nature microbiology

Article

https://doi.org/10.1038/s41564-025-02034-3

Single-cell image-based screens identify host regulators of Ebola virus infection dynamics

Received: 11 March 2024

Accepted: 15 May 2025

Published online: 24 July 2025

Check for updates

Rebecca J. Carlson[1,2,15], J. J. Patten[3,4,15], George Stefanakis[5], Brian Y. Soong[2,13], Adityanarayanan Radhakrishnan[2,6], Avtar Singh[2,14], Naveen Thakur[7], Kathleen C. F. Sheehan[8], Gaya K. Amarasinghe[8], Nir Hacohen[2,9], Christopher F. Basler[7], Daisy W. Leung[10], Caroline Uhler[2,5], Robert A. Davey[3,4,16] ✉ & Paul C. Blainey[2,11,12,16] ✉

Filoviruses such as Ebola virus (EBOV) give rise to frequent epidemics with high case fatality rates while therapeutic options remain limited. Earlier genetic screens aimed to identify potential drug targets for EBOV relied on systems that may not fully recapitulate the virus life cycle. Here we applied an image-based genome-wide CRISPR screen to identify 998 host regulators of EBOV infection in 39,085,093 cells. A deep learning model associated each host factor with a distinct viral replication step. From this we confirmed UQCRB as a post-entry regulator of EBOV RNA replication and show that small-molecule UQCRB inhibition reduced virus infection in vitro. Using a random forest model, we found that perturbations on STRAP (a spliceosome-associated factor) disrupted the equilibrium between viral RNA and protein. STRAP was associated with VP35, a viral RNA processing protein. This genome-wide screen coupled with 12 secondary screens including validation experiments with Sudan and Marburg virus, presents a rich resource for host regulators of virus replication and potential targets for therapeutic intervention.

Filoviruses such as Ebola virus (EBOV), Sudan virus (SUDV) and the distantly related Marburg virus (MARV) are single-stranded, negative-sense viruses responsible for outbreaks with high case fatality rates, predominantly in West or equatorial Africa. While monoclonal antibody therapy has demonstrated partial efficacy, and a recently FDA-approved vaccine resulted in EBOV-specific protection, gaps in therapy remain and there are no approved treatments against other filoviruses such as MARV or SUDV.

EBOV is taken up into cells via macropinocytosis and then trafficked through endolysosomes, where the viral glycoprotein (GP) binds to the host receptor NPC1, enabling cytoplasmic release[1]. Transcription from the negative stranded virus genome produces viral mRNAs and, in turn, production of viral proteins leads to formation of inclusion bodies (IBs), cytoplasmic foci that serve as sites for viral RNA synthesis[1–3]. At later stages of infection, these proteins exhibit a diffuse cytoplasmic localization pattern and, finally, localize to the cell periphery during

[1]Massachusetts Institute of Technology, Department of Health Sciences and Technology, Cambridge, MA, USA. [2]Broad Institute of MIT and Harvard, Cambridge, MA, USA. [3]Department of Virology, Immunology, and Microbiology, Boston University School of Medicine, Boston, MA, USA. [4]National Emerging Infectious Diseases Laboratories, Boston University, Boston, MA, USA. [5]Laboratory for Information and Decision Systems, Massachusetts Institute of Technology, Cambridge, MA, USA. [6]Harvard School of Engineering and Applied Sciences, Cambridge, MA, USA. [7]Department of Microbiology, Icahn School of Medicine at Mount Sinai, New York, NY, USA. [8]Department of Pathology and immunology, Washington University School of Medicine, St Louis, MO, USA. [9]Massachusetts General Hospital Cancer Center, Boston, MA, USA. [10]Department of Medicine, Washington University School of Medicine, St Louis, MO, USA. [11]Department of Biological Engineering, Massachusetts Institute of Technology, Cambridge, MA, USA. [12]Koch Institute for Integrative Cancer Research, MIT, Cambridge, MA, USA. [13]Present address: Icahn School of Medicine at Mount Sinai, New York, NY, USA. [14]Present address: Department of Cell and Tissue Genomics, Genentech, San Francisco, CA, USA. [15]These authors contributed equally: Rebecca J. Carlson, J. J. Patten. [16]These authors jointly supervised this work: Robert A. Davey, Paul C. Blainey. ✉e-mail: radavey@bu.edu; pblainey@broadinstitute.org

virus budding[3]. These distinct stages of infection are not readily distinguishable by simple intensity-based measurements.

Genetic screens are high-throughput methods that enable identification of host targets that modulate viral infection. However, due to the challenges of screening high-consequence viruses, previous genetic screens for EBOV often relied on the use of pseudotyped viruses[4–6] or on reverse genetics-based life cycle modelling systems to recapitulate viral RNA synthesis and protein expression[7]. The former precludes identification of post-entry viral modulators, while the latter may not fully recapitulate the life cycle of authentic virus. Here we set out to assess cell survival-independent phenotypes, allowing for recovery of host factors affecting any part of the viral replication cycle.

Our optical pooled screening (OPS)[8] approach enables image-based pooled genetic screens in tens of millions of cells[9,10]. OPS couples high-resolution images of single cells with a targeted in situ sequencing readout of each cell's specific genetic perturbation identity. Here we present presumably the first genome-wide multiparametric genetic screen for EBOV, evaluating viral protein and RNA synthesis as markers of infection to characterize how hundreds of host genes influence distinct aspects of EBOV replication. We apply machine learning approaches to this dataset of nearly 40 million cell images to identify regulators of distinct stages of the EBOV life cycle, from cell entry, to inclusion body formation and viral RNA transcription and replication. These include deep learning architectures such as an autoencoder to encode unsupervised image information as well as supervised approaches. We confirmed hits via secondary screening in two cell lines with EBOV and across two distantly related filoviruses (SUDV and MARV) to validate and contextualize our results. UQCRB (a complex III subunit) and STRAP (Serine/Threonine Kinase Receptor Associated Protein), proteins that are respectively involved in the mitochondrial respiratory chain and the spliceosome, were selected for further mechanistic analysis of their roles in viral replication.

Results

An image-based screen identifies EBOV replication regulators

Here we applied OPS to define the contributions of host factors to distinct steps of the EBOV replication cycle at the genome-wide scale; we obtained images of nearly 40 million individual HeLa-TetR-Cas9 cells infected with EBOV and transduced with a pool of ~80,000 single-guide RNAs (sgRNAs) targeting ~20,000 genes, including 454 non-targeting control sgRNAs (ntgRNAs, Fig. 1a). Cells were infected with EBOV at a multiplicity of infection (MOI) of ~3 and assayed for EBOV VP35 protein and the EBOV VP35 positive-sense RNA, representing mainly viral mRNA[11], using hybridization chain reaction fluorescence in situ hybridization (HCR-FISH) optimized for the first time as part of the OPS workflow (Extended Data Fig. 1 and Methods). In addition to measuring viral protein and RNA, we stained for the host transcription factor c-Jun (whose activity is increased in EBOV- and MARV-infected cells)[12,13], as well as LAMP1 (a lysosomal protein), vimentin, and a nuclear stain (DAPI) (Fig. 1b). c-Jun's mechanistic role in EBOV infection is not well established but others have noted upregulation of c-Jun and cellular mRNA responsive to c-Jun binding upon EBOV and MARV infection[12,13].

We identified 998 genetic knockouts (KO) that significantly impact EBOV infection (false discovery rate (FDR)-adjusted $P < 1 \times 10^{-4}$) as measured by a VP35 protein intensity statistic based on the difference in cumulative area under the curve (ΔAUC) between cumulative per-cell VP35 median intensity for each sgRNA and ntgRNAs as previously described[8] (Fig. 1c–e and Supplementary Table 1). Aside from the entry receptor NPC1, other known factors were identified as top hits, including all six members of the HOPS complex and cathepsins[5,6,14,15], in addition to CAD (a pyrimidine biosynthesis factor that was previously identified in a screen using a synthetic genome replication system[7]) and PIK3C3 (a recently reported factor, involved in autophagy and membrane trafficking[16]) (Extended Data Fig. 1c). In addition, many previously unidentified factors scored, including TIMM10, and entire protein complexes including the COG, GET, GARP and CIA complexes (Fig. 1f). We also identified 57 negative regulators, none of which were previously shown to affect EBOV infection. These included chaperones such as HSP90B1, MESDC2 and UNC45A (UNC45A interacts with VP30 (ref. 17)), integrin subtypes, mRNA deadenylases and SAGA complex members[6] (Fig. 1f).

Next we identified genes that differentially regulated VP35 RNA or c-Jun nuclear translocation relative to VP35 protein levels. This identified factors that, when targeted, resulted in uncoupling of these typically correlated features and might regulate viral replication through previously unidentified mechanisms of action. We specified features from the VP35 RNA and protein channels using CellProfiler-derived image features[18] as previously described[10]. We then trained two random forest regression models (see Methods) to predict c-Jun nuclear intensity from VP35 RNA and protein features (Fig. 1g,h) or VP35 RNA levels predicted from VP35 protein staining intensities (Fig. 1i and Supplementary Table 2). Factors where c-Jun levels were poorly predicted by viral VP35 and RNA levels included c-Jun itself as well as known MAP kinase pathway members (Fig. 1h). The gene associated with the strongest reduction in VP35 RNA relative to VP35 protein levels was the mRNA binding protein PABPC1 (Supplementary Table 2). A number of other genes decreased EBOV VP35 RNA relative to protein levels (Fig. 1j), while only two genes (*SRPK1* and *STRAP*) had the opposite effect, showing reduced VP35 protein relative to RNA levels (Fig. 1i) which suggests disruption of mRNA translation.

A deep neural network identifies regulators of EBOV localization

While overall VP35 protein and RNA levels were informative, the single-cell images contain more information about virus replication. Due to the complexity of these localization patterns, we sought to identify regulators of EBOV infection dynamics using a deep learning model. Autoencoders are classical deep learning models used to generate informative representations of data in an unsupervised fashion[19]. We used a convolutional autoencoder to reconstruct all six-channel segmented masked images (the unsupervised model in Fig. 2a and Extended Data Fig. 4a,b). We then fine-tuned this model with a supervised objective function to predict four human-labelled categories of EBOV VP35 protein subcellular localization: (1) faint, indicative of poorly or uninfected cells, (2) punctate, representing cells with viral inclusion bodies, (3) cytoplasmic, representing a later stage of infection with diffuse protein localization and (4) peripheral, representing viral budding. These were manually annotated for over 3,000 cells for model training. PHATE[20] was then used to visualize the 2,048-dimensional embedding of cells obtained from the unsupervised and fine-tuned autoencoders. While the unsupervised cell embeddings showed a lack of separation of the four phenotypic classes, the fine-tuned embeddings showed clear separation among the four classes (Fig. 2b). An ordinal chi-square test identified genes that significantly altered the proportion of cells in each of the four phenotypic classes (Fig. 2c and Supplementary Table 3).

While the most significant genes were the same as those identified by the simpler intensity metric, the fine-tuned autoencoder model allowed separation of cells that would otherwise have similar protein staining intensities but instead had distinct faint or punctate VP35 protein localization patterns (Fig. 2d). Genes with a high proportion of faint cells included many known entry or early infection regulators, as well as genes identified as regulators of infection such as PIK3C3 and GET complex members (Fig. 2d,e). In contrast, many of the genes associated with a significant increase in punctate VP35 localization, probably reflecting inhibition of the replication cycle after initial infection, have not been previously linked to EBOV infection (Fig. 2e, green). Key enzymes in de novo pyrimidine biosynthesis previously shown to positively regulate EBOV replication[7,21] were significantly

https://doi.org/10.1038/s41564-025-02034-3

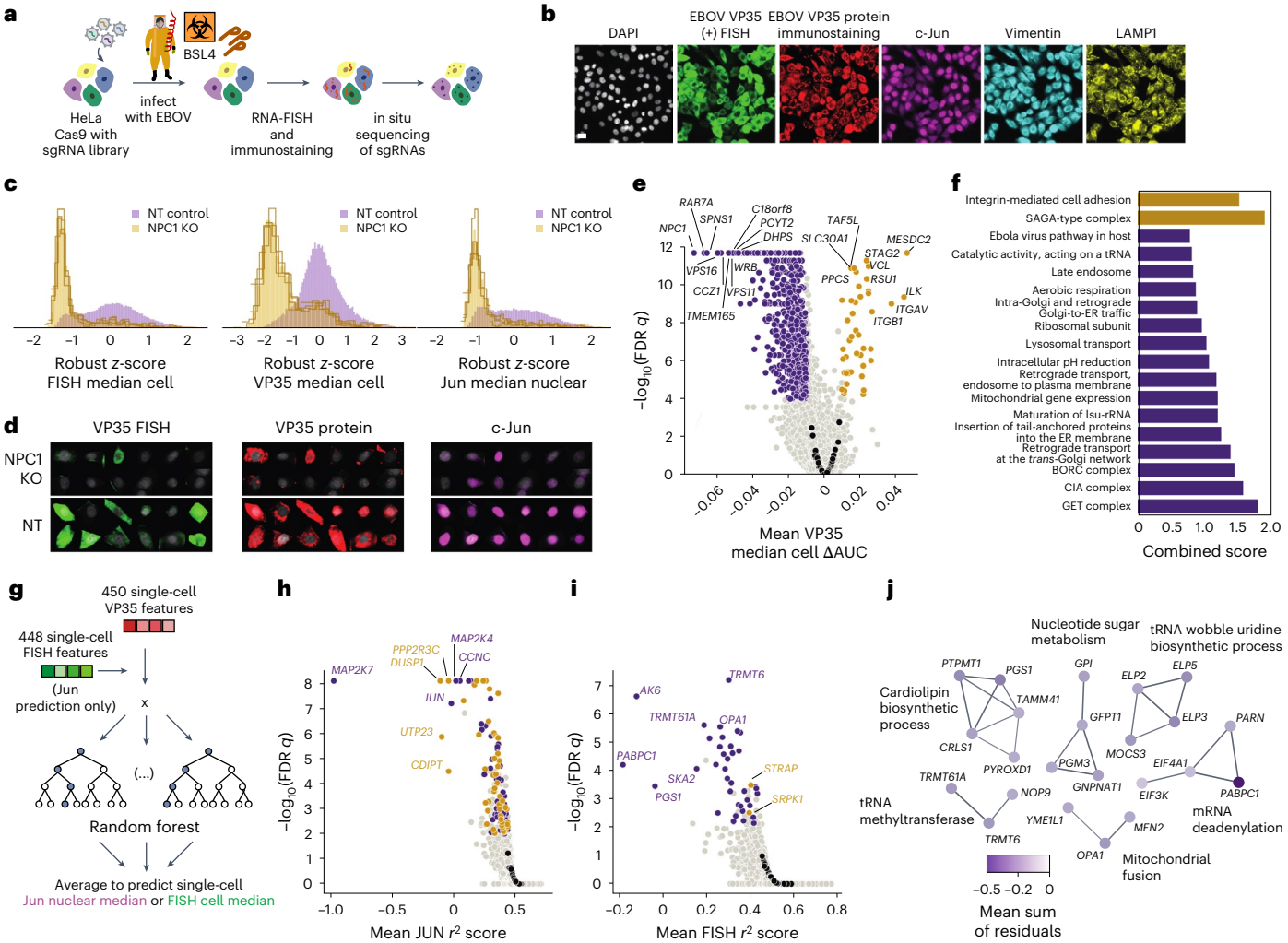


Fig. 1 | Genome-wide optical pooled screening reveals regulators of multiple responses to Ebola virus infection. a, Workflow for genome-wide optical pooled screen. **b**, Example images (random subselection from tens of thousands of screen FOVs) of infected cells assayed across six distinct markers in the pooled screen. Scale bar, 20 µm. **c**, Histograms of VP35 RNA FISH, VP35 protein and c-Jun transcription factor intensity levels in non-targeting or NPC1 KO cells from the screen; each histogram trace represents a distinct sgRNA targeting NPC1. **d**, Randomly selected cells transduced with a non-targeting sgRNA or NPC1-targeting sgRNA show reduced levels of VP35 RNA (FISH), VP35 protein and c-Jun in NPC1 knockout cells relative to non-targeting cells. Cells were masked to highlight the cell of interest using seeded watershed segmentation. **e**, Volcano plot of the per-cell median VP35 protein intensity ΔAUC between each gene and non-targeting control cells; black points represent distinct non-targeting control sgRNAs. Purple and gold points indicate top genes that decreased or increased EBOV levels, respectively, while grey points are additional screen datapoints. **f**, Enrichr gene ontology analysis of top terms significantly enriched in genes that showed reduced VP35 protein intensity upon knockout (purple) or increased VP35 protein intensity (gold). **g**, Workflow for random forest regression model trained to predict either c-Jun nuclear intensity from VP35 RNA and protein features or VP35 RNA intensity from VP35 protein features. **h**, Volcano plot of random forest regression model coefficients of determination for the c-Jun prediction task; black points represent distinct non-targeting control sgRNAs; genes in purple had a negative mean sum of residuals, indicating decreased c-Jun relative to model prediction, while genes in gold had a positive mean sum of residuals. **i**, Volcano plot of random forest regression model coefficients of determination for the VP35 RNA prediction task; black points represent distinct non-targeting control sgRNAs. **j**, STRING analysis of genes that had a negative mean sum of residuals for the VP35 RNA FISH prediction task; purple shade denotes magnitude of mean sum of residuals, indicating the amount that EBOV VP35 positive-sense RNA was decreased relative to protein levels. Edge thickness corresponds to confidence score; only interactions with a confidence score ≥0.7 in the full STRING v.11.5 were considered.

enriched for the punctate phenotype, as well as genes involved in purine biosynthesis not previously associated with EBOV replication. Knockdown of Elongator complex genes (ELP2-6) also led to an increase in punctate VP35 (Fig. 2e).

