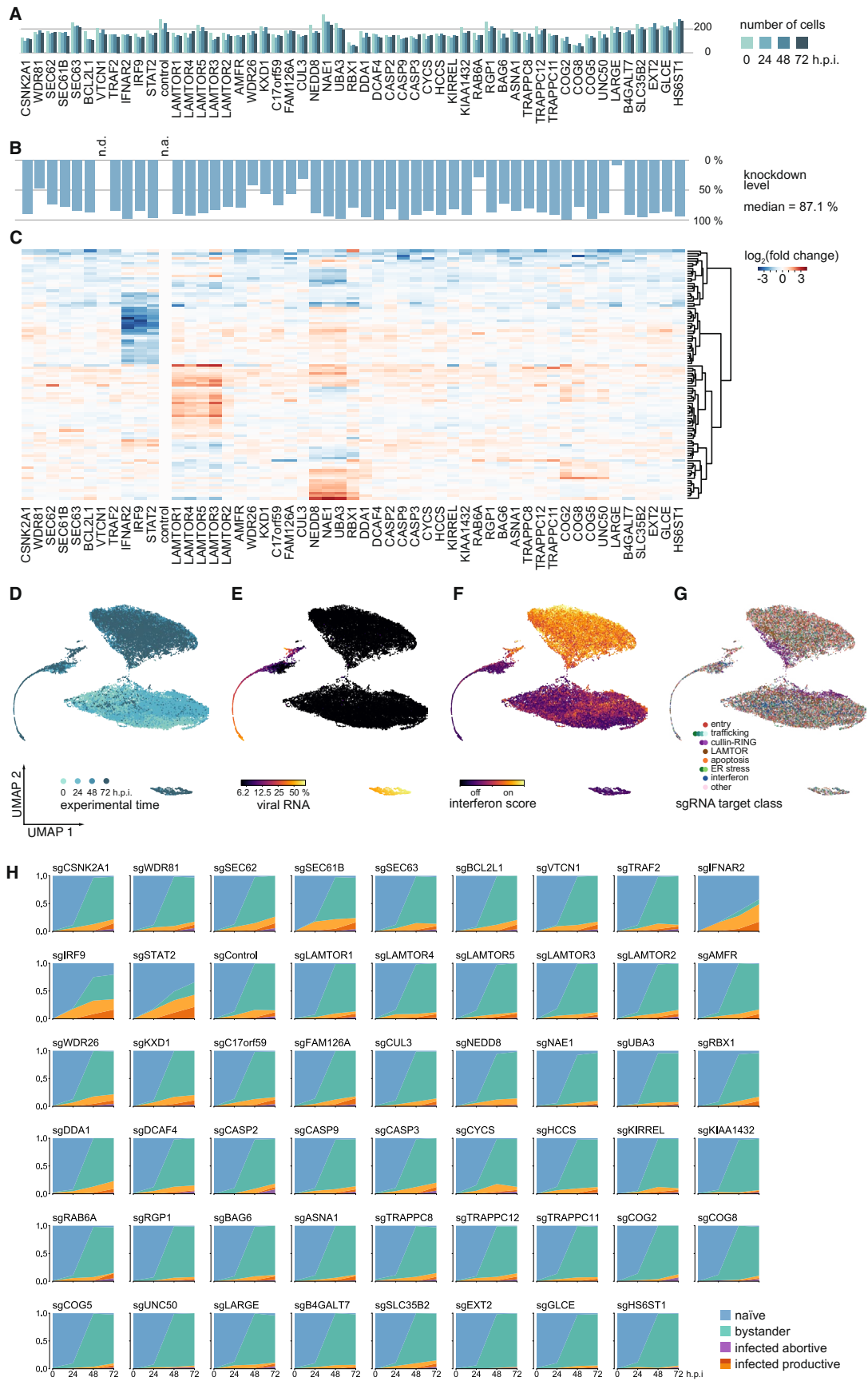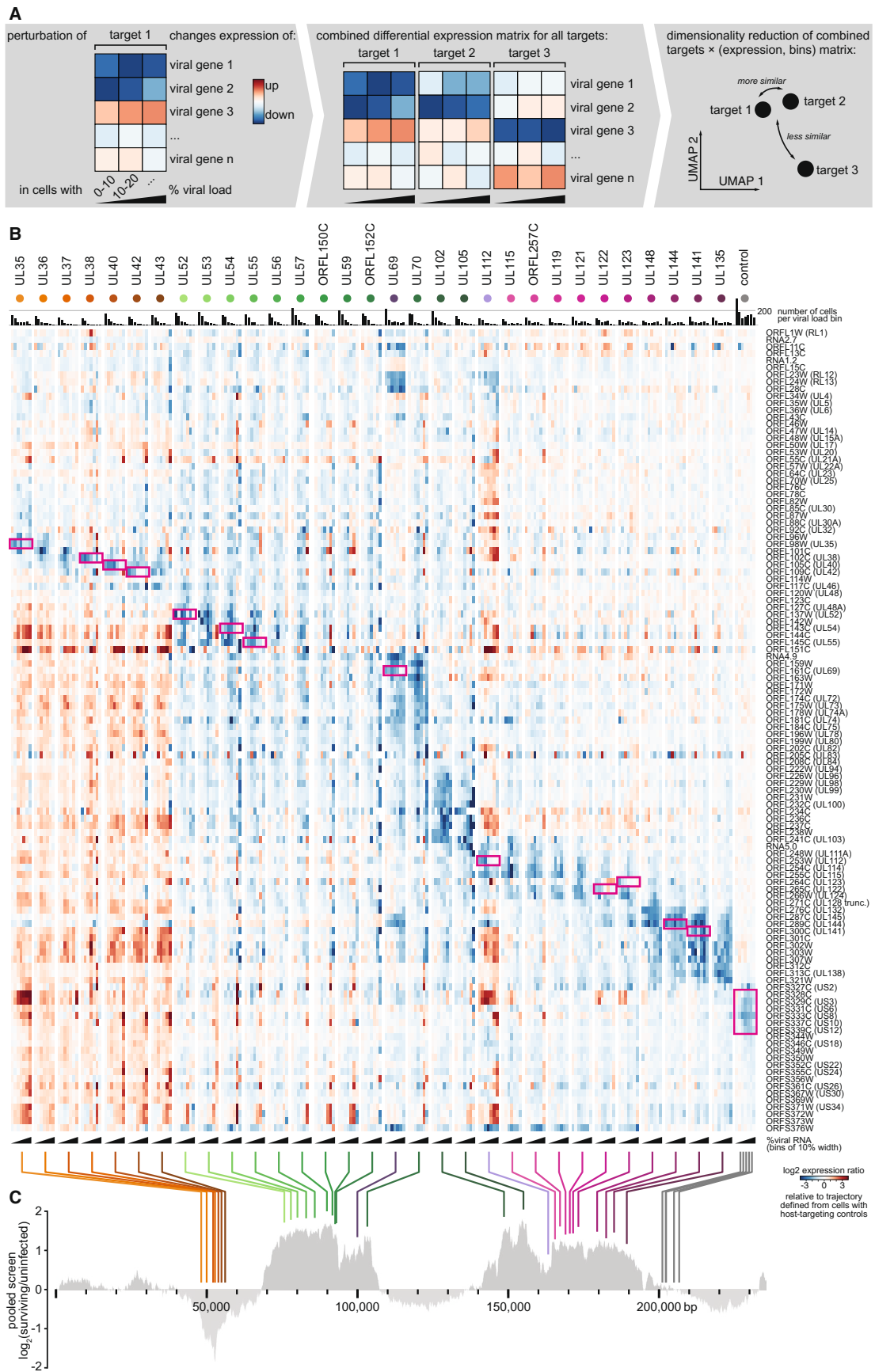
Extended Data Fig. 5 | See next page for caption.

