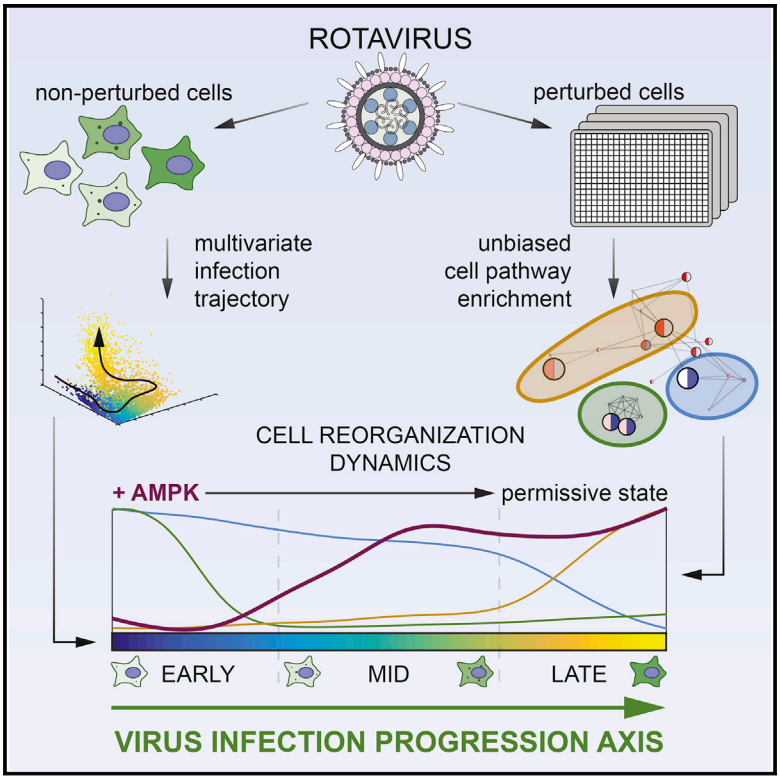
Resource

Cell Host & Microbe

A Systems Survey of Progressive Host-Cell Reorganization during Rotavirus Infection

Graphical Abstract



Highlights

- High-quality genome-wide resource of host pathways involved in rotavirus infection
- Mapping of cellular reorganization dynamics along an infection progression trajectory
- Trajectory shifts reveal ordered host-factor functions during virus infection
- AMPK promotes a cellular environment permissive to rotavirus infection

Authors

Victoria A. Green, Lucas Pelkmans

Correspondence

victoria.green@uzh.ch (V.A.G.),
lucas.pelkmans@imls.uzh.ch (L.P.)

In Brief

Genetic perturbation screens generate large maps of host factors involved in pathogen infection but without dynamic information. Green and Pelkmans now harness infection heterogeneity to map newly identified host factors and pathways along a rotavirus infection progression trajectory, and find ordered, widespread host reorganization that creates a replication-permissive cellular state.

Green & Pelkmans, 2016, Cell Host & Microbe *20*, 107–120
July 13, 2016
http://dx.doi.org/10.1016/j.chom.2016.06.005

CellPress

Cell Host & Microbe
Resource

A Systems Survey of Progressive Host-Cell Reorganization during Rotavirus Infection

Victoria A. Green[1,*] and Lucas Pelkmans[1,*]
[1]Faculty of Sciences, Department of Molecular Life Sciences, University of Zurich, Zurich, Switzerland
*Correspondence: victoria.green@uzh.ch (V.A.G.), lucas.pelkmans@imls.uzh.ch (L.P.)
http://dx.doi.org/10.1016/j.chom.2016.06.005

SUMMARY

Pathogen invasion is often accompanied by widespread alterations in cellular physiology, which reflects the hijacking of host factors and processes for pathogen entry and replication. Although genetic perturbation screens have revealed the complexity of host factors involved for numerous pathogens, it has remained challenging to temporally define the progression of events in host cell reorganization during infection. We combine high-confidence genome-scale RNAi screening of host factors required for rotavirus infection in human intestinal cells with an innovative approach to infer the trajectory of virus infection from fixed cell populations. This approach reveals a comprehensive network of host cellular processes involved in rotavirus infection and implicates AMPK in initiating the development of a rotavirus-permissive environment. Our work provides a powerful approach that can be generalized to order complex host cellular requirements along a trajectory of cellular reorganization during pathogen invasion.