Another major category of genes that elevated VP35 puncta when knocked out were mitochondria-related genes. Specifically, knockouts of UQCRB and HARS2 (a mitochondrial tRNA synthetase) (Fig. 2f,g), resulted in the most enhanced punctate phenotype. Mitochondrial function has not been directly linked to EBOV infection; however, EBOV VP30 and VP35 proteins physically interact with the mitochondrial ribosome and inner membrane components[22], and oxidative phosphorylation and expression of mitochondrial translation genes were found to increase upon EBOV infection[23]. To further investigate the effect of UQCRB inhibition on EBOV infection, we treated EBOV-infected HeLa cells with terpestacin, a small-molecule inhibitor of UQCRB[24]; we observed clear reductions in EBOV infection (Fig. 2h) with little effect on cell viability as measured by nuclear count at the treatment half-maximal inhibitory concentration (IC_{50}) of 5 µM, highlighting an immediate practical application of results from our deep learning approach.

Profile clustering reveals relationships between EBOV modulators

We next more closely examined the 998 genes that significantly altered overall virus infection as measured by VP35 protein or viral RNA levels

https://doi.org/10.1038/s41564-025-02034-3

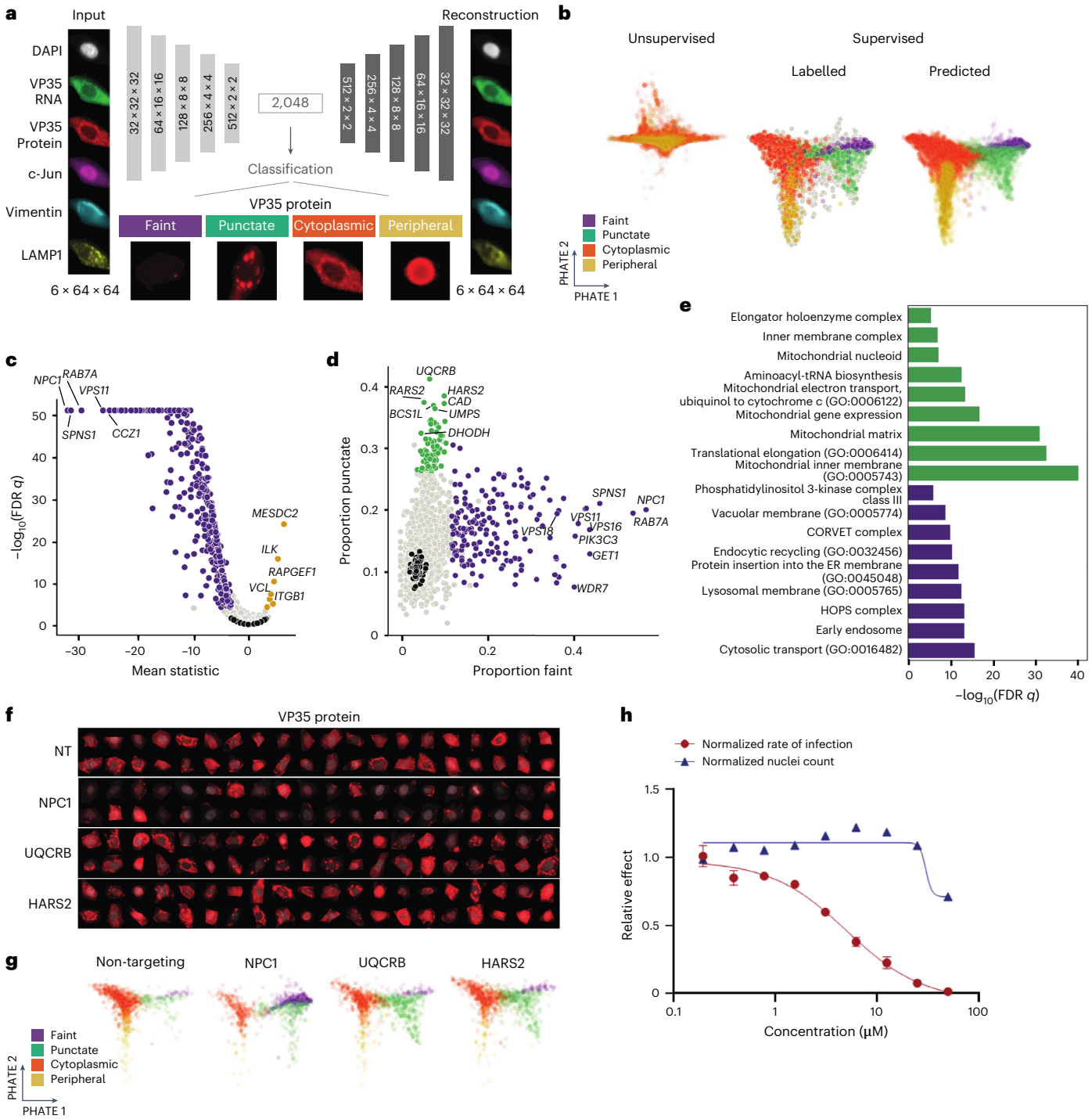


Fig. 2 | A deep neural network model reveals regulators of Ebola virus VP35 protein subcellular localization. a, Architecture of the neural network model, trained first as an autoencoder with a latent space of 2,048 dimensions to reconstruct input cell images (example input cell and autoencoder reconstruction shown) and subsequently fine-tuned to classify cells on the basis of the manually annotated four classes of VP35 protein localization, with one example image shown for each class. **b**, PHATE visualization of the single-cell autoencoder embedding (2,048 dimensions per cell) obtained from the two deep learning modes: fully unsupervised autoencoder model (left) coloured by the predicted class labels, and fine-tuned autoencoder model using the supervised objective function (right) showing both the hand-labelled training data as well as the predicted labels. **c**, Volcano plot of per-gene mean ordinal chi-square statistics and FDR-corrected *P* values. Black points represent distinct non-targeting sgRNAs; points in yellow indicate perturbations with a significant increase in the mean chi-square statistic; and points in purple indicate perturbations with a significant decrease in the statistic. **d**, Scatterplot of the proportion of faint versus punctate VP35 staining patterns of cells for each gene; black points denote individual non-targeting sgRNAs. **e**, Significantly enriched gene ontology terms for sets of genes with a high proportion of faint cells (purple) or punctate cells (green). **f**, Images of VP35 protein expression (red) overlaid with nuclear mask (grey) in randomly selected cells with sgRNAs targeting the indicated genes. **g**, Fine-tuned autoencoder embedding visualized using PHATE for cells with indicated gene knockouts; colours indicate the class labels predicted by the neural network. **h**, Ebola virus infection and cell survival (normalized nuclei count relative to untreated) after 48 h in HeLa cells treated with terpestacin. Error bars indicate s.d.

(Supplementary Table 1). For each autoencoder model, we extracted embeddings for individual cells and calculated cumulative ΔAUCs for each embedding feature between cells with a given sgRNA and non-targeting control cells (Supplementary Tables 5 and 6). We then clustered the resulting embeddings to identify genes with similar effects using the Leiden algorithm (Fig. 3a,c and Methods)[25], determining the optimal cluster resolution by assessing the Adjusted Rand Score between subsets of the data (Extended Data Fig. 4a). While the clusters resulting from the two autoencoder models were similar (Fig. 3b), the fully unsupervised autoencoder model resulted in a larger number of clusters, which were significantly enriched for at least one ontology term using Enrichr (23 versus 7). Single-cell images for genetic knockouts from distinct clusters are shown in Fig. 3d. Some genes that scored significantly for increased VP35 protein expression, such as integrins ITGAV (Fig. 3d) and ITGB1 (Extended Data Fig. 4b), showed increased intensity in channels not directly related to VP35, such as vimentin and LAMP1, as well as cell rounding. We anticipate that the change in VP35 expression pattern for such genes represents impact on cell morphology or cell health resulting in redistribution of the marker, rather than a change primarily affecting virus replication.

To prioritize genes for further follow-up experimentation, we sought to control observed effects for the level of viral protein expression as a proxy for the level of infection in each cell. We expected that this could highlight genes influencing the typical course of infection because we observed a wide spread of viral protein expression in both control and perturbed cells. Controlling for the degree of infection as a major 'nuisance' variable allows us to identify changes in the cells that differ from changes associated with the nuisance variable. To implement this, we sampled non-targeting control cells whose distribution of VP35 protein median intensities matched the VP35 protein median intensity for each sgRNA in our screen (Fig. 3e), and recomputed ΔAUCs between each sgRNA and the matched non-targeting control cells. Next, we plotted the correlation of ΔAUCs for features from all six imaging channels computed without matching (Supplementary Table 1) against those from the matched condition (Fig. 3f). If the primary phenotypic effect of the genetic perturbation is to alter viral infection levels, we expect that the matched AUCs will be close to 0, as the perturbed cells would resemble non-targeting cells with the same infection levels without further morphological changes. Therefore, Pearson correlation between matched and unmatched ΔAUCs will be modest, since single-cell images of the genetic perturbation of interest closely match images of non-targeting control cells with a similar infection level. Indeed, we see that features for NPC1 KO cells are much more modestly correlated between matched and unmatched conditions, consistent with NPC1's primary effect being on the infection level. In contrast, features for knockout of the integrin ITGAV, which affects global cell morphology rather than viral replication specifically, show higher correlation between the matched and unmatched AUCs, reflecting morphological changes largely independent of infection (Fig. 3f). Using this approach, we calculated Pearson correlations between matched and unmatched conditions for each of the genes in our screen (Fig. 3g) and used this metric to select genes with primary effects on viral replication rather than on cell health for targeted follow-up screening. We illustrate our analytical approaches in Extended Data Fig. 5.

Targeted image-based genetic screens identify cell- and virus-specific filovirus regulators

To evaluate the contributions of cell type, filovirus species and infection timing to outcomes of our screen, we targeted 113 hit genes (manually subselected from hits with VP35 protein intensity FDR-adjusted $P < 1 \times 10^{-4}$, as well as genes with lower impact on cell health (Fig. 3f) to prioritize hits that primarily affect EBOV replication without impairing cell morphology) of interest from our genome-wide screen with 6 sgRNAs per gene and >475,000 cells per condition. Screen conditions varied: (1) timepoints (16 and 24 h), both earlier than the primary screen (28 h) to increase power for identification of negative regulators of viral replication; (2) cell lines, the HeLa cells used in the genome-wide screen and Huh7 cells, a liver cell line relevant to filovirus disease; and (3) distinct filoviruses from the genera *Orthoebolavirus* (Ebola virus Mayinga, Sudan virus Gulu) and *Orthomarburgvirus* (Marburg virus Musoke) (Fig. 4a and Supplementary Table 5) at MOIs of ~3. As in our primary screen, we assayed a filovirus protein (VP35 for EBOV and SUDV; VP40 for MARV) and VP35 RNA via probes designed against the VP35 sequences of each virus, as well as c-Jun nuclear translocation.

Principal component analysis (PCA) of features from non-targeting control sgRNAs from each experimental condition showed separation of outcomes by cell type across PC1 (Fig. 4b). Comparing the cumulative ΔAUCs from the genome-wide screen with the 24 h infection condition from the secondary screen resulted in strong correlations (Pearson $r \geq 0.87$) for VP35 protein and RNA levels (Fig. 4c). False-positive rates for changes in VP35 protein intensity in our secondary screen were only modestly higher than expected and are probably due to technical factors associated with executing a separate screen (Extended Data Fig. 6a). Known regulators such as NPC1 had concordant effects on viral protein across conditions (Fig. 4d,e), as well as positive regulators PIK3C3, GET4 and the GARP complex, and the negative regulator COP1 (Fig. 4d,e).

We next examined virus-specific regulators (Fig. 4f,g), identifying differential dependencies for retriever and retromer genes VPS35L and VPS35 across viruses (Fig. 4f,g). Finally, we investigated cell type-specific regulators (Fig. 4h,i). Many of the cell-type-specific regulators were related to mitochondrial function (BCS1L, PRMT1 and HARS2, among others) (Fig. 4h and Extended Data Fig. 6f), where these positive regulators in HeLa cells did not regulate infection in Huh7 cells, highlighting potential metabolic differences between cell lines.

Characterization of STRAP as a positive regulator of EBOV replication

One hit that stood out across multiple screening analyses was STRAP. STRAP was one of only two proteins identified in the random forest analysis as having increased viral RNA relative to VP35 protein upon knockout (Fig. 1i), and was identified by the fine-tuned autoencoder model as having a significantly different distribution of VP35 protein staining phenotypes compared with non-targeting controls (Fig. 5a). Examination of single-cell intensity histograms from the genome-wide screen revealed that while there was minimal change in the intensity distribution for protein staining, the RNA levels in STRAP KO cells took on a bimodal distribution, with distinct populations of cells having less viral RNA or more viral RNA compared with the NT controls—an observation only possible at the single-cell level and where single-cell-matched perturbation identities were available (Fig. 5b).

To further study the effects of STRAP KO on viral infection, we generated single-cell clones that showed undetectable levels of STRAP by immunoblot. To confirm the phenotype of overall increased RNA in the KOs, cells were incubated with virus at a high MOI of 3 for 72 h. qPCR of cell lysates showed a strong increase in viral RNA in cells (Fig. 5c). By contrast, viral RNA in the supernatant was reduced by 3–8 fold (Fig. 5d) and a 5-fold reduction in infectious titre was observed (Fig. 5e). Images of the infected cell monolayer stained for EBOV protein showed a greater proportion of infected cell foci, indicating that there was limited virus spread, consistent with cells producing fewer overall infectious virus particles (Fig. 5f).