Extended Data Fig. 5 | Host-directed CRISPRi Perturb-seq experiment. a, Numbers of single cells for each sgRNA target for each experimental time point in the host-directed CRISPRi Perturb-seq experiment. The average is 165 ± 50 (mean $\pm$ standard deviation) cells per sgRNA per time point. **b**, Knockdown levels for each sgRNA target calculated from the expression of the target gene in cells with a given sgRNA target relative to cells with control sgRNAs. No transcript at all was detected for VTCN1. Median knockdown level was 87.1%. **c**, Hierarchical clustering of expression changes of the most variable 100 genes (excluding the targeted factors) in response to host factor knockdown in naïve cells, relative to naïve cells with control sgRNAs. **a–g**, UMAP projections of single-cell transcriptomes of cells from the host-directed Perturb-seq experiment (same as in Fig. 4c), color-coded by experimental time post infection (**d**), percentage of viral transcripts per cell (**e**), interferon score, calculated from the normalized expression of interferon stimulated genes (**f**), and by pathway of the targeted host factor in each cell (**g**). **h**, Cluster membership as a function of sgRNA target and time post infection.
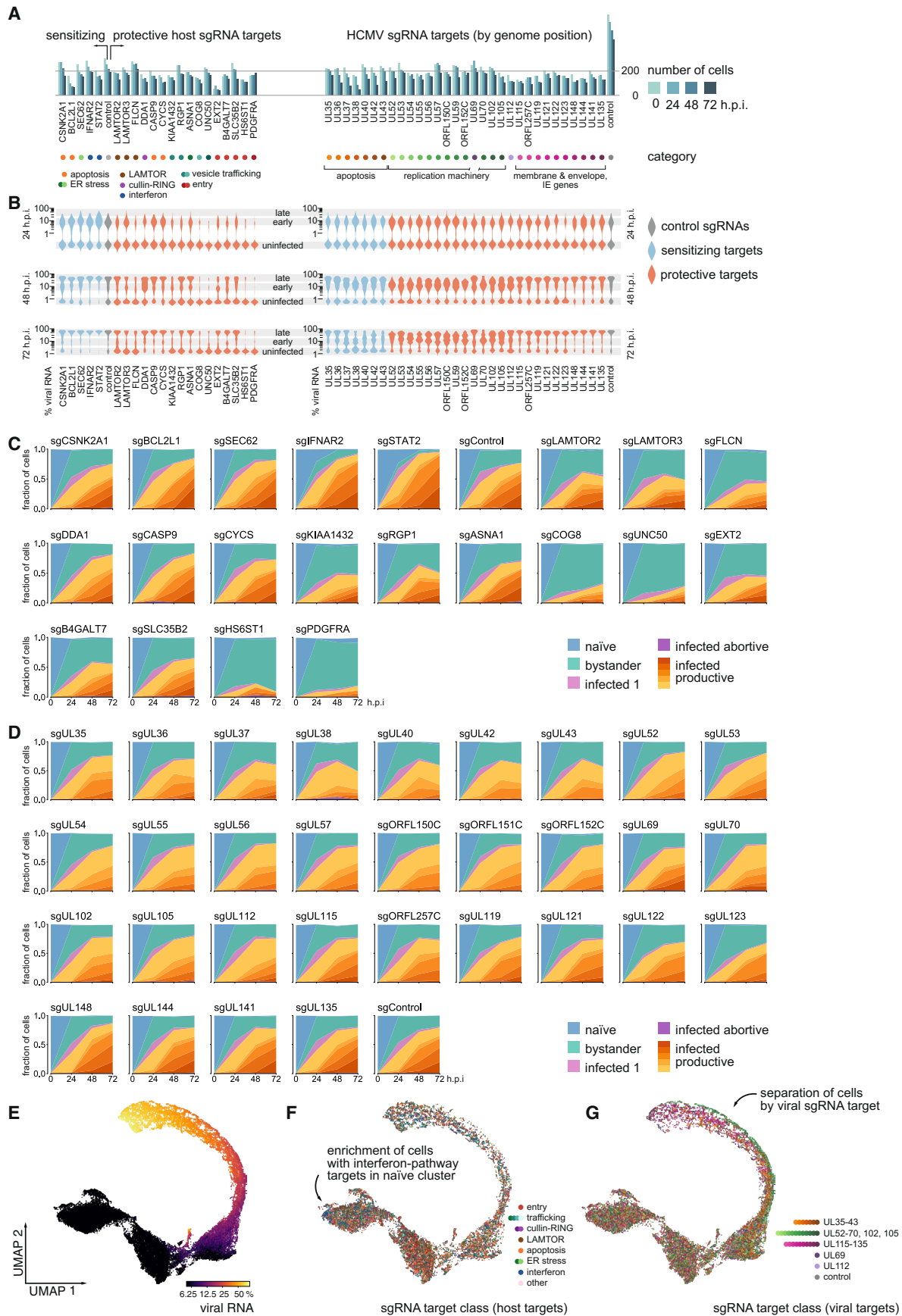
Extended Data Fig. 6 | See next page for caption.