INTRODUCTION

Pathogen invasion is often accompanied by widespread host cellular changes, caused by the pathogen to promote replication and spread, and by the host cell to fight the invasion. Genetic perturbation screens have been powerful in revealing large sets of candidate host factors (Cherry et al., 2005; Pelkmans et al., 2005). Combined with databases on functional genetic interactions, such datasets have allowed the creation of maps of host factors linked to various aspects of cellular physiology (Ryan et al., 2013), but generally these maps do not place the complexity of host factors in the context of ordered infection progression. Trajectories derived from fixed cell populations, utilizing multivariate features of single cells, have proved a powerful approach for the ordering of cells along a dynamic process, including cell differentiation (Bendall et al., 2014) and the cell cycle (Gut et al., 2015), but have not been applied to pathogen infection progression.

By combining large-scale perturbation screens with infection progression trajectories, we here generate an ordered view of the widespread cellular changes induced by rotavirus infection. Rotavirus is the leading cause of gastroenteritis in children, causing significant morbidity worldwide. It is a non-enveloped, triple-layered particle (TLP) encompassing the 11 segments of the dsRNA genome, which encode six structural (VP1–6) and six non-structural (NSP1–6) proteins. Rotavirus infects mature enterocytes of the small intestine. Following endocytosis, calcium flux from the endocytic compartment, facilitated by the proton gradient generated by the vesicular ATPase (v-ATPase) (Chemello et al., 2002), leads to release of a transcriptionally competent double-layered particle (DLP) into the cytosol of infected cells. The active viral polymerase complex (VP1-VP3) generates capped, positive-stranded viral mRNAs, which serve as templates for both viral protein synthesis and new genome segments on minus-strand synthesis. New DLPs are assembled in RNA/protein-rich cytoplasmic aggregations termed viroplasms, from which the newly formed DLP buds into ER-derived membranes in which the TLP is assembled. Several host factors have been linked to rotavirus infection: in entry, glycosphingolipid-modified cell attachment factors (Martínez et al., 2013), ESCRT machinery (Silva-Ayala et al., 2013), dynamin-2, and caveolin-1 (Gutiérrez et al., 2010); casein kinase 1 alpha is required for viroplasm formation (Eichwald et al., 2004), and calcium/calmodulin-dependent protein kinase beta-AMP kinase (AMPK) signaling is important for viroplasm membrane recruitment (Arnoldi et al., 2014; Crawford et al., 2012). However, a global analysis of human host factors and associated cellular processes required for rotavirus infection is lacking.

Using different gene silencing approaches in human intestinal cells, we generate a high-confidence genome-wide view on human host factors involved in rotavirus infection. Through probabilistic aggregation of the multiple datasets, we generate an unbiased systems-level map of the host cellular processes involved, including RNA processing and translation, mTOR and MAPK signaling, organelle and membrane organization, and lipid metabolism. We then developed a complementary approach that harnesses cell-to-cell variability in infection progression to infer a trajectory of virus infection progression from a fixed population of cells. This revealed an ordered reorganization of host cellular processes important for infection progression. Finally, by integrating the large-scale gene perturbation dataset with single-cell trajectories, we propose a model for host-factor-mediated cellular reorganization during rotavirus infection, many elements of which can be linked to calcium-activated AMPK signaling.

RESULTS

RNAi Screens for Rotavirus Infection Are Reproducible, Robust, and Unbiased

To identify host genes involved in rotavirus replication, we performed arrayed, image-based high-throughput RNAi screens in