While STRAP KO cells produced fewer infectious virions, the mechanism by which STRAP was modulating viral infection remained unclear. We used a two-step PCR targeting viral genomic RNA (gRNA) to distinguish this from viral mRNA. After challenge at an MOI of 3, EBOV gRNA was significantly reduced in STRAP KO cells (Fig. 5g). Contrasting with the overall 2-fold increase seen in total viral RNA, the decrease in genomic RNA suggested dysregulation of viral RNA transcription.

https://doi.org/10.1038/s41564-025-02034-3

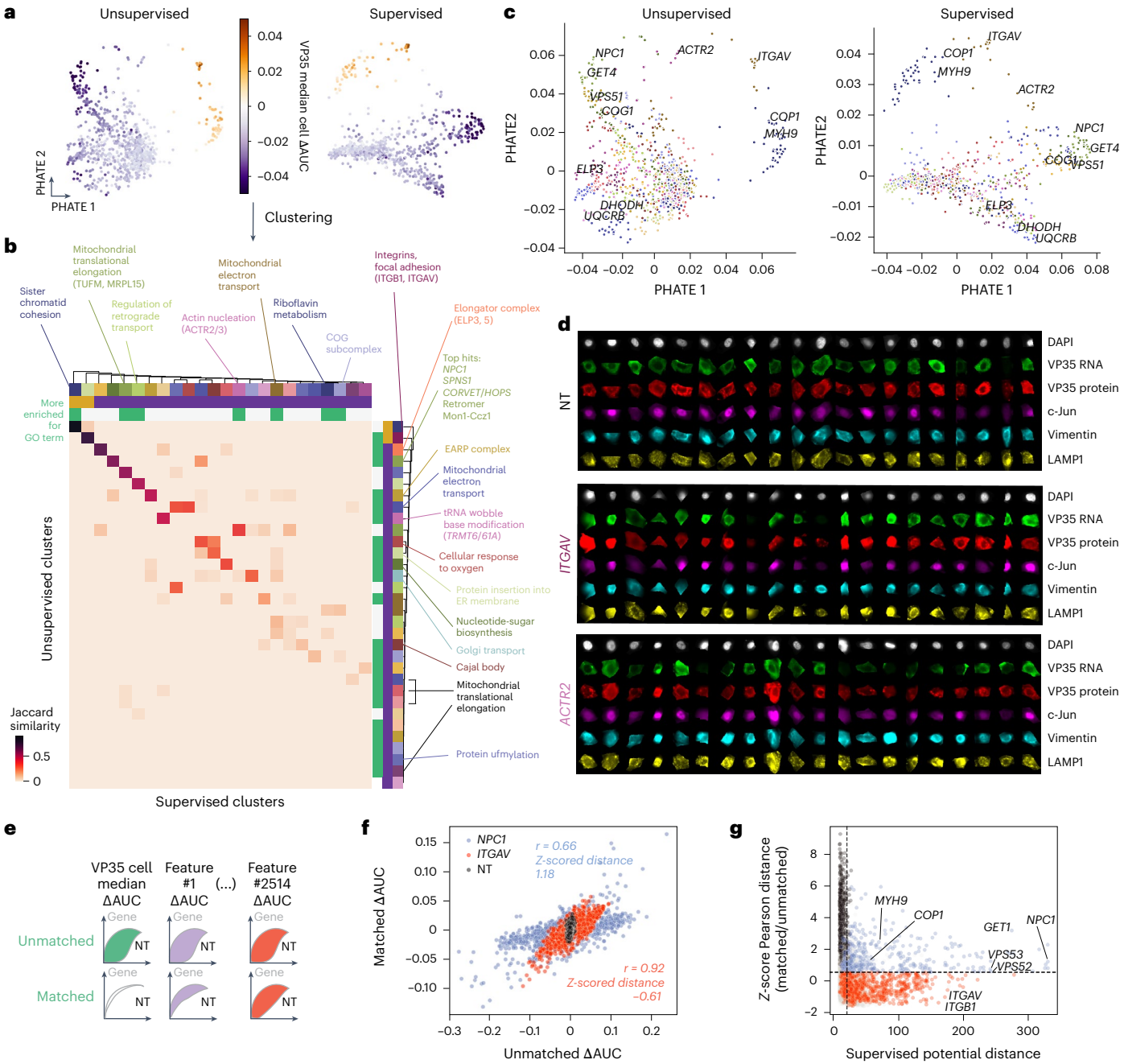


Fig. 3 | Phenotypic profile clustering and matching on infection level reveal relationships between Ebola virus infection modulators. a, PHATE visualization of the embeddings derived from the fully unsupervised and the fine-tuned autoencoder models; each point represents the average embedding of a gene knockout of interest. Points in purple indicate genes that decreased Ebola VP35 protein intensity upon knockout, while points in gold increased intensity. Clustering was performed using the Leiden algorithm. **b**, Heat map of Jaccard similarities between cluster memberships for all clusters with at least one significant GO term identified using the fully unsupervised model as well as the fine-tuned model. Green and white colours indicate binary classification of which clustering approach was more enriched for each GO term, while yellow and purple colours indicate whether the cluster primarily contained genes that increased or decreased virus infection, respectively. **c**, Unsupervised and supervised PHATE visualizations as in **a** coloured by cluster membership with select genes highlighted. **d**, Single-cell images of representative gene knockouts from distinct clusters. **e**, Schematic of recalculation of cumulative ΔAUCs for autoencoder embeddings between each sgRNA and non-targeting control cells after matching on VP35 protein levels in non-targeting controls. **f**, Correlations between ΔAUCs calculated without matching on VP35 protein levels compared to with matching on VP35 protein levels for non-targeting controls, NPC1 knockout and ITGAV knockout cells; each point represents one individual feature. **g**, Correlation between the mean PHATE potential distance from non-targeting controls in the fine-tuned autoencoder embedding (indicative of the EBOV infection phenotype strength) and the *z*-scored Pearson correlation distance between matched and unmatched features from **e**; each point represents a gene of interest. Points in red indicate genes that have abnormal cell health, while points in blue represent hit phenotypes under conditions where overall cell health was less affected. Dashed lines indicate regions beyond the 95th percentile of non-targeting control sgRNAs.

https://doi.org/10.1038/s41564-025-02034-3

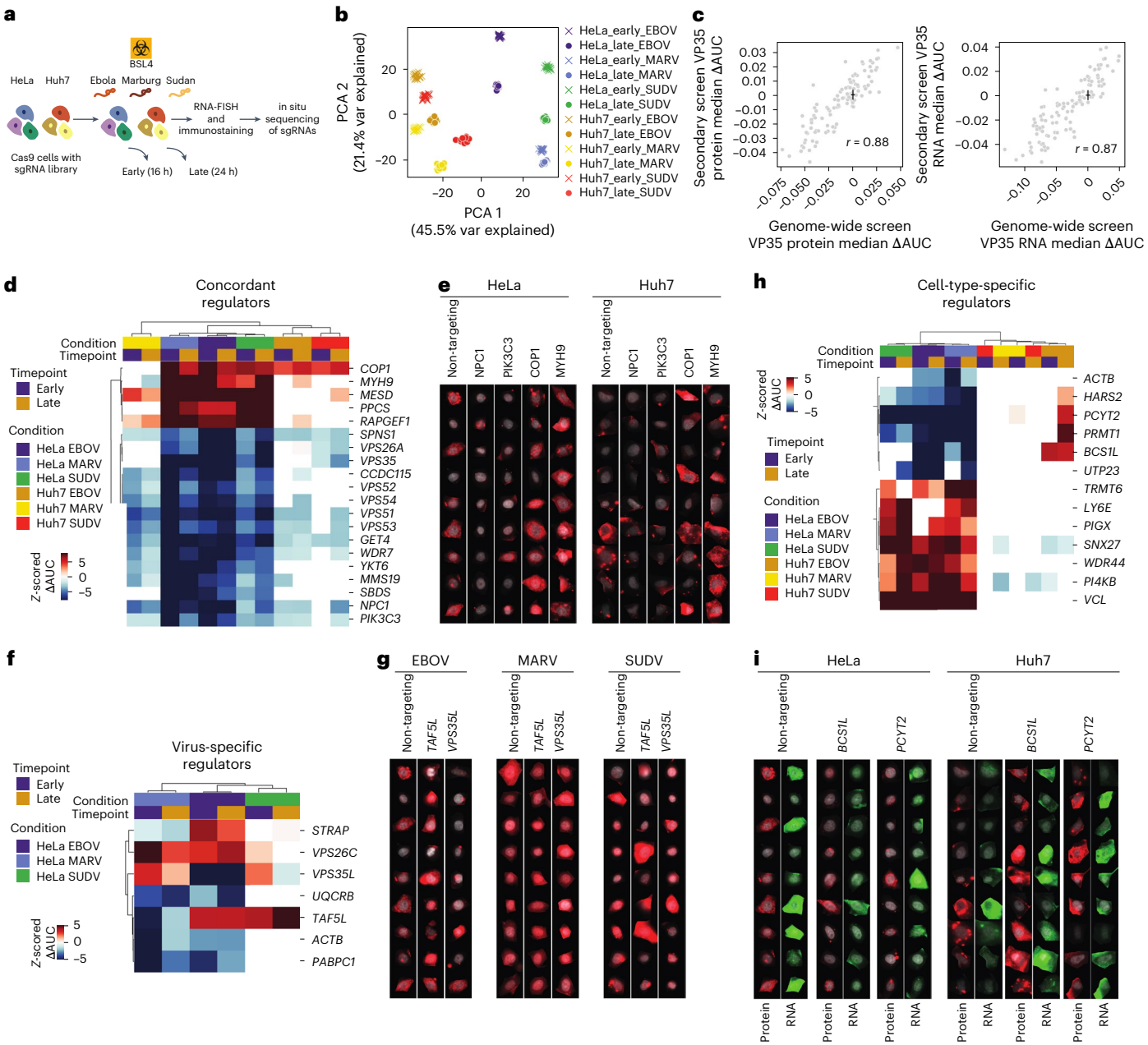


Fig. 4 | Targeted follow-up screens identify concordant and cell- and virus-specific Ebola virus regulators. a, Workflow for targeted secondary screens with 2 cell lines, 3 viruses and 2 timepoints for a total of 12 screening conditions[42]. **b**, PCA of non-targeting control sgRNA phenotypic profiles from each of indicated screening conditions. Early and late indicate 16 and 24 h infection times, respectively. **c**, Correlation between genome-wide and secondary screen VP35 protein and RNA ΔAUCs; black lines indicate standard deviation for non-targeting control sgRNAs in each screen centred around the mean value for non-targeting sgRNAs in the screen. **d**,**f**,**h**, Heat maps showing *z*-scored ΔAUC values for VP35 or VP40 protein. *Z*-scores were calculated on ΔAUC values for all genes in each screen condition relative to means and standard deviations for non-targeting sgRNAs. Hierarchical clustering performed using Pearson correlations. White cells indicate conditions where $P > 0.05$ relative to non-targeting controls in the same condition. **d**, Heat map for genes concordant across screen conditions. **e**, Single-cell images from the secondary screen of select concordant genes (DAPI in grey, VP35 protein in red). **f**, Heat map as in **d** for genes with virus-specific effects. **g**, Single-cell images from the secondary screen of select virus-specific genes (DAPI in grey, VP35 protein for EBOV/SUDV or VP40 protein for MARV in red). **h**, Heat map as in **d** for genes with cell type-specific effects. **i**, Single-cell images from the secondary screen of select cell type-specific genes (DAPI in grey, VP35 protein in red, VP35 RNA in green).

STRAP is commonly found throughout the cytoplasm and nucleus of the cell, with a diffuse immunofluorescence staining pattern. However, upon EBOV infection, STRAP was recruited to the viral inclusion bodies (Fig. 5h) as indicated by STRAP staining directly overlapping VP35, a major component of inclusion bodies. To confirm the close association of STRAP and VP35, we performed a proximity ligation assay (PLA) for STRAP and VP35 (Fig. 5i). The PLA assay yields a fluorescent signal when two proteins are within 40 nm of each other[26], and the intensity and volume of the signal in the cell increase proportionally with a greater number of associative events. PLA with VP35 and STRAP showed signal comparable to that between VP35 and VP30, indicating a similarly close association. Punctate and diffuse phenotypes were also visible, suggesting that the interaction can occur outside of virus inclusion bodies.