Extended Data Fig. 6 | Host- and virus directed CRISPRn Perturb-seq experiment. a, Numbers of single cells for each sgRNA target for each experimental time point in the host and virus-directed CRISPRn Perturb-seq experiment. The average is 188 ± 77 (mean $\pm$ standard deviation) cells per sgRNA per time point. Note the over-proportional drop in numbers in late time points of cells with apoptosis-related sgRNA targets. 'Control' denotes all safe-targeting sgRNAs, which are 4 and 5 distinct sgRNAs targeting the host and virus, respectively. **b**, Violin plots of the distribution of viral RNA fraction per cell as a function of time post infection and the sgRNA target (red, protective phenotype; blue, sensitizing phenotype; grey, control). Regions of the violin plot corresponding to uninfected cells, as well as early and late stages of infection are highlighted. Note that uninfected cells have non-zero background amounts of viral RNA, and those background levels are higher in later time points, indicating leaking of viral RNA from dying cells. **c, d**, Cluster membership as a function of sgRNA target and time post infection for cells with host-targeting sgRNAS (**c**) and virus-targeting sgRNAs (**d**). **e–g**, UMAP projections of single-cell transcriptomes of cells from the host and virus-directed Perturb-seq experiment (same as in Fig. 5c), color-coded by percentage of viral transcripts per cell (**e**), by pathway of the targeted host factor in each cell (**f**), and by viral target in each cell (**g**) for cells with host and viral targets, respectively.