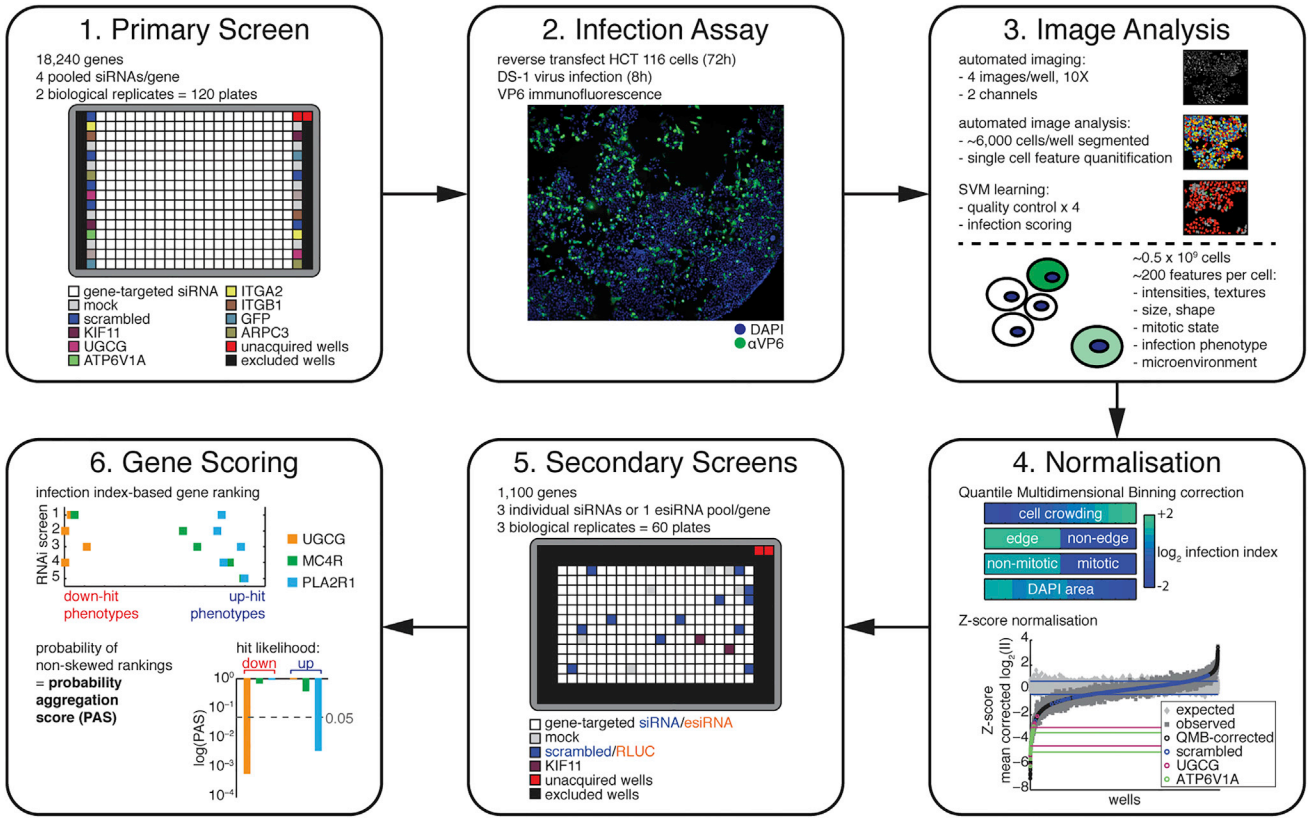


Figure 1. Genome-Wide Genetic Perturbation Screens for Rotavirus Infection

Overview of the experimental and computational workflow of the RNAi screens analyzed in this study. See also Supplemental Experimental Procedures. See also Figure S1 and Table S1.

human HCT 116 cells (Figure 1). The first screen consisted of a genome-wide set of siRNA pools and was performed in duplicate (Table S1). The secondary screen, for result validation, consisted of three independent siRNAs and an esiRNA per gene in triplicate. Our experimental and computational workflows were highly reproducible (Figure S1A), and our steps of data normalization and confounding factor reduction markedly improved host factor detection (see Supplemental Experimental Procedures; Figures S1B–S1E). No plate bias was detected (Figures S1F–S1I), and correlations between datasets, in both infection index (II) and total cell number (TCN), were consistently high (Figures S1J–S1L).

We integrated the information from all five screen datasets into two probability aggregation scores (PAS) for an individual gene, PAS_{down} and PAS_{up}, indicating the likelihood the gene is required for, or inhibits, infection, respectively, with values nearest to zero indicating a stronger likelihood (see Supplemental Experimental Procedures; Figure S1M; Table S2). The PAS outperformed other screen aggregation methods (Figures S1N and S1O), and the results displayed higher validation rates than previous genome-wide RNAi screens on pathogen infection (Figure S1P) (Hao et al., 2013; Rämö et al., 2014). Finally, our experimental validation with multiple RNAi effectors was more effective at host factor identification than correcting for off-target siRNA "seed effects" (Figures S1Q–S1T).