https://doi.org/10.1038/s41564-025-02034-3

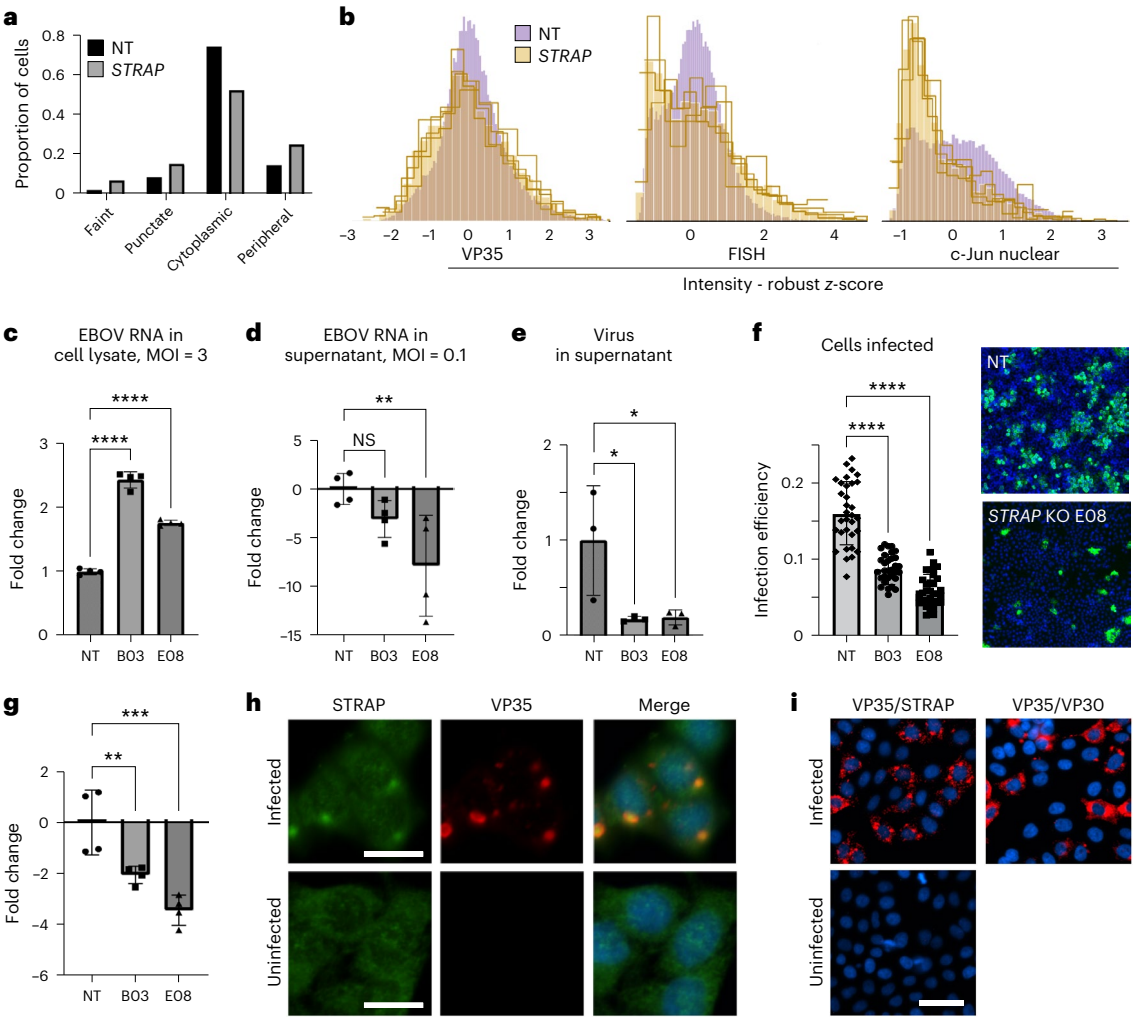


Fig. 5 | STRAP KO impacts virus RNA expression levels and virus infectivity. **a**, Distribution of phenotypes from the primary genome-wide screen as predicted by the fine-tuned autoencoder model for non-targeting controls and STRAP. **b**, Histograms of staining intensity for VP35 protein (ICF), VP35 mRNA by FISH and nuclear c-Jun at single-cell resolution from the primary screen. Each yellow line represents a distinct sgRNA targeting STRAP. **c**, Measurement of total viral RNA in clonal cell lines that were KO for STRAP (B03 or E08) or non-targeting cell line (NT) after challenge with virus at MOI = 3 at 72 h by qPCR. One-way analysis of variance (ANOVA) comparing each clonal line to the non-targeting control, $^{****}P < 0.0001$. **d**, Measurement of total viral RNA in the supernatant after challenge with MOI = 0.1 after 72 h by qPCR. $^{**}P = 0.0086$; NS, not significant. **e**, Measurement of infectious virus in the supernatant after 72 h by f.f.u. assay, $^{*}P < 0.025$. **f**, Measurement of number of cells infected after challenge and allowing 72 h to spread. Left: infection efficiency was calculated as the number of infected cells divided by total nuclei normalized to the non-targeting control, $^{****}P < 0.0001$. Right: representative images of non-targeting and STRAP knockout cells. **g**, Measurements of EBOV negative-sense genomic RNA synthesis in cells after 72 h at MOI = 3. One-way ANOVA was used to compare two different STRAP knockout cell lines to non-targeting controls. $^{***}P = 0.0002$, $^{**}P = 0.0067$. All data are average ± s.d. **h**, Immunostaining for native STRAP protein and EBOV VP35 in infected cells showing accumulation of STRAP within viral inclusion bodies as marked by VP35 staining. Scale bar, 20 µm. **i**, Duolink proximity ligation assay using antibodies against VP35 and STRAP in infected and uninfected cells indicating close association of STRAP with viral proteins within and outside the inclusion bodies. VP35 and VP30 specific antibodies were used on infected cells as a positive control. Scale bar, 50 µm.

Discussion

Genome-wide screens for host factors impacting EBOV infection have mostly been limited to cytotoxicity-based approaches where cells susceptible to infection succumb to cytopathic effects of the virus. While productive, this approach biases toward strong early inhibitors of infection. Here we used image-based pooled screening to directly link genetic knockouts with diverse single-cell phenotypes.

We evaluated our data with three complementary analytical approaches and expect the data to be a resource for future analyses. A first analysis separately examined the intensity of each marker of infection. Unlike previous cell survival screens, we identified negative regulators of filovirus replication, in addition to positive regulators including the GARP and GET complexes and PIK3C3, reported recently[16]. We also validated these effects through 12 secondary screens.

To leverage our high-content imaging data, we applied deep learning approaches. We trained autoencoders directly on single-cell images and fine tuned the resulting model to classify four distinct categories of VP35 protein subcellular localization. This enabled linkage of altered VP35 localization to specific host factors, including UQCRB. A small-molecule inhibitor of UQCRB decreased EBOV infection in cells, presenting an actionable therapeutic hypothesis for EBOV infection therapy. Our fine-tuned autoencoder approach generated several feature embeddings for unsupervised analyses. However, future work focused on a similar supervised classification task could consider using predefined CellProfiler features or a simpler deep learning approach to limit resource utilization in model training, as the fine-tuned autoencoder did not significantly outperform these approaches (Extended Data Fig. 3d).

The third analytical approach we employed was random forest regression analysis to identify host factors that, when knocked out, altered relationships between readouts of infection that are typically correlated. Using this approach, we revealed that knockout of STRAP, a spliceosome regulator controlling cellular distribution of the complex, increased viral positive-sense RNA. Furthermore, STRAP was closely associated with VP35 in inclusion bodies, indicating that it may directly impact timing of negative-strand synthesis. The necessity of STRAP for infectious particle production represents a potential target for therapeutic intervention.

High-content single-cell screens with multidimensional outputs are becoming routine and were demonstrated at genome-scale with a variety of imaging and even transcriptomic readouts[6,27–29]. However, high-content profiling screens of host–pathogen interactions are still limited. Our image-based genetic screens and analytical approaches revealed multiple genes that modulate distinct stages of the EBOV life cycle from viral entry to EBOV RNA replication and the formation of viral inclusion bodies. Small-molecule ligands of potential therapeutic value are known for several of these genes. This work also serves as a general framework for systematically identifying regulators of distinct steps in host–virus interaction dynamics directly from data produced in a single genetic screen, enabling rapid identification of a broader set of potential therapeutic targets for diverse pathogens.

Limitations

Given the complexity and scale of the data derived from our optical pooled screening approach, we implemented a variety of analytical approaches, illustrated in Extended Data Fig. 5, to expand the discovery potential of this study. Despite applying distinct approaches designed to identify a variety of regulators, we note that the analyses presented are not exhaustive and other approaches could identify additional regulators. Thus, we have shared the complete raw image data and processed image features to enable further analysis by the research community. This study dataset, however expansive, cannot report on all aspects of filovirus–host interaction, and further mapping of phenotypic classes representing distinct stages of virus replication and other data analytical techniques will further enhance understanding of important aspects of filovirus biology and associated host responses. While our primary and secondary screens showed high correlation, the false-positive rates in the secondary screen were modestly higher than expected, which may reflect subtle differences in cell types used and culture conditions, and are detailed in Extended Data Fig. 6a.

Methods

Library cloning, lentivirus production and transduction

Libraries were cloned into a CROP-seq-puro-v2 (Addgene, 127458) backbone and lentivirus was then produced and transduced as previously described[29]. Multiplicity of infection for library transductions was estimated by counting colonies following sparse plating and antibiotic selection with puromycin.

Virus infection, phenotyping and in situ sequencing for genome-wide screen

HeLa-TetR-Cas9 clonal cells previously described[8] were used for primary screening. Following transduction, cells were selected with puromycin (1 μg ml^{-1}) for 3 days after transduction and library representation was validated by next-generation sequencing. Cas9 expression was induced with 1 μg ml^{-1} doxycycline for 1 week and cells were then seeded in ten 6-well glass-bottom dishes at 400,000 cells per well 2 days before fixation. Zaire Ebola virus (strain Mayinga) was added at a MOI of ~3 in 2 ml media per well for 28 h before fixation. Cells were fixed by removing media and adding 10% neutral-buffered formalin (Fisher Scientific, LC146705) for >6 h.

Cells were permeabilized with 100% methanol for 20 min; subsequently, the permeabilization solution was exchanged with PBS-T wash buffer (PBS + 0.05% Tween-20) by performing six 50% volume exchanges followed by three quick washes. Cells were incubated for 30 min at 37 °C with homemade probe hybridization buffer (30% formamide, 5× SSC, 0.1% Tween), since dextran sulfate inhibits the reverse transcriptase and polymerase activities necessary for readout of sgRNA sequences by in situ sequencing[18]. Cells were then incubated with primary probes against VP35 positive-sense RNA (purchased from Molecular Instruments and diluted 1:250 in probe hybridization buffer with 1:100 Ribolock) at 37 °C for 4 h. Samples were then washed for 4 × 5 min in probe hybridization buffer at 37 °C, washed 3× in PBS-T and incubated in reverse transcription mix at 37 °C overnight. The reverse transcription mix consisted of 1× RevertAid RT buffer, 250 μM dNTPs, 0.2 mg ml^{-1} BSA, 1 μM RT primer with LNA bases A+CT+CG+GT+GC+CA+CT+TTTTCAA, 0.8 U μl^{-1} Ribolock RNase inhibitor and 4.8 U μl^{-1} RevertAid H minus reverse transcriptase in 750 μl per well. After reverse transcription, cells were washed 5× with PBS-T and post-fixed using 3% paraformaldehyde and 0.1% glutaraldehyde in PBS for 30 min, followed by washing with PBS-T 3 times. Samples were then incubated in homemade HCR-FISH amplification buffer (5× SSC 0.1% Tween) at room temperature for 30 min. Meanwhile, HCR hairpins (Molecular Instruments B1 probes conjugated to Alexa Fluor 488) were separately prepared by heating at 95 °C for 90 s and then cooling to room temperature in the dark for 30 min. Next, samples were incubated with probes diluted 1:125 in probe amplification buffer for 2 h at room temperature. Following incubation, excess hairpins were removed by washing 5× for 5 min with probe amplification buffer. Primary antibodies against VP35 (1:3,200 dilution; developed in-house at Washington University School of Medicine), c-Jun (1:1,800 dilution, Cell Signaling Technology, 9165, RRID:AB_2130165) and vimentin (1:1,300 dilution, Abcam, ab24525, RRID:AB_778824) were added by incubating samples for 3.5 h at 37 °C in 3% BSA (VWR, 97061-422) in PBS. Samples were then washed 3× in PBS-T for 3 min and incubated with secondary antibodies: 1:1,800 donkey anti-mouse antibody (Jackson ImmunoResearch Labs, 715-006-151, RRID:AB_2340762) disulfide-linked to Alexa Fluor 594 (Thermo Fisher, A10270) via a custom conjugation, 1:1,800 donkey anti-rabbit antibody (Jackson ImmunoResearch Labs, 711-006-152, RRID:AB_2340586) disulfide-linked to Alexa Fluor 647 (Thermo Fisher, A10277) via a custom conjugation, and goat anti-chicken DyLight 755 (Thermo Fisher, SA5-10075), in 3% BSA for 3 h at 37 °C. The conjugation approach was previously described[30]. Finally, samples were washed 6× with PBS-T for 3 min each, and 2× SSC with 200 ng ml^{-1} DAPI was added to visualize nuclei to image samples. Following imaging, Alexa Fluor 594 and 647 antibodies were destained with 50 mM TCEP in 2× SSC for 45 min at room temperature and FISH signal removed through treatment with 80% formamide in 2× SSC for 30 min at room temperature. In situ amplification of sgRNA sequences was then completed by incubating samples in a padlock probe and extension-ligation reaction mixture (1× Ampligase buffer, 0.4 U μl^{-1} RNase H, 0.2 mg ml^{-1} BSA, 100 nM padlock probe: /5Phos/GTTTCAGAGCTATGCTCTCCTGTTCGCCAAATTCTACCCACCACCCACTCTCCAAAGGACGAAACACCG, 0.02 U μl^{-1} TaqIT polymerase, 0.5 U μl^{-1} Ampligase and 50 nM dNTPs) for 5 min at 37 °C and 90 min at 45 °C, and washing 2× with PBS-T. Circularized padlocks were amplified using a rolling circle amplification mix (1× Phi29 buffer, 250 μM dNTPs, 0.2 mg ml^{-1} BSA, 5% glycerol and 1 U μl^{-1} Phi29 DNA polymerase) at 30 °C overnight. AlexaFluor 488-conjugated LAMP1 (Cell Signaling Technology, 58996, RRID:AB_2927691) was visualized via incubation for 2 h at 37 °C in 3% BSA at 1:500 dilution. Following imaging, in situ sequencing was performed as previously described using sequencing primer GCCAAATTCTACCCACCACCCACTCTCCAAAGGACGAAACACCG for 12 cycles.

Virus infection, phenotyping and in situ sequencing for secondary screen

HeLa-TetR-Cas9 clonal cells used in the genome-wide screen and polyclonal Huh7 cells transduced with 311-Cas9 (Addgene, plasmid

96924) and selected using 10 μg ml^{-1} blasticidin for 7 days were used for secondary screening. EBOV (strain Mayinga), Marburg virus (strain Musoke), Sudan virus (strain Gulu) were added at MOIs of ~3 in 2 ml media per well for 16 h or 24 h before fixation. Viral RNA was amplified as described for the primary screen using virus strain-specific probes purchased from Molecular Instruments for positive-sense VP35. Antibody staining and in situ sequencing were performed as described for the primary screen except for Marburg virus, where an antibody against VP40 (1:1,000, Integrated BioTherapeutics, 0203-012) rather than VP35 was used since no antibody for Marburg VP35 was available. In addition, antibodies for vimentin and LAMP1 were omitted to allow for slower higher-magnification imaging. For Sudan virus, the same VP35 antibody was used as it recognized both EBOV and Sudan VP35 protein. In situ sequencing was performed for 6 cycles.

Fluorescence microscopy for primary genome-wide and secondary targeted follow-up screens

All in situ sequencing images were acquired using a Ti2 Eclipse inverted epifluorescence microscope (Nikon) with automated *XYZ* stage control and hardware autofocus. The Lumencor CELESTA Light Engine was used for fluorescence illumination and all hardware was controlled using NIS elements software with the JOBS module. In situ sequencing cycles were imaged using a ×10 0.45 NA CFI Plan Apo λ objective (Nikon, MRD00105) with the following filters (Semrock) and exposure times for each base: G (546 nm laser at 40% power, emission 575/30 nm, dichroic 552 nm, 200 ms); T (546 nm laser at 40% power, emission 615/24 nm, dichroic 565 nm, 200 ms); A (637 nm laser at 40% power, emission 680/42 nm, dichroic 660 nm, 200 ms); C (637 nm laser at 40% power, emission 732/68 nm, dichroic 660 nm, 200 ms). For the genome-wide primary screen, phenotyping images were acquired using a ×20 0.75 NA CFI Plan Apo λ objective (Nikon, MRD00205) with the following filters (Semrock unless otherwise noted) and exposure times: DAPI (405 nm laser at 5% power, Chroma Multi LED set 89402, 50 ms), AF488 (477 nm laser at 30% power, Chroma Multi LED set 96372, 200 ms), AF594 (546 nm laser at 10% power, emission 615/24 nm, dichroic 565 nm, 200 ms), AF647 (637 nm laser at 10% power, emission 680/42 nm, dichroic 660 nm, 200 ms), Dylight 755 (749 nm laser at 10% power, emission 820/110 nm, dichroic 765 nm, 200 ms).

For the secondary screen, phenotyping images were acquired using a ×40 0.95 NA CFI Plan Apo λ objective (Nikon, MRD70470) with the following filters and exposure times: DAPI (405 nm laser at 5% power, Chroma Multi LED set 89402, 50 ms), AF594 (546 nm laser at 10% power, emission 615/24 nm, dichroic 565 nm, 200 ms) and AF647 (637 nm laser at 10% power, emission 680/42 nm, dichroic 660 nm, 200 ms).

Quantification and statistical analysis

No statistical methods were used to predetermine sample sizes, but our sample sizes are similar to those reported in previous publications[9,10,28]. Data collection and analysis were not performed blind to the conditions of the experiments except for pooled screens. Data distribution was assumed to be normal, but this was not formally tested.

In OPS, the incorporation of sgRNAs into individual cells is inherently stochastic, and cells are distributed randomly within wells, ensuring effective randomization. During secondary and tertiary assays, specific sgRNAs are selected for further evaluation and are systematically allocated to wells and slides in a randomized manner, effectively mitigating positional biases. This deliberate randomization ensures that any positional effects are accounted for, thereby attributing observed phenotypic outcomes directly to the sgRNA-induced perturbations rather than to extraneous spatial variables.