Extended Data Fig. 7 | See next page for caption.

Extended Data Fig. 7 | Trajectories in viral gene expression space upon perturbation of viral factors. a, Cartoon explaining the analytical workflow for comparing viral trajectories across the different sgRNA targets. **b**, Heatmaps of viral gene expression for all cells with virus-targeting sgRNAs, corresponding to the middle panel of the workflow cartoon. For each sgRNA target, cells were grouped in bins of 10% of viral RNA fraction, and the expression of viral genes plotted relative to a corresponding bin defined by cells with host-directed, safe-targeting sgRNAs (similar to Supplementary Fig. 4b), representing the unperturbed, dominant trajectory. Both the columns (viral sgRNA targets) as well as the rows (expressed viral genes) are ordered by genome position. This facilitates the distinction of gene expression effects in cis, that is the immediate effect of cutting on genes adjacent to the cut site, as opposed to in trans, which are reflecting an altered trajectory of infection. Pink boxes indicate the sgRNA target genes. **c**, Mapping the sgRNA targets onto the phenotypic landscape of the HCMV genome, indicating genome position and phenotype in the CRISPRn tiling screen.

nature research

Corresponding author(s):	Jonathan S. Weissman
	Marco Y. Hein
Last updated by author(s):	Aug 6, 2021

Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see our Editorial Policies and the Editorial Policy Checklist.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a	Confirmed	
☐	☒	The exact sample size (n) for each experimental group/condition, given as a discrete number and unit of measurement
☐	☒	A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☒	☐	The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☒	☐	A description of all covariates tested
☐	☒	A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☐	☒	A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☒	☐	For null hypothesis testing, the test statistic (e.g. F, t, r) with confidence intervals, effect sizes, degrees of freedom and P value noted *Give P values as exact values whenever suitable.*
☒	☐	For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☒	☐	For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☐	☒	Estimates of effect sizes (e.g. Cohen's d, Pearson's r), indicating how they were calculated

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection	Pooled screen data preprocessing: https://github.com/mhorlbeck/ScreenProcessing Single-cell data preprocessing: Cellranger v.2.0.1 (10x Genomics)
Data analysis	Single-cell data processing: https://github.com/thomasmaxwellnorman/perturbseq_demo using umap (v0.1.1), leidenalg (v0.8.0), hdbscan (v0.8.26)

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:
- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

Raw and processed sequencing data for all experiments are publicly accessible on GEO (accession code: GSE165291).

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences ☐ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	Pooled screens were conducted at 500-1000x coverage (i.e. ~500-1000 cells per library element per time point). Perturb-seq experiments were conducted at ~160-180x coverage (i.e. 160-180 cells per library element per time point). These coverages were selected based on previous studies (Gilbert & Horlbeck et al., 2014; Adamson & Norman et al., 2016).
Data exclusions	For all single-cell experiments, we removed cells that could not be uniquely assigned to one element of the library (i.e. cells that had no sgRNA assigned or that were assigned multiple sgRNAs, indicating doublets or cells that got infected with two or more separate lentiviruses). A few library elements were excluded during data analysis because these sgRNAs were either inactive or not informative (see explanation in the Methods section).
Replication	The pooled CRISPRi host screen and pooled CRISPRn virus-targeting screens were performed as two biological replicates, which were well correlated. Hits from the CRISPRi host-directed screen were validated by one replicate of a CRISPRn screen. For the unperturbed single-cell time course, we infected cells separately with different MOIs for the same time points and confirmed excellent reproducibility of the transcriptomes of cells in similar states of infection across those samples (see Extended Data Fig. 3F). Perturb-seq experiments were done once, and we confirmed reproducibility of the responses of the two separate guides for each target gene.
Randomization	Randomization not relevant to our study. All comparisons were performed against coherent sets of negative control elements of libraries or cells with negative control sgRNAs.
Blinding	No blinding was performed.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
☒	☐ Antibodies
☐	☒ Eukaryotic cell lines
☒	☐ Palaeontology and archaeology
☒	☐ Animals and other organisms
☒	☐ Human research participants
☒	☐ Clinical data
☒	☐ Dual use research of concern

Methods

n/a	Involved in the study
☒	☐ ChIP-seq
☒	☐ Flow cytometry
☒	☐ MRI-based neuroimaging

Eukaryotic cell lines

Policy information about cell lines

Cell line source(s)	Fibroblasts were purchased from the ATCC (CRL-1634)
Authentication	Fibroblasts were not authenticated after obtaining them from the ATCC.
Mycoplasma contamination	Fibroblasts were tested negative for Mycoplasma infection.
Commonly misidentified lines (See ICLAC register)	No commonly misidentified cell lines were used.