Probabilistic Aggregation of Multiple Screens Reveals Rotavirus Host Factors

Several genes with well-known roles in rotavirus infection were identified as host factors by our screening approach: *DNM2*, *CAV1* (Gutiérrez et al., 2010), *CSNK1A1*, *CSNK2A1* (Eichwald et al., 2004; Eichwald et al., 2002), and *CALM2* (Crawford et al., 2012) (Figures 2A and S2A). No infection phenotype was observed on knockdown of *CAMKK2* (Figure 2B), but there was a strong infection phenotype observed on depletion of *CAB39* (Figure S2A), another AMPK activator. In accordance with the only published rotavirus RNAi screen, using human siRNAs in African green monkey cells (Silva-Ayala et al., 2013), two negative regulators of the ESCRT machinery, *PTPN23* and *STAMBP*, were identified as up-hits (Figures 2A and S2A). Although not all AMPK- and v-ATPase-associated genes gave rise to a phenotype on perturbation (3/7 or 12/21 genes, respectively; Figure S2B), we obtained infection or viability phenotypes for at least one isoform for 67% and 85% of AMPK and v-ATPase subunits, respectively (Figure 2B), with all infection phenotypes scoring low PAS values (Figure 2A). Thus, our data identify known host factors, machinery regulators, and multiple subunits of complexes required for infection.

For several of the rotavirus host factors identified here, *ABCF1*, *COPG1*, *WDR46*, and *DDX52*, we confirmed that they also reduce infection in Caco-2 cells (Figure S2D) and confirmed that knockdown of expression of these genes in HCT 116 cells