Image analysis

Images of cell phenotype and in situ sequencing of perturbations were manually aligned during acquisition using nuclear masks to calibrate the plate position to each of the four corner wells during screening. Alignment was then refined computationally via cross-correlation of DAPI signal between imaging acquisitions. Nuclei and cell detection and segmention as well as in situ sequencing read calling were performed as previously described[29]. Data analysis functions were written in Python, using Snakemake for workflow control[31]. The image analysis code is available on GitHub. Briefly, for segmentation of phenotyping images from the primary screen, nuclei were segmented using the following parameters: nuclei smooth = 4, nuclei radius = 15, nucleus area 90–1,200, DAPI intensity threshold = 1,350. Cells were segmented using signal in the vimentin channels at an intensity threshold = 3,000. For segmentation of in situ sequencing images from the primary and secondary screens, nuclei were segmented using the following parameters: nuclei smooth = 1.15, nuclei radius = 15, nucleus area 20–400, DAPI intensity threshold = 1,000–2,000, adjusted differently for each plate. Cells were segmented using signal in the four sequencing channels at intensity thresholds adjusted for each plate, between 2,500 and 4,200. For segmentation of phenotyping images from the secondary screen, nuclei were segmented using the following parameters: nuclei smooth = 9, nuclei radius = 100, nucleus area 200–18,000 for HeLa cells or 200–50,000 for Huh7 cells, DAPI intensity threshold = 4,000 for HeLa cells or 3,000 for Huh7 cells. Cells were segmented using background cell signal in the Jun channel at an intensity threshold = 1,525–1,900 for Hela cells (depending on the assay plate) or 1,625–1,825 for Huh7 cells. All other parameters used for analysis were set to default settings.

Optical pooled screen analysis

Data filtering and preprocessing. From an initial number of phenotyped cells equal to 66.5 million, cells with at least one in situ sequencing read were selected (96% of cells; Extended Data Table 1, Extended Data Fig. 2a). Since both phenotyping and in situ sequencing images were acquired with 15% field of view (FOV) overlap at different magnifications (×20 and ×10, respectively), some phenotyping cells aligned to multiple cells with sequencing information (4%). We then filtered out cells with in situ sequencing reads not mapped to any sgRNA in the library (expected to be 4^{12}/80,833 = 0.48% by chance), resulting in an additional 24% of cells filtered out (Extended Data Fig. 2a). After this initial filtering, we plotted mean values for each channel from 2 out of the 8 plates screened, averaging across 2,562 fields of view (~33% of all fields of view) per plate (Extended Data Fig. 2b). We noticed that illumination was not even and differed by channel and microscope. We therefore used a previously published illumination-correction algorithm (described in Singh et al.[32] and equivalent to CellProfiler's CorrectIlluminationCalculate module with options 'Regular', 'All' and 'Median Filter') to correct illumination on a per-plate and per-channel basis (results shown in Extended Data Fig. 2c). To mitigate the impact of illumination variation remaining as a consequence of vignetting, we averaged post-illumination correction factors from Extended Data Fig. 2c across channels for each plate and masked out the bottom 10th percentile, eliminating strongly impacted cells (see Extended Data Fig. 2d for the resulting mask for plates A and B). After filtering, we plotted the VP35 median cell intensity and vimentin median cell intensity for 10,000 cells randomly sampled across all plates in the screen and found no systematic variation with i and j coordinates within the field of view (Extended Data Fig. 2e).

Next, we considered that although we had already filtered out cells that were phenotyped and mapped to multiple cells with sequencing reads (due to the 15% acquisition overlap for the sequencing images), we had not yet filtered out cells that were phenotyped twice (phenotyping images were also acquired with 15% acquisition overlap). We therefore kept only one phenotyped cell per cell with a sequencing read (Extended Data Fig. 2f) and additionally removed phenotyped cells whose centres were within 10 pixels of the field of view border. We noted that there was well-level variability in the data (see example

in Extended Data Fig. 2g for VP35 RNA FISH channel plate I well A1). We therefore normalized features on a per-cell basis relative to cells in the same field of view by subtracting the median and dividing by the median absolute deviation (MAD) × 1.4826 (ref. 18) of non-targeting cells ±1,480 pixels (width of a field of view) in i or j from each cell of interest (minimum 464, mean 3,257 cells). Extended Data Fig. 2g shows an example of the per-FOV mean post normalization.

We also manually annotated 2,520 cells on the basis of DAPI signal into three categories (apoptic, interphase and mitotic, obtaining mean receiver operating characteristic area under the curve (ROC AUCs) of 0.94, 0.97 and 0.99 for these categories, respectively). Annotated data were split into 5 stratified splits using scikit-learn 1.1.3 and the sklearn.model_selection.StratifiedKFold function. The sklearn.ensemble.RandomForestClassifier function with default parameters was separately used to predict each class with a OneVsRestClassifier. Only interphase cells (99.3% of cells) were used for further downstream analysis (Extended Data Fig. 2h).

Following filtering, the majority of targeting sgRNAs had hundreds of cells per sgRNA: median 456 cells per sgRNA, 25th percentile 267 cells per sgRNA (Extended Data Table 2, Extended Data Fig. 2i). Only genes with a minimum of 2 sgRNAs with at least 50 cells per sgRNA were considered for further analyses (20,338 out of 20,392, >99.7%).

ΔAUC and random forest regression analysis. Scores for features relative to non-targeting controls were determined by calculating differences in cumulative AUCs. These ΔAUCs were averaged over sgRNAs for a given gene and significance was determined by comparing ΔAUCs for individual sgRNAs to distributions bootstrapped from non-targeting control cells (bootstrapped 100,000 times). Gene-level *P* values were calculated using Stouffer's method and then corrected using the Benjamini–Hochberg procedure. Random forest regression models were trained on features from the VP35 protein channel only (for predicting VP35 RNA FISH levels) or the VP35 protein and RNA levels (for predicting c-Jun nuclear intensity) for 50,000 randomly selected non-targeting control cells using sklearn.ensemble.RandomForestRegressor with scikit-learn 1.1.3, with random state set to 7, n_estimators = 100 and max_features = 'sqrt'[33]. Statistical significance was determined as described for the cumulative AUC analysis above.

Dimensionality reduction, clustering and gene enrichment analysis. In Figs. 1 and 2, Enrichr results[34] were determined using gseapy (0.14.0)[35] with the 2021 KEGG and GO gene sets and the 2016 Reactome gene sets. PHATE (1.0.10)[21] was used to perform dimensionality reduction on single cells with Euclidean distance, cosine mds distance, gamma = 1, knn = 5 and 20 PCs in Fig. 2. For gene-level clustering in Fig. 3, PHATE with Euclidean distance, cosine mds distance, gamma = 1, knn = 3 and the number of PCs giving 95% of the variance (282 for unsupervised and 372 for supervised features) were used. Before performing PCA, feature sets were standardized using sklearn.preprocessing.StandardScaler. Following dimensionality reduction, Leiden clustering[25] was performed with the resolution parameter determined on the basis of the Adjusted Rand Score computed by subsampling 95% of the data and reclustering at least ten times for each resolution value (Extended Data Fig. 4a).

Deep learning model

Unsupervised training. First, we trained an unsupervised convolutional autoencoder on 40 million 64 × 64 6-channel single-cell images. We implemented a U-Net style architecture[36] with an encoder containing 5 convolutional layers and a decoder containing 5 convolutional layers (Fig. 2a). We used strided convolution to reduce dimensionality of the images in the encoder, resulting in a 2,048-dimensional embedding. Bilinear upsampling was used in the decoder to map from this 2,048-dimensional latent space back to the original image space.

For model selection, we used an 80/20 train/test split. In each epoch of training, at least 256 cells were sampled from each training field of view. As a loss function, we used the mean squared error (MSE) over all reconstructed pixels within the watershed mask for each cell of interest. We used the Adam optimizer with a learning rate of 0.001, trained for 50 epochs with a random seed. Training and test loss curves are shown in Extended Data Fig. 3a.

Supervised training. We then fine-tuned the encoder of the trained autoencoder by training a classification head with the negative log-likelihood loss using Adam with a learning rate of 0.001. We used 3,889 cell images manually labelled with one of four phenotypic categories in the set (faint, punctate, cytoplasmic, peripheral). We used stratified dataset splitting implemented in scikit-learn[33] to ensure a balanced distribution of phenotypes across classes. Namely, we reserved 25% of the dataset, or 973 cell images, as a held-out test set, and the remaining 2,916 images were split into 4 stratified folds (train/validation) and used for fine tuning the autoencoder.

For fine tuning the autoencoder, we applied the following data augmentations using the available image transforms in PyTorch: (1) rotations up to 180 degrees, (2) random vertical and horizontal flips, (3) random perspective with a distortion scale of 0.1, (4) random affine transformations with a shear of 10 and a scale of (0.75, 1.25) and (5) Gaussian blur with a kernel size of 5 and standard deviation uniformly selected in the interval (0.05, 0.5). For training our model, we constructed a batch by sampling 5 images from each of the four classes, applying 5 random transformations to each image, resulting in 100 images per batch. We sampled 100 batches in any given epoch. The balanced sampling strategy was used to account for any class imbalances. We trained for a total of 50 epochs and chose the model that achieved the highest balanced accuracy defined as accuracy across each class weighted by the proportion of validation samples in the class.

To fairly compare model performance across embeddings from various standard deep learning models on the held-out test set, we trained support vector machines (SVMs) with a linear kernel on the embeddings extracted from these models on the held-out test set. In particular, we split our test embeddings according to a 95–5 split, trained the SVM on the 95% and evaluated model performance on the remaining 5%. We trained SVMs on the embeddings obtained from the fully unsupervised autoencoder model, the fine-tuned autoencoder model, a pretrained ResNet model and hand-crafted features (Extended Data Fig. 4d).

To determine sgRNAs that significantly altered the proportion of cells in each phenotypic category, an ordinal chi squared test was performed using R 4.2.2 with the coin package (v.1.0.9), and the results were combined at the gene level using Stouffer's method.

Identification of terpestacin as an antiviral compound

The evening before the infection, ~15,000 HeLa cells were seeded into each well of a 96-well plate. The following day, terpestacin (Aurora Fine Chemicals), a known small-molecule inhibitor of UQCRB, was dosed onto cells in a 9-point, 2-fold dose curve beginning at 50 µM. Cells were infected with EBOV at an MOI of 0.1–0.2 and allowed to infect for 48 h. The plates were then fixed in 10% neutral-buffered formalin for >6 h and removed from containment. The cells were washed, permeabilized with 0.1% Triton X-100, blocked with 3.5% BSA and immunostained with an anti-EBOV GP monoclonal antibody (IBT). After several hours incubating at 37 °C, the cells were washed in PBS and incubated in anti-mouse Alexa Fluor 488 secondary antibody (Thermo Fisher). Cells were again washed in PBS and nuclei were counterstained with Hoechst 33342. The plate was imaged with a BioTek Cytation 1 automated plate imager. Images were fed into a custom pipeline in CellProfiler (Broad Institute) used to count the number of infected cells and nuclei. Infection efficiency was calculated as the number of infected cells divided by the total number of nuclei (as a proxy for total cell count) and normalized to the average of cells treated with dimethylsulfoxide only. The total nuclei

count, normalized to the average of the negative controls, was used to check for potential cytotoxicity. The normalized infection efficiency and normalized nuclei counts were plotted in GraphPad Prism 8.0.0 (GraphPad Software). A 4-parameter variable slope nonlinear regression was used to fit the data and calculate the antiviral IC_{50} of the compound.

Generation of STRAP KO cells

The evening before transduction, 250,000 HeLa-TetR-Cas9 cells were seeded into a 12-well plate in DMEM medium supplemented with 10% FBS. The following morning, the medium was replaced with DMEM containing 8 μg ml^{-1} polybrene. A lentiviral vector containing an sgRNA targeting STRAP purchased from the Broad Institute's Genetic Perturbation Platform was added and the cells were spinoculated at 1,000 × *g* for 2 h at 33 °C. After spinoculation, the cells were transferred to the incubator for 3 h, followed by medium replacement. At 24 h after spinoculation, the cells were split into medium containing 2 μg ml^{-1} puromycin to begin selection. At 48 h after beginning selection, the medium was replaced with that containing 2 μg ml^{-1} puromycin and 1 μg ml^{-1} doxycycline to induce Cas9 activation. Once the cells reached confluency, they were split into medium containing only 1 μg ml^{-1} doxycycline and were maintained in doxycycline for 1 week. After 1 week, the cells were split into basic medium and single-cell clones were obtained. Clonal populations were identified and verified as being STRAP KOs by capillary electrophoresis using a Jess automated immunodetection machine (Bio-Techne). Two clones, identified as B03 and E08, had no detectable STRAP expression and were used for subsequent experiments.

qPCR of EBOV RNA

To detect and quantify total cellular EBOV RNA, NT or STRAP KO cells were infected at the indicated MOI and incubated for ~16 h. Cells were washed once with PBS and collected in TRIzol (ThermoFisher). RNA was processed using a Zymo Direct-zol RNA Miniprep kit following manufacturer instructions. RT–qPCR was performed using the NEB Luna Universal Probe One-Step RT–qPCR kit (E3006L) on a Bio-Rad CFX Opus 96 Real Time PCR system. Primer/probe sequences for EBOV NP (forward: GCAGAGCAAGGACTGATACA, reverse: GTTCGCATCAAACGGAAAAT, probe- FAM-CAACAGCTT-ZEN-GGCAATCAGTAGGACA-IABkFQ) and human GAPDH (forward: ACATCGCTCAGACACCATG, reverse: GTAGTGAGGTCAATGAAGGG, probe: Cy5-AAGGTCGGAGTCAACGG ATTTGGTC-IAbRQSp) were previously established. The thermocycler protocol was as follows: 55 °C for 10 min, 95 °C for 1 min, then 40 cycles of 95 °C for 10 s and 60 °C for 30 s. Primers and probes were synthesized by IDT. Four concentrations of synthetic RNA standards were used to calculate genome equivalents from Cq values. Each standard and sample were run in duplicate technical replicates and averaged. To detect extracellular EBOV RNA, NT or STRAP KO cells were infected at the indicated MOIs and incubated for ~72 h. Supernatants from infected cells were collected into TRIzol LS (ThermoFisher) in a 3:1 TRIzol LS to supernatant ratio. RNA was processed and PCR was performed as above using the same primers.

To quantify genome-specific EBOV RNA, whole-cell lysates of infected NT or STRAP cells were generated and processed as above. A two-step PCR was performed to specifically target EBOV genomic RNA, which has been previously described. For reverse transcription, an Invitrogen SuperScript III Reverse Transcriptase kit was used following manufacturer protocol. Briefly, 4 μl purified RNA was mixed with 1 μl 2 μM gene-specific primer (EBOV-18046 forward: 5′-GAGTTGATTAGTGTGTGCAATAGGTTTAC-3′), 1 μl 10 mM dNTP mix and 4 μl nuclease-free water. The mixture was denatured for 5 min at 65 °C, then cooled on ice for >1 min. After cooling, 4 μl First Strand Buffer, 1 μl 0.1 M dithiothreitol, 1 μl RNase inhibitor, 1 μl SuperScript III reverse transcriptase and 3 μl nuclease-free water were added (total reaction volume of 20 μl). Reverse transcription was performed at 55 °C for 1 h, and the enzyme was then inactivated at 85 °C for 5 min. A volume of 4 μl of the first reaction was used as the template for the subsequent qPCR using the Bio-Rad i*Taq* Universal SYBR Green Supermix according to manufacturer protocol.

EBOV immunostaining and infection efficiency calculation

Cells were infected as described above with a low MOI. After a 72 h infection, the plates were fixed in 10% formalin and removed from containment. Cells were immunostained with an anti-EBOV GP mouse antibody (IBT) and Alexa Fluor 488 anti-mouse secondary antibody (Thermo Fisher). Nuclei were counterstained with Hoechst 33342. Images were taken with a Biotek Cytation 1 automated plate imager. The numbers of infected cells and nuclei were quantified using a customized pipeline on CellProfiler (Broad Institute) and infection efficiency was calculated by dividing the total number of infected cells by the total number of nuclei.