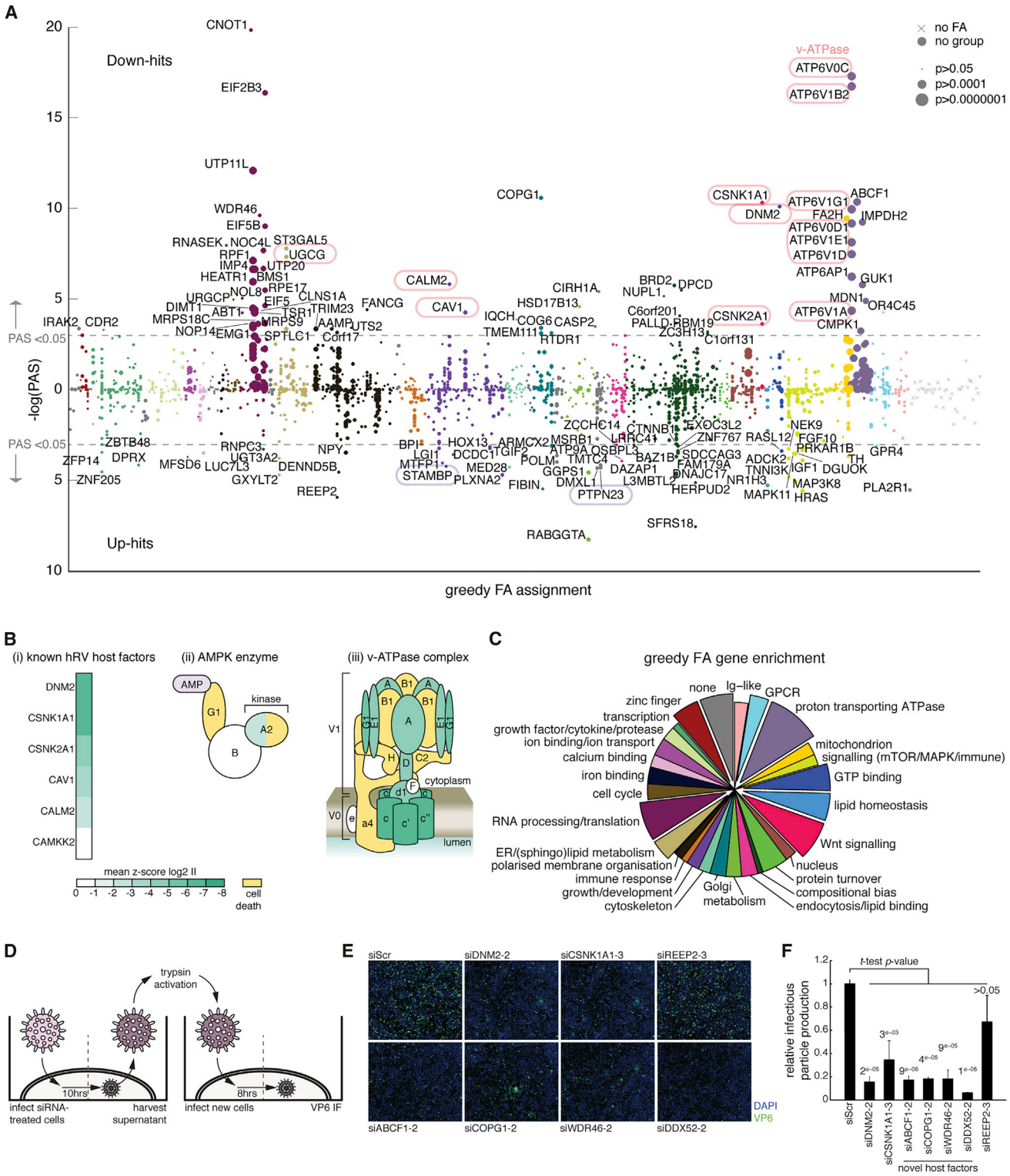


Figure 2. Probabilistic Aggregation of RNAi Screens Reveals Host Factors in Rotavirus Infection

(A) Overview of the gene phenotype scores. A single, negative natural logarithm of PAS value, the smallest of PAS_{down} and PAS_{up}, was plotted per gene, with the latter values on an inverted axis. Dashed gray lines correspond to a PAS of 0.05. Genes are grouped on the x axis based on similar functional annotations, derived from a greedy functional annotation (FA) assignment. Circle size indicates the FA enrichment score (FAES) of the FA to which the gene is greedily assigned. Previously reported down- (red) and up-hits (blue) are circled.

(legend continued on next page)

also significantly inhibits infectious particle production (Figures 2D–2F and S2E). Together with the high reproducibility, robustness, and consistency of the dataset, this demonstrates that we provide a powerful resource of rotavirus host factors that both facilitate and inhibit infection.

Functional Analyses of Host Determinants of Rotavirus Infection

To visualize cellular pathways that genes with the strongest infection phenotypes are associated with, genes were assigned to a single functional annotation (FA) using a greedy assignment method (Figure 2A; Table S3). Although this retrieves enrichment of the v-ATPase (Figure 2C), it ignores PAS values. Performing a threshold-based (PAS < 0.1) FA enrichment analysis also revealed some expected results (e.g., translation, vesicular transport, and the v-ATPase) (Table S5). However, such a method does not consider the distribution of gene rankings within a FA group, rather equally weighting genes beyond an arbitrary threshold. It may thus ignore much of the information present in such datasets.

Therefore, we used a rank-based approach in which the aggregated gene data was used to calculate the probability of whether each FA listed in the DAVID database is significantly enriched, as given by the FA enrichment score (FAES), in genes reducing ($FAES_{down}$), as well as increasing ($FAES_{up}$), infection (see Supplemental Experimental Procedures; Figure S3A; Table S4). The most enriched FAs were then used to construct a network, in which FA nodes were connected by edges representing the degree of gene overlap between FAs. Incorporating all screen data in this unbiased way, cellular processes important to infection are readily detected as clusters of significantly enriched, functionally related annotations (Figure 3; browse at http://rotavirus.infectome.org).

From such a rank-based, probabilistic network analysis, we recovered enrichment for numerous expected FAs not found to be significant ($p < 0.05$) by the threshold-based approach, including calcium-binding region ($FAES_{up} = 0.0035$) and WNT Signaling Pathway ($FAES_{down} = 0.0028$) (Figure 3; Tables S4 and S5). Furthermore, the FA network reveals significant enrichment (FAES < 0.05) for annotations related to vesicular transport and membrane and organelle localization that were unexpected, including annotations related to the endoplasmic reticulum (ER) (Figure S3A), Golgi complex, and mitochondrion. In addition, there is significant enrichment for annotations associated with lipid synthesis and storage, translation and RNP binding, and MAPK and mTOR signaling (Figure S3B). Interestingly, host factors with a role in viral infection are enriched for compositionally biased low-complexity domains (LCDs), including lysine, serine/threonine, and acidic (Asp/Glu-rich) regions, which are associated with RNA binding, regulation by phosphorylation, and casein kinase substrates, respectively. Also intriguing was the enrichment for RNA degradation, particularly given that rotavirus replication involves the generation of capped viral mRNAs. Thus, the network provides a valuable tool for the discovery of pathways important for rotavirus infection, revealing roles for multiple cellular processes, suggestive of widespread reorganization and redirection of cell resources on infection.

Trajectories of Infection Progression from Heterogeneous Fixed Populations of Cells

The order by which such large-scale cellular reorganization occurs during the infection process remains unclear. In our attempts to address this, we noticed that virus infection progression rapidly becomes asynchronous (even in cells washed 1 hr after virus exposure), with both the number of NSP5 spots and NSP5 intensity within cells varying significantly at 9 hr post-infection (p.i.) (Figure 4A), as well as the abundance of viral transcripts at 4 hr p.i. (Figures S4A and S4B). This heterogeneity can confound attempts to reliably quantify cellular changes over the course of infection in multiple cells, even if using live cell imaging. We therefore hypothesized that a time-lapse-independent method that can (a) harness the variability in infection progression among single cells in a fixed cell population and (b) order them along an axis of infection progression may be more accurate.

To identify which single-cell features best reflect infection progression, we clustered time course mean values of various measurements, including texture, intensity, and spot features of the NSP5 staining, as well as general cellular features such as nuclear and cell size (Figures S4C and S4D; Table S6). We selected features changing over time without highly overlapping single-cell distributions. NSP5 concentration showed the least overlap between 6 and 9 hr p.i. of any feature, while, in contrast, NSP5 texture features that capture the emergence of early viroplasms showed least overlap between early time points (Figure S3D). Thus, combined, multiple features allow an ordering of infected single cells along a trajectory, or virus infection progression axis (VIX), that reproducibly resolves different stages of the infection process better than time alone (Figures 4B–4D and S4E–S4G; see Supplemental Experimental Procedures). The distribution over time of the resulting VIX values behaved as expected, with earlier time points (3–6 hr p.i) heavily enriched for cells with low VIX values, which were depleted at later time points (9–15 hr p.i; Figure 4E). We therefore used trajectory feature dynamics, rather than time, to define three main phases of infection progression: early (0–0.3), mid (0.3–0.7) and late (0.7–1).

Next, we tested whether the cell-to-cell variability in infection progression within a single fixed population is sufficient to construct a trajectory. Indeed, at 8 or 9 hr p.i, the VIX is identical to one obtained from all time points (Figures 4E and S4H). In addition, two infection features not used in trajectory

(B) Summary of results for known rotavirus host factors. The mean of GW duplicate corrected log2 II for known human rotavirus (hRV) host factors (i)–(iii), with a selected isoform for each subunit of the AMPK enzyme (ii) and v-ATPase complex (iii). For the α-2 subunit of AMPK, the two colors refer to replicates 1 and 2. Yellow indicates gene perturbation resulted in a significant loss of cell viability (TCN < 625).
(C) Enrichment for grouped functional annotations among validation genes. The proportion of validation genes greedily assigned to the different FA groups, with exploded segments indicating a higher proportion in the validation screen than in the genome.
(D) Schematic of the experimental approach for image-based quantification of infectious particle production.
(E) Selected images of cells infected using virus produced from cells with reduced host factor expression.
(F) Knockdown of down-hits reduces infectious particle production. Single infected cells on infection with virus from cells with host factor knockdown, relative to siScrambled control. Depicted are mean ± SD of quadruplicate wells.
See also Figure S2 and Tables S2 and S3.