Focus-forming unit (f.f.u.) assay

Non-targeting and STRAP knockout HeLa cells were infected as described above with a low MOI in triplicate. After a 72 h infection, supernatant was collected from samples and diluted out onto HeLa cells in a 96-well plate in a 2-fold series. Cells were incubated for another 72 h to allow viral replication and spread. The plates were formalin fixed, removed from containment and immunostained for EBOV GP as above. One f.f.u. was defined as a cluster of >5 infected cells. The lowest dilution with an average of >50 f.f.u.s was used as the endpoint to calculate viral titres. The data were normalized to the average of the non-targeting controls.

Proximity ligation assay

The night before infection, ~7,000 HeLa cells were seeded into an 18-well chamber slide (Ibidi). The following evening, cells were infected with EBOV at an MOI of ~3. After 16–18 h, the cells were fixed in 10% formalin and removed from containment. Cells were washed thoroughly with PBS and the Duolink PLA kit was used (Sigma-Aldrich) according to manufacturer protocol. Briefly, blocking solution was added for 1 h at 37 °C. The anti-EBOV VP35 mouse antibody and anti-STRAP rabbit antibody were diluted 1:3,000 and 1:1,000, respectively, in antibody diluent and incubated at 37 °C for 2–3 h. The cells were washed in wash buffer A and diluted Duolink PLA probes were added for 1 h at 37 °C. Cells were washed in wash buffer A and ligase was added in ligase buffer for 30 min at 37 °C. Cells were again washed in wash buffer A and polymerase diluted in amplification buffer was added for 100 min at 37 °C. The cells were finally washed in wash buffer B and nuclei were counterstained with Hoechst 33342. Images were taken on a Nikon Ti2 Eclipse microscope.

Co-immunostaining of EBOV VP35 and STRAP proteins

The night before infection, ~30,000 HeLa cells were seeded into 8-well chamber slides. Cells were infected with EBOV at an MOI of ~3 for 16–18 h, then fixed in 10% formalin and removed from containment. Cells were permeabilized in 0.1% Triton X-100, blocked in 3.5% BSA, and immunostained with our anti-EBOV VP35 mouse antibody and an anti-STRAP rabbit antibody (Atlas Antibodies) at 1:3,000 and 1:1,000, respectively, for several hours at 37 °C. Cells were washed, then anti-rabbit Alexa Fluor 488 and anti-mouse Alexa Fluor 594 secondary antibodies were added (Thermo Fisher). After 1–2 h, the cells were washed and nuclei were counterstained with Hoechst 33342. Images were taken on a Nikon Ti2 Eclipse microscope.

Reporting summary

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

Data availability

Raw screen images and single-cell features are available in their entirety on Google Cloud Storage at https://console.cloud.google.com/storage/browser/opspublic-east1/EBOVOpticalPooledScreen. A curated

subset of raw images from the genome-wide screen in HeLa cells (https://doi.org/10.7910/DVN/YHVWXY), and the targeted screens in HeLa and Huh7 cells at 16 h post infection (https://doi.org/10.7910/DVN/6FQNUA) and at 24 h post infection (https://doi.org/10.7910/DVN/W9WVHG) are available at Harvard's Dataverse[37–39]. Additional tables are available on Zenodo at https://doi.org/10.5281/zenodo.14741479 (ref. 40). Source data are provided with this paper.

Code availability

Code is available on Zenodo at https://zenodo.org/records/15725070 (ref. 41).

References

1. Hoenen, T., Groseth, A. & Feldmann, H. Therapeutic strategies to target the Ebola virus life cycle. *Nat. Rev. Microbiol.* **17**, 593–606 (2019).
2. Hoenen, T. et al. Inclusion bodies are a site of Ebolavirus replication. *J. Virol.* **86**, 11779–11788 (2012).
3. Nanbo, A., Watanabe, S., Halfmann, P. & Kawaoka, Y. The spatio-temporal distribution dynamics of Ebola virus proteins and RNA in infected cells. *Sci. Rep.* **3**, 1206 (2013).
4. Bruchez, A. et al. MHC class II transactivator CIITA induces cell resistance to Ebola virus and SARS-like coronaviruses. *Science* **370**, 241–247 (2020).
5. Carette, J. E. et al. Ebola virus entry requires the cholesterol transporter Niemann-Pick C1. *Nature* **477**, 340–343 (2011).
6. Cheng, H. et al. A parallel genome-wide RNAi screening strategy to identify host proteins important for entry of Marburg virus and H5N1 influenza virus. *Virol. J.* **12**, 194 (2015).
7. Martin, S. et al. A genome-wide siRNA screen identifies a druggable host pathway essential for the Ebola virus life cycle. *Genome Med.* **10**, 58 (2018).
8. Feldman, D. et al. Optical pooled screens in human cells. *Cell* **179**, 787–799.e17 (2019).
9. Carlson, R. J., Leiken, M. D., Guna, A., Hacohen, N. & Blainey, P. C. A genome-wide optical pooled screen reveals regulators of cellular antiviral responses. *Proc. Natl Acad. Sci. USA* **120**, e2210623120 (2023).
10. Funk, L. et al. The phenotypic landscape of essential human genes. *Cell* **185**, 4634–4653.e22 (2022).
11. Galão, R. P. et al. TRIM25 and ZAP target the Ebola virus ribonucleoprotein complex to mediate interferon-induced restriction. *PLoS Pathog.* **18**, e1010530 (2022).
12. Hölzer, M. et al. Differential transcriptional responses to Ebola and Marburg virus infection in bat and human cells. *Sci. Rep.* **6**, 34589 (2016).
13. Wynne, J. W. et al. Comparative transcriptomics highlights the role of the activator protein 1 transcription factor in the host response to Ebolavirus. *J. Virol.* **91**, e01174-17 (2017).
14. Filone, C. M., Dower, K., Cowley, G. S., Hensley, L. E. & Connor, J. H. Probing the virus host interaction in high containment: an approach using pooled short hairpin RNA. *Assay Drug Dev. Technol.* **13**, 34–43 (2015).
15. Flint, M. et al. A genome-wide CRISPR screen identifies *N*-acetylglucosamine-1-phosphate transferase as a potential antiviral target for Ebola virus. *Nat. Commun.* **10**, 285 (2019).
16. Gong, M. et al. Genome-wide CRISPR/Cas9 screen identifies SLC39A9 and PIK3C3 as crucial entry factors for Ebola virus infection. *PLoS Pathog.* **20**, e1012444 (2024).
17. Fang, J. et al. Functional interactomes of the Ebola virus polymerase identified by proximity proteomics in the context of viral replication. *Cell Rep.* **38**, 110544 (2022).
18. Bray, M.-A. et al. Cell Painting, a high-content image-based assay for morphological profiling using multiplexed fluorescent dyes. *Nat. Protoc.* **11**, 1757–1774 (2016).
19. Baldi, P. Autoencoders, unsupervised learning, and deep architectures. In *Proc. ICML Workshop on Unsupervised and Transfer Learning* 37–49 (JMLR, 2012).
20. Moon, K. R. et al. Visualizing structure and transitions in high-dimensional biological data. *Nat. Biotechnol.* **37**, 1482–1492 (2019).
21. Luthra, P. et al. Inhibiting pyrimidine biosynthesis impairs Ebola virus replication through depletion of nucleoside pools and activation of innate immune responses. *Antiviral Res.* **158**, 288–302 (2018).
22. Batra, J. et al. Protein interaction mapping identifies RBBP6 as a negative regulator of Ebola virus replication. *Cell* **175**, 1917–1930. e13 (2018).
23. Woolsey, C. et al. A VP35 mutant Ebola virus lacks virulence but can elicit protective immunity to wild-type virus challenge. *Cell Rep.* **28**, 3032–3046.e6 (2019).
24. Jung, H. J. et al. Terpestacin inhibits tumor angiogenesis by targeting UQCRB of mitochondrial complex III and suppressing hypoxia-induced reactive oxygen species production and cellular oxygen sensing. *J. Biol. Chem.* **285**, 11584–11595 (2010).
25. Traag, V. A., Waltman, L. & van Eck, N. J. From Louvain to Leiden: guaranteeing well-connected communities. *Sci. Rep.* **9**, 5233 (2019).
26. Söderberg, O. et al. Direct observation of individual endogenous protein complexes in situ by proximity ligation. *Nat. Methods* **3**, 995–1000 (2006).
27. Replogle, J. M. et al. Mapping information-rich genotype-phenotype landscapes with genome-scale Perturb-seq. *Cell* **185**, 2559–2575. e28 (2022).
28. Gentili, M. et al. Classification and functional characterization of regulators of intracellular STING trafficking identified by genome-wide optical pooled screening. *Cell Syst.* **15**, 1264–1277. e8 (2024).
29. Feldman, D. et al. Pooled genetic perturbation screens with image-based phenotypes. *Nat. Protoc.* **17**, 476–512 (2022).
30. Ramezani, M. et al. A genome-wide atlas of human cell morphology. *Nat. Methods* **22**, 621–633 (2025).
31. Köster, J. & Rahmann, S. Snakemake—a scalable bioinformatics workflow engine. *Bioinformatics* **28**, 2520–2522 (2012).
32. Singh, S., Bray, M.-A., Jones, T. R. & Carpenter, A. E. Pipeline for illumination correction of images for high-throughput microscopy. *J. Microsc.* **256**, 231–236 (2014).
33. Pedregosa, F. et al. Scikit-learn: machine learning in Python. *J. Mach. Learn. Res.* **12**, 2825–2830 (2011).
34. Kuleshov, M. V. et al. Enrichr: a comprehensive gene set enrichment analysis web server 2016 update. *Nucleic Acids Res.* **44**, W90–W97 (2016).
35. Fang, Z., Liu, X. & Peltz, G. GSEApy: a comprehensive package for performing gene set enrichment analysis in Python. *Bioinformatics* **39**, btac757 (2023).
36. Ronneberger, O., Fischer, P. & Brox, T. U-Net: convolutional networks for biomedical image segmentation. *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2015: 18th International Conference, Proc. Part III* 18 (pp. 234–241) (Springer International Publishing, 2015).
37. Carlson, R. Ebola virus optical pooled screen | genome-wide screen images. *Harvard Dataverse* 10.7910/DVN/YHVWXY (2025).
38. Carlson, R. Ebola virus optical pooled screen | secondary screen early timepoint (16 hours) images. *Harvard Dataverse* 10.7910/DVN/6FQNUA (2025).
39. Carlson, R. Ebola virus optical pooled screen | secondary screen late timepoint (24 hours) images. *Harvard Dataverse* 10.7910/DVN/W9WVHG (2025).
40. Carlson, R. Additional data for single-cell image-based screens identify host regulators of Ebola virus infection dynamics [Data set]. *Zenodo* https://doi.org/10.5281/zenodo.14741479 (2025).

41. Stefanakis, G. & Carlson R. beccajcarlson/EBOVOptical-PooledScreen: v01. *Zenodo* https://zenodo.org/records/15725070 (2025).
42. Carroll, S. A. et al. Molecular evolution of viruses of the family Filoviridae based on 97 whole-genome sequences. *J. Virol.* **87**, 2608–2616 (2013).

Acknowledgements

We thank members of the Blainey, Davey, Uhler and Hacohen labs for critical feedback and discussions; C. Diaz and J. Bauman in the lab of J. T. Neal at the Broad Institute for assistance in developing custom antibody conjugations; and Y. Qin for assistance with data management. The HeLa cell line was used in this research. Henrietta Lacks, and the HeLa cell line that was established from her tumour cells without her knowledge or consent in 1951, have made significant contributions to scientific progress and advances in human health. We thank Lacks, now deceased, and the Lacks family for their contributions to biomedical research. This work was supported by the Broad Institute through startup funding (to P.C.B.) and the BN10 programme, and two grants from the National Human Genome Research Institute (HG009283 and RM HG006193). P.C.B. was supported by a Career Award at the Scientific Interface from the Burroughs Wellcome Fund. R.J.C. was supported by a Fannie and John Hertz Foundation Fellowship and an NSF Graduate Research Fellowship. C.F.B. was supported by R01AI148663 and P01AI120943. A.R. was supported by a George F. Carrier Postdoctoral Fellowship. A.R. and C.U. acknowledge support from the Eric and Wendy Schmidt Center at the Broad Institute, NCCIH/NIH (1DP2AT012345), and ONR (N00014-22-1-2116). G.K.A., D.W.L, C.F.B and R.A.D. are supported by NIH P01AI120943.

Author contributions

R.J.C. and J.J.P. designed the approach with input from all authors. R.J.C, J.J.P., B.Y.S. and N.T. performed experiments. R.J.C. performed analysis aside from developing the deep learning model. A.R. and G.S. developed the deep learning methodology and designed the architecture for it with input from R.J.C., J.J.P. and C.U., and G.S. trained the model to obtain the single-cell embeddings. A.S. provided critical feedback and performed custom antibody conjugation. D.W.L. and K.C.F.S. developed the VP35 antibody. G.K.A., K.C.F.S., D.W.L., C.F.B., N.H., C.U., R.A.D. and P.C.B. supervised the research. R.J.C. and J.J.P. wrote the manuscript with contributions from all authors.

Competing interests

P.C.B. is a consultant to or holds equity in 10X Genomics, General Automation Lab Technologies/Isolation Bio, Next Gen Diagnostics, Cache DNA, Concerto Biosciences, Stately, Ramona Optics, Bifrost Biosystems, and Amber Bio. His laboratory received research funding from Calico Life Sciences, Merck, and Genentech for work related to genetic screening. N.H. holds equity in and advises Danger Bio/Related Sciences, owns equity in BioNtech and receives research funding from Bristol Myers Squibb. C.U. serves on the Scientific Advisory Board of Immunai, Relation Therapeutics, and Focal Biosciences, and receives research funding from AstraZeneca and Janssen Pharmaceuticals. The Broad Institute and MIT may seek to commercialize aspects of this work, and related applications for intellectual property have been filed. A.S. is an employee at Genentech and R.J.C. is an employee at Flagship Pioneering. The remaining authors declare no competing interests.

Additional information

Extended data is available for this paper at https://doi.org/10.1038/s41564-025-02034-3.

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41564-025-02034-3.

Correspondence and requests for materials should be addressed to Robert A. Davey or Paul C. Blainey.

Peer review information *Nature Microbiology* thanks Stefan Bonn, Thomas Hoenen 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.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.

© The Author(s), under exclusive licence to Springer Nature Limited 2025

https://doi.org/10.1038/s41564-025-02034-3

Extended Data Table 1 | Statistics for fraction and number of cells post initial filtering steps based on in situ sequencing data

	Number of cells	Fraction of phenotyped cells	Fraction of category above
Phenotyped cells	6.65E+07		
Cells with >=1 read	6.39E+07	0.96	0.96
Cells uniquely aligned	6.16E+07	0.93	0.96
Cells with majority read matching sgRNA in library	4.65E+07	0.70	0.76

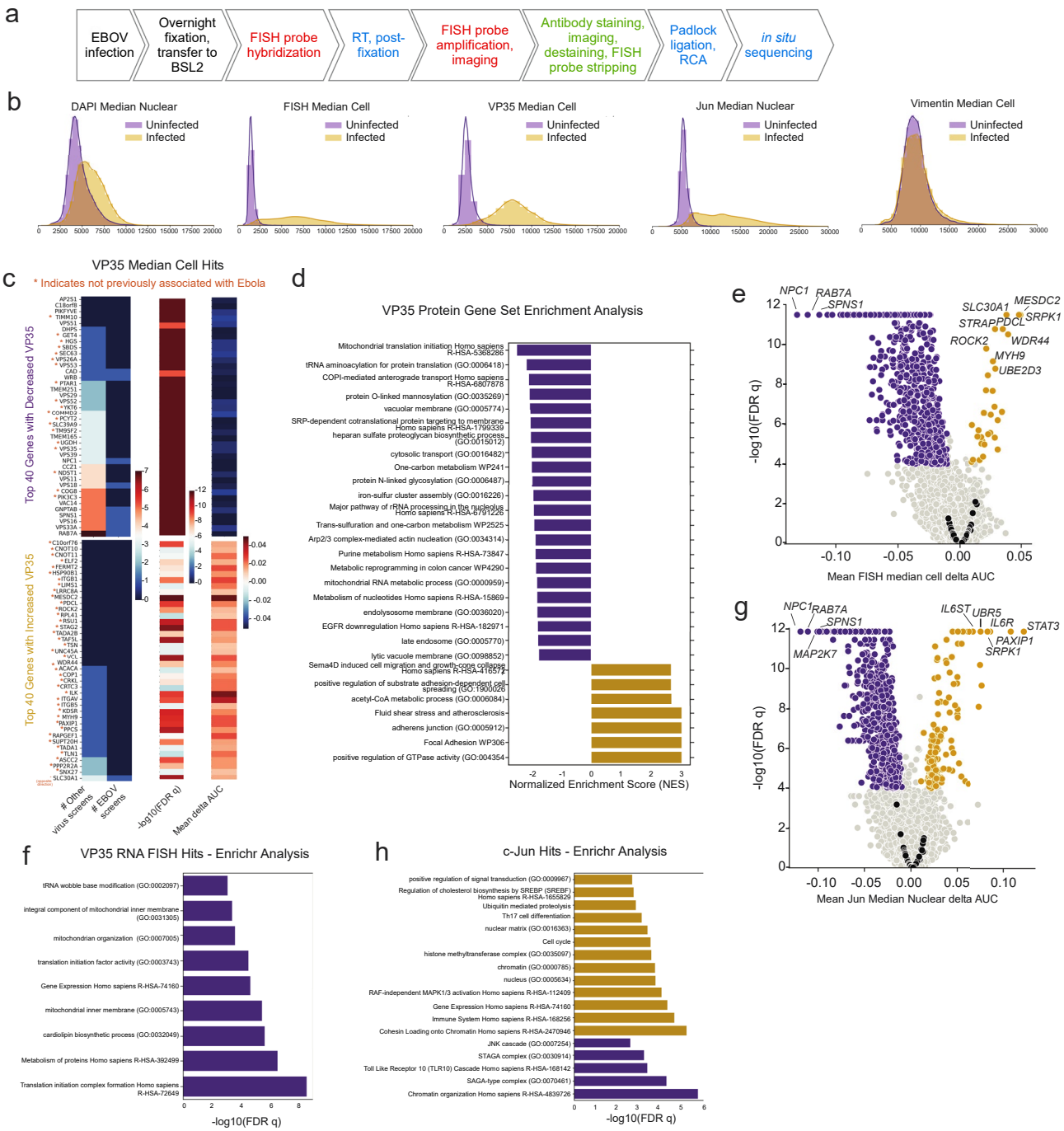
https://doi.org/10.1038/s41564-025-02034-3

Extended Data Table 2 | Statistics on cell coverage for targeting and non-targeting sgRNA post filtering

	Cells/sgRNA	
	Targeting	**Non-targeting**
Count	80739	454
Mean	540.2	6773.4
s.d.	403.2	7426.7
Min	1	74
25th Percentile	267	2502
50th Percentile	456	4768
75th Percentile	710	8515.5
Max	7446	85343

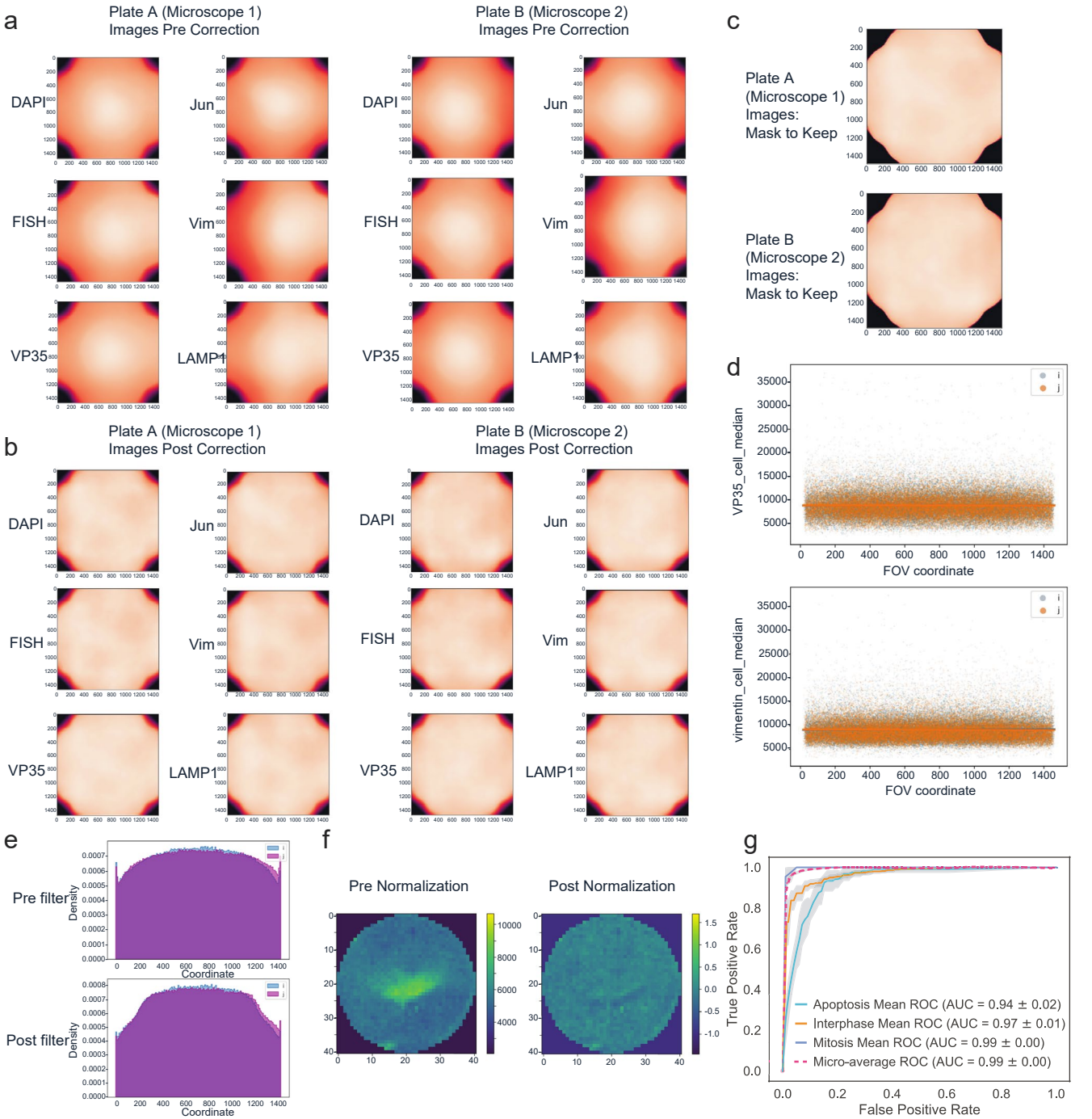
Cells/sgRNA for targeting and non-targeting sgRNAs post-filtering.

https://doi.org/10.1038/s41564-025-02034-3



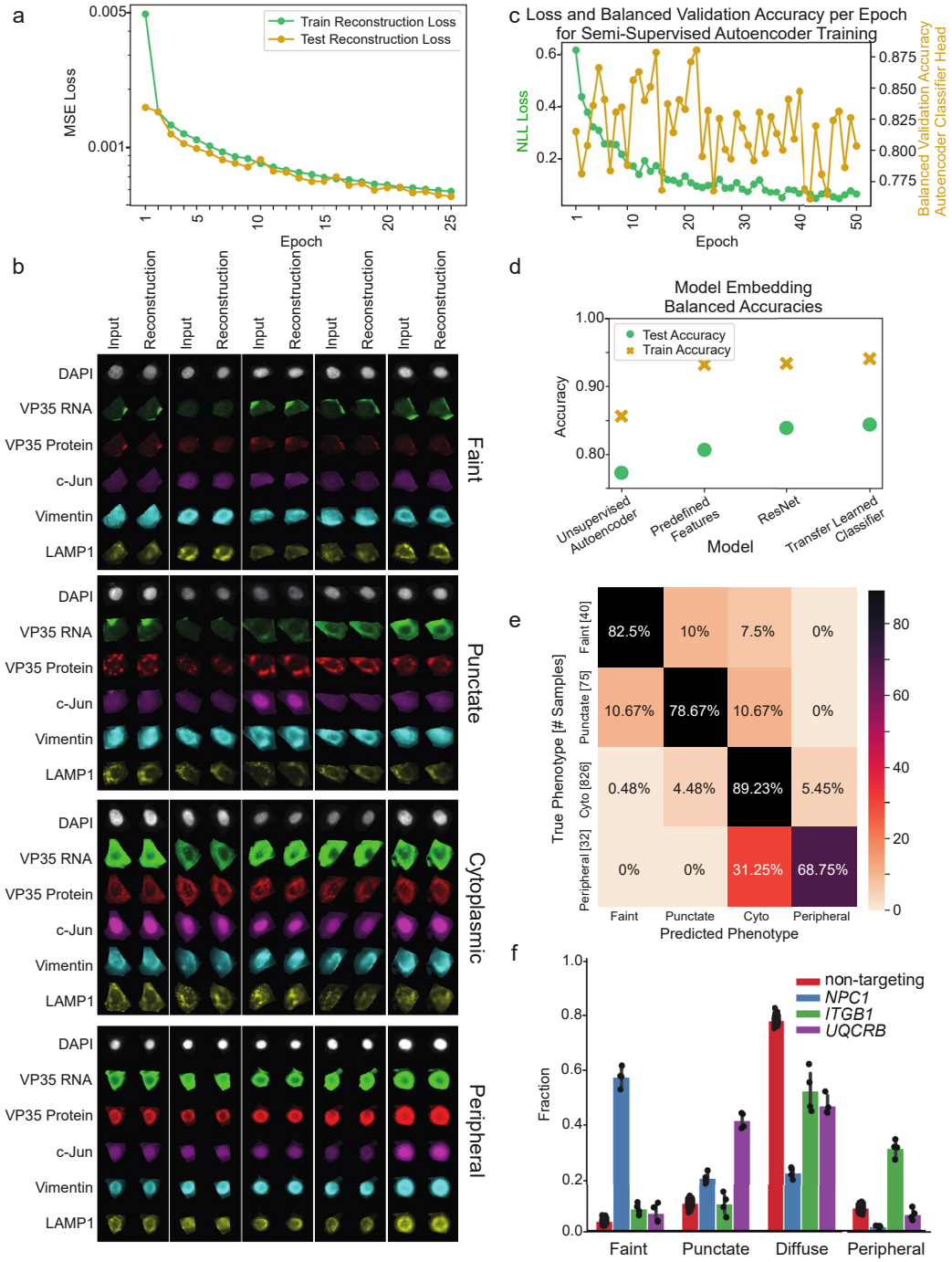
Extended Data Fig. 1 | A genome-wide optical pooled screen identifies genes affecting VP35 protein, RNA, and c-Jun. (**a**) Integration of optical pooled screening workflow with RNA FISH detection using HCR amplification. (**b**) Histograms of intensity features in five channels for non-targeting controls cells that were infected or not infected in the genome-wide optical pooled screen at 28 h. (**c**) Top 40 hits with increased or decreased VP35 protein levels and the number of non-Ebola virus genetic screens or Ebola-specific genetic screens they scored in. Genes not previously associated with Ebola in the literature are marked with an orange asterisk. (**d**) Gene set enrichment analysis of genes with significantly decreased (purple) or increased (gold) Ebola virus VP35 protein levels. (**e**) Volcano plot showing genes that scored significantly for changes in VP35 RNA levels by FISH. (**f**) Enrichr analysis of gene ontology terms significantly enriched in genes that reduced VP35 RNA levels. (**g**) Volcano plot showing genes that scored significantly for changes in c-Jun levels. (**h**) Enrichr analysis of gene ontology terms significantly enriched in genes that reduced or enhanced c-Jun levels.

https://doi.org/10.1038/s41564-025-02034-3



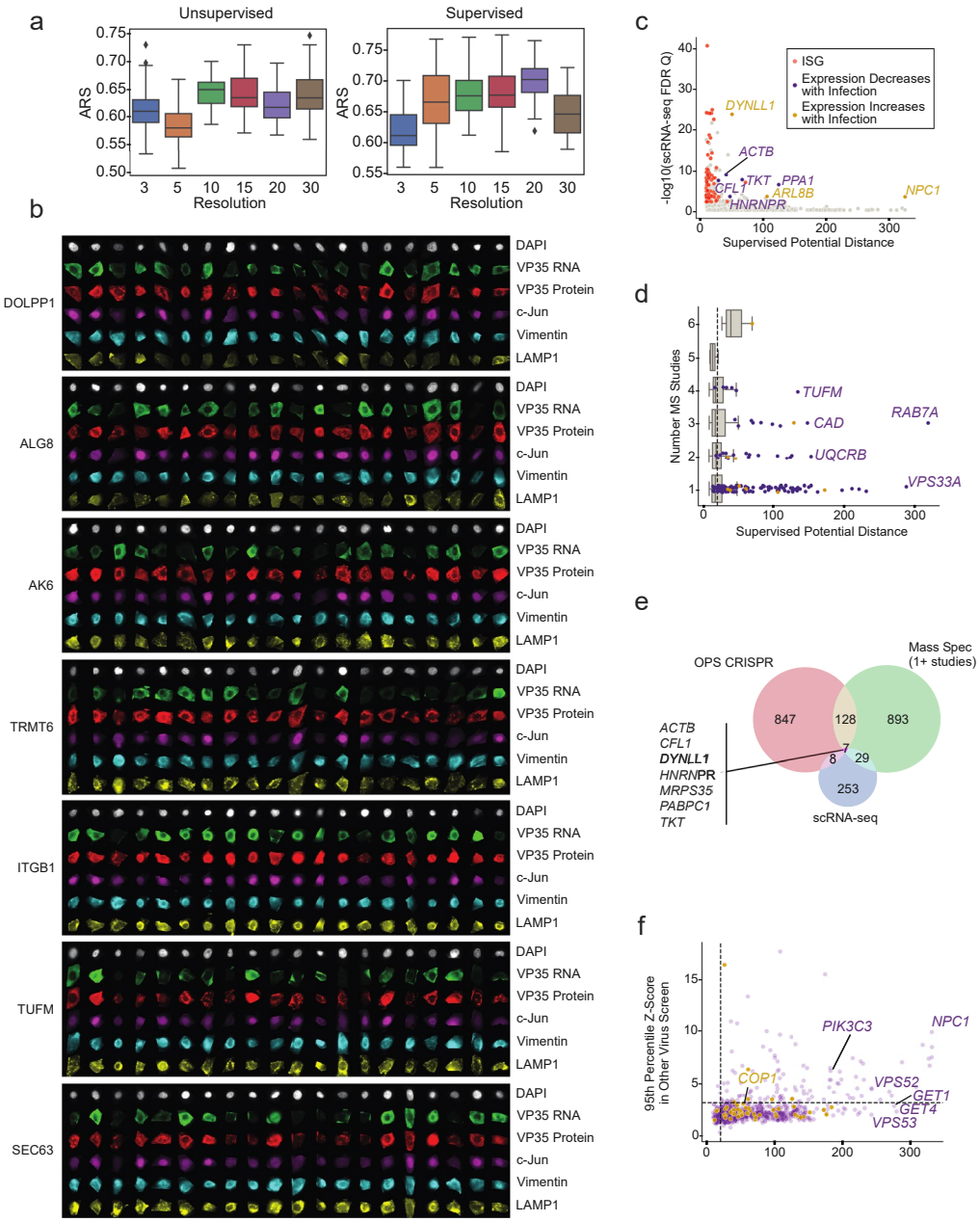
Extended Data Fig. 2 | Quality control and filtering information for genome-wide optical pooled screen. (**a**) Per-channel and per-plate mean intensities across 2562 fields of view for two out of eight plates in the genome-wide screen. (**b**) Same as (**a**) but post illumination-correction. (**c**) Same as (**b**) but after averaging across all channels for each plate and removing the bottom 10th percentile of each image. (**d**) Scatterplots of i and j coordinates for VP35 median intensity per-cell (top) and vimentin median intensity per-cell (bottom) for 10,000 randomly selected cells across the entire genome-wide screen post post-illumination correction. (**e**) Frequency of i and j coordinates for cells pre-filtering for duplicate phenotyping cells and cells at field of view edges and post-filtering. (**f**) Pre-normalization mean per-field of view FISH median intensity (plate I, well A1) pre-normalization and post-scaling to mean and unit variance based on features for non-targeting cells within -/+ the width of a field of view of each cell of interest. (**g**) Performance for the random forest classification of apoptotic, mitotic, and interphase cells trained on manual annotations.

https://doi.org/10.1038/s41564-025-02034-3



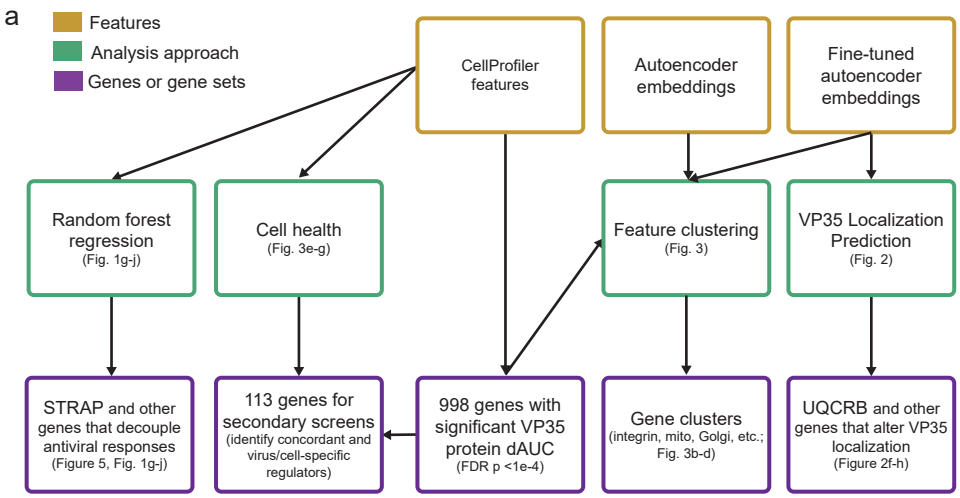
Extended Data Fig. 3 | Additional unsupervised and fine-tuned autoencoder metrics. (**a**) Fully unsupervised autoencoder reconstruction losses for training and test sets across 25 epochs. (**b**) Examples of manually labelled faint, punctate, cytoplasmic, and peripheral input cell images with accompanying unsupervised autoencoder reconstructions. (**c**) Fine-tuned autoencoder trained using negative log likelihood loss with balanced validation accuracy also reported across 50 epochs of training. (**d**) Best model train and test set accuracies for the VP35 protein localization prediction task using SVMs on latent embeddings from the unsupervised autoencoder, predefined features, a Resnet-50 architecture trained on the prediction task, or the fine-tuned autoencoder. Predefined features include intensity, correlation, and texture morphological features similar to those previously described for Cell Painting[18]. (**e**) Confusion matrix of model predictions vs manually labelled classifications on model test set. (**f**) Proportion of cells in each VP35 localization category for non-targeting controls and the genes with the largest proportion of faint (NPC1), punctate (UQCRB), and peripheral (ITGB1) cells. Error bars indicate SEM across sgRNAs targeting the same gene.

https://doi.org/10.1038/s41564-025-02034-3



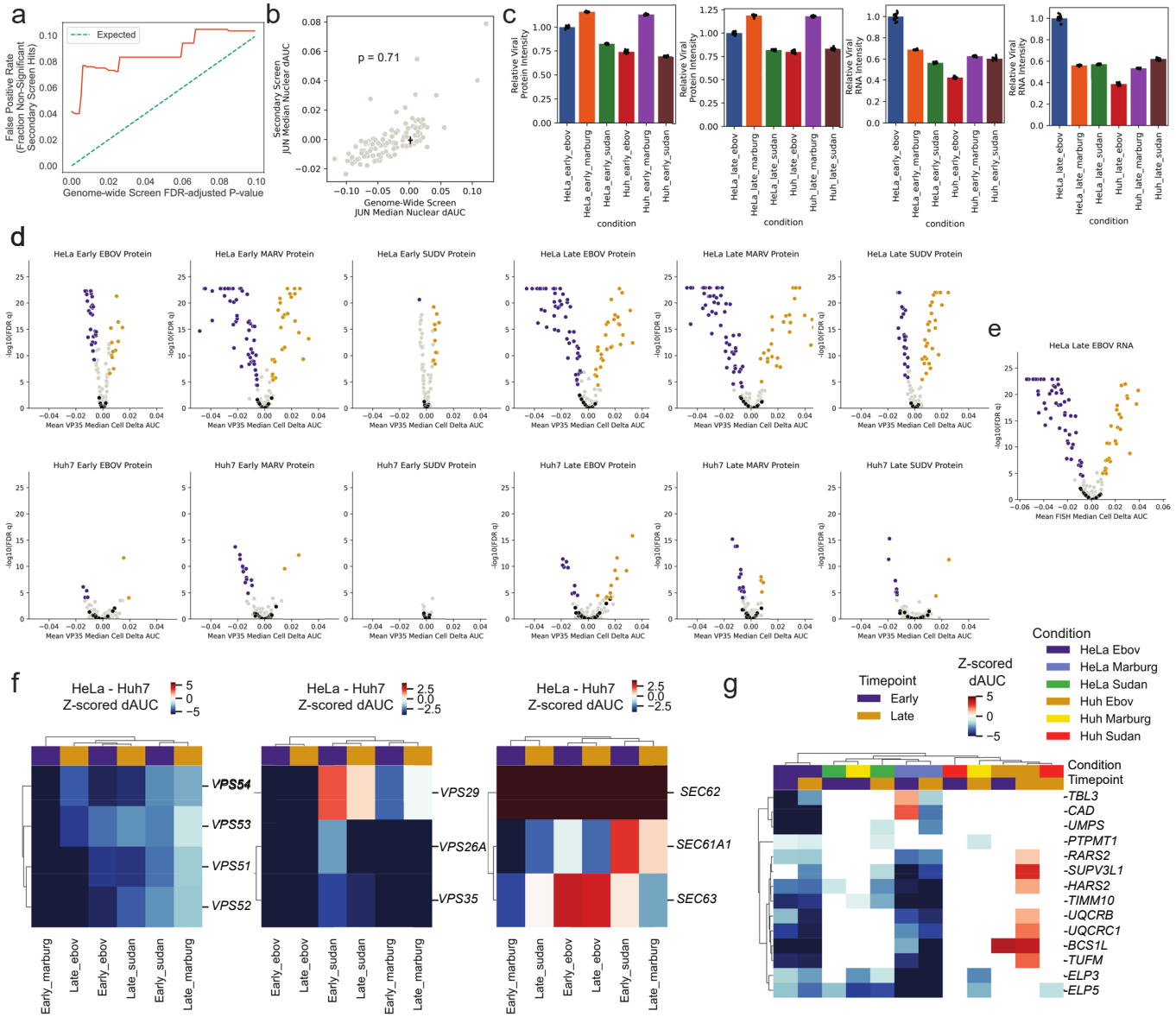
Extended Data Fig. 4 | Clustering and dimensionality reduction identify information. (**a**) Adjusted Rand score for Leiden clustering at different resolutions for 50 folds of 90% of the input data. Box plots here and in (**d**) indicate median (middle line), 25th, 75th percentile (box) and 1.5 times the IQR (whiskers) as well as outliers (single points). (**b**) Additional single-cell images of select genetic knockouts from the genome-wide optical pooled screen. (**c**) Correlation between the PHATE potential distance from the clustering using the fine-tuned model and the adjusted FDR p-value from the Kotliar study, noting genes whose expression significantly increased or decreased along with infection. (**d**) Correlation between the PHATE potential distance from the supervised clustering and the number of mass spectrometry studies that identified the genes as an interactor with an Ebola virus protein. (**e**) Venn diagram showing overlap between top optical pooled screen hits, genes that were present in at least one mass spectrometry study, and differentially expressed genes from Kotliar et al's single-cell RNA sequencing study. (**f**) Correlation between the PHATE potential distance from the supervised clustering and the 95th percentile z-score for each gene in other virus genetic screens.

https://doi.org/10.1038/s41564-025-02034-3



Extended Data Fig. 5 | Summary of analytical approaches used in the manuscript. (**a**) Workflow of analytical approaches used in this manuscript.

https://doi.org/10.1038/s41564-025-02034-3



Extended Data Fig. 6. Additional secondary screen metrics. | (**a**) Fraction of non-significant secondary screen hits (p >=0.05) at varying thresholds of the genome-wide FDR-adjusted p-value for the HeLa late infection timepoint VP35 protein intensity. (**b**) Correlation between genome-wide c-Jun median nuclear delta AUC scores and secondary screen delta AUC scores; black lines indicate standard deviation for non-targeting control sgRNAs in each screen centred around the mean value for non-targeting sgRNAs in the screen. (**c**) Secondary screen mean viral protein (VP35 for EBOV and SUDV or VP40 for MARV) and RNA intensities in non-targeting control sgRNAs relative to HeLa cells infected with EBOV. (**d**) Volcano plots for VP35 (EBOV, SUDV) or VP40 (MARV) protein expression in each of the twelve screening conditions. (**e**) Volcano plot for viral VP35 RNA levels in HeLa cells at the late timepoint condition. (**f**) Heatmaps showing the difference between HeLa cell and Huh7 cell z-scored delta AUCs for members of the GARP, retromer, and the Sec61 complex. Hierarchical clustering performed using Euclidean distance. (**g**) Heatmap showing z-scored delta AUC values for genes identified as enriched for a punctate phenotype in the genome-wide screen and also included in secondary screens (white cells indicate conditions where p > 0.05). z-scored dAUC values for VP35 or VP40 protein were calculated on delta AUC values for all genes in each screen condition relative to means and standard deviations for non-targeting sgRNAs. White cells indicate conditions where p > 0.05 relative to non-targeting controls in the same condition. Hierarchical clustering performed using Pearson correlations.

nature portfolio

Corresponding author(s): Robert Davey and Paul Blainey

Last updated by author(s): Feb 5, 2025

Reporting Summary

Nature Portfolio 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 Portfolio 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	All code is available through our GitHub repository at https://github.com/beccajcarlson/EBOVOpticalPooledScreen
Data analysis	Data analysis was performed using Cell Profiler (version ##), Graph pad Prism (version ##), NIS-ELEMENTS AR-MQS31000, and custom code available on GitHub.

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 Portfolio 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 description of any restrictions on data availability
- For clinical datasets or third party data, please ensure that the statement adheres to our policy

Raw screen images and single-cell features are available on Google Cloud Storage at https://console.cloud.google.com/storage/browser/opspublic-east1/EBOVOpticalPooledScreen.

Research involving human participants, their data, or biological material

Policy information about studies with human participants or human data. See also policy information about sex, gender (identity/presentation), and sexual orientation and race, ethnicity and racism.

Reporting on sex and gender	N/A
Reporting on race, ethnicity, or other socially relevant groupings	N/A
Population characteristics	N/A
Recruitment	N/A
Ethics oversight	N/A

Note that full information on the approval of the study protocol must also be provided in the manuscript.

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	There was no specific statistical method used to determine sample size.
Data exclusions	All data were included. Exceptions were when microscope images were of poor quality due to foreign material in the image (dust or similar debris).
Replication	The primary analysis identified hundreds of cells expressing the same sgRNA by barcode and looked for similar phenotypic outcomes. For those that gave statistically significant outcomes (FDRs are indicated in the text), were included as hits. We replicated all secondary hits at least 3 times. We validated outcomes in multiple cell types and are described in the text and figures.
Randomization	The screen is inherently random as there is no way to predict which cells will express a specific sgRNA from the library. A population of cells expressing all sgRNA variants is evaluated at once in the same wells and across multiple wells. Secondary and tertiary assays involve selection of sgRNAs to be evaluated but are plated in random wells across plates and slides without specific bias. Biases in outcomes based on position on plates and slides are ruled out during assay construction.
Blinding	The AI assessed evaluation of cell phenotypes involved blinded evaluation of outcomes to select a set of 4 groups distinguishable by humans. This training set was then applied by the AI system without it understanding the differences between the image sets and so is blinded. Evaluation of predictive power was performed by statistical testing without human bias in decisions of significance.

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
☒	☐ Clinical data
☒	☐ Dual use research of concern
☒	☐ Plants

Methods

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

Antibodies

Antibodies used	Primary antibodies: c-Jun (1:1800 dilution, Cell Signaling Technology Cat# 9165, RRID:AB_2130165), vimentin (1:1300 dilution, Abcam Cat# ab24525, RRID:AB_778824), AlexaFluor 488-conjugated LAMP1 (Cell Signaling Technology Cat# 58996, RRID:AB_2927691), and VP40 (1:1000, Integrated BioTherapeutics 0203-012). The VP35 antibody was obtained from Dr. Leung, Washington University and is available upon request. Secondary antibodies: Donkey anti-mouse antibody (Jackson ImmunoResearch Labs Cat# 715-006-151, RRID:AB_2340762) disulfide-linked to Alexa Fluor 594 (Thermo Fisher A10270) via a custom conjugation, 1:1800 donkey anti-rabbit antibody (Jackson ImmunoResearch Labs Cat# 711-006-152, RRID:AB_2340586) disulfide-linked to Alexa Fluor 647 (Thermo Fisher Scientific A10277) via a custom conjugation, and goat anti-chicken Dylight 755 (Thermo Fisher Scientific SA5-10075).
Validation	Validation of the antibodies has been reported by the commercial suppliers. In addition, c-Jun, vimentin, and LAMP1 scored highly for reduced fluorescence upon knockout in our pooled screen, providing internal validation of these signals. For the VP35 antibody, this was independently validated in this paper with cells lacking virus showing no signal and those with virus showing strong staining that is represented by inclusion bodies characteristic of virus structures in infected cells.

Eukaryotic cell lines

Policy information about cell lines and Sex and Gender in Research

Cell line source(s)	HeLa-TetR-Cas9 cells were obtained from Iain Cheeseman's lab at the Whitehead Institute. Huh7 cells ###
Authentication	*Describe the authentication procedures for each cell line used OR declare that none of the cell lines used were authenticated.*
Mycoplasma contamination	We use PCR-based mycoplasma testing. Cell lines were negative.
Commonly misidentified lines (See ICLAC register)	Cell lines were not authenticated.

Plants

Seed stocks	N/A
Novel plant genotypes	N/A
Authentication	N/A