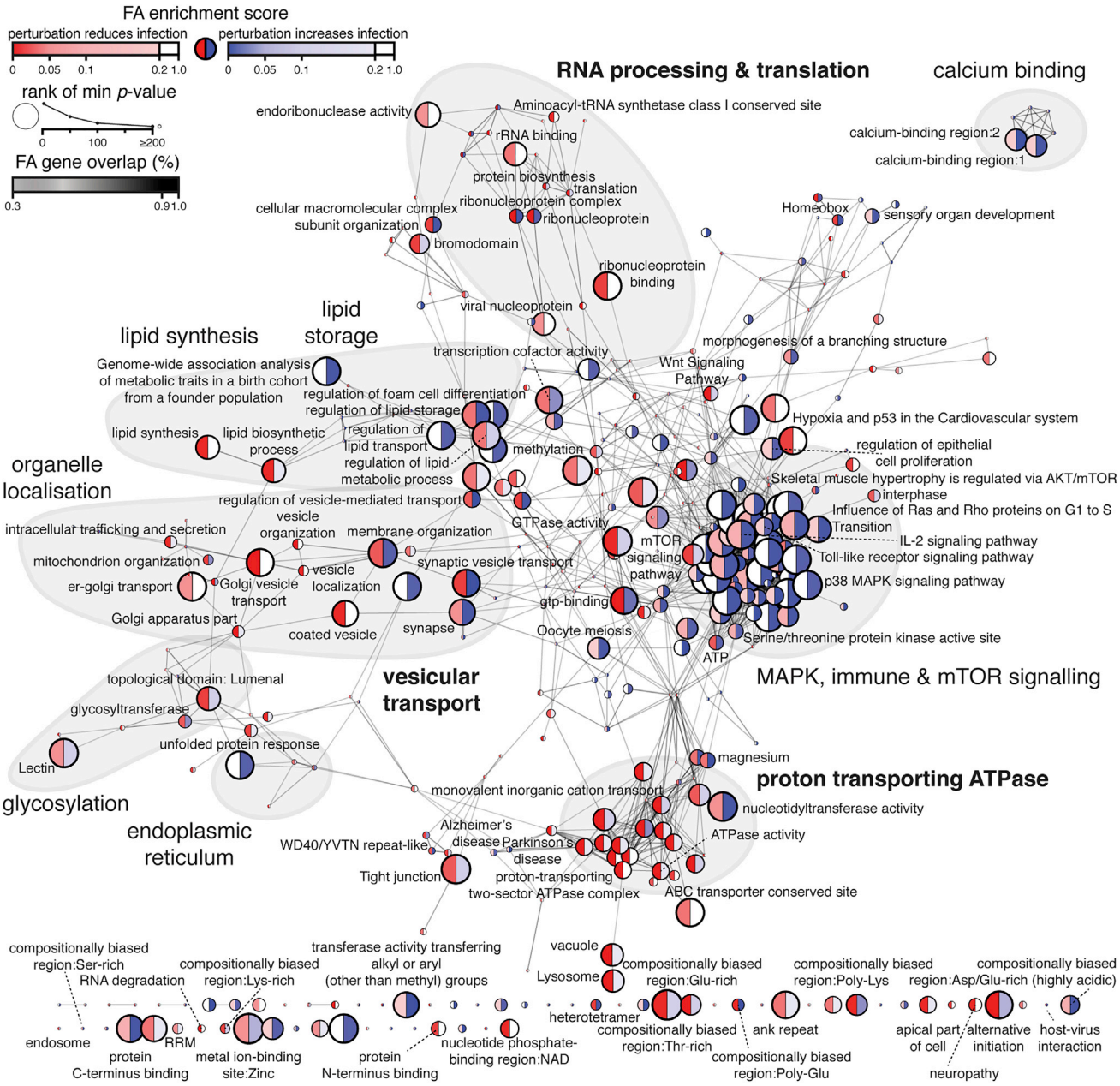


Figure 3. Multiple Cellular Processes Are Determinants of Rotavirus Infection

Overview of significantly enriched functional annotations (FAs) derived from the aggregated screening data. FA enrichment scores (FAES) were derived from PAS-based rankings of both down- and up-hits for all FAs listed in the DAVID database. Those with more than 20 genes, and a $FAES_{down} \leq 0.05$ (red) or $FAES_{up} \leq 0.02$ (blue), are displayed as nodes. Node size indicates the rank at which the FAES, a minimum p value, was obtained. Edges indicate the percentage of gene overlap between FAs. Functionally related FAs are manually clustered into clouds. Those clouds containing FAs also found as significantly enriched by a threshold-based enrichment approach (PAS < 0.1) are indicated in bold.

See also Figure S3 and Tables S4 and S5.

construction, namely an SVM-based classification of late-infected cells, and the concentration of VP6, were ordered more accurately along a VIX derived from fixed cells at 9 hr p.i only, than by monitoring their behavior over a time course experiment of 0–15 hr p.i (Figure 4F). Thus, our trajectory allows single-cell activities to be accurately quantified along infection progression without the need for laborious time course assays.

Virus infection Trajectories Reveal Widespread Physiological Changes in Host Cells with Variable Dynamics

To characterize alterations in the pathways and organelles identified from the FA enrichment analyses (Figure 3), we quantified a number of relevant single-cell readouts and plotted them along the VIX (Figures 5A and 5B). Membrane organization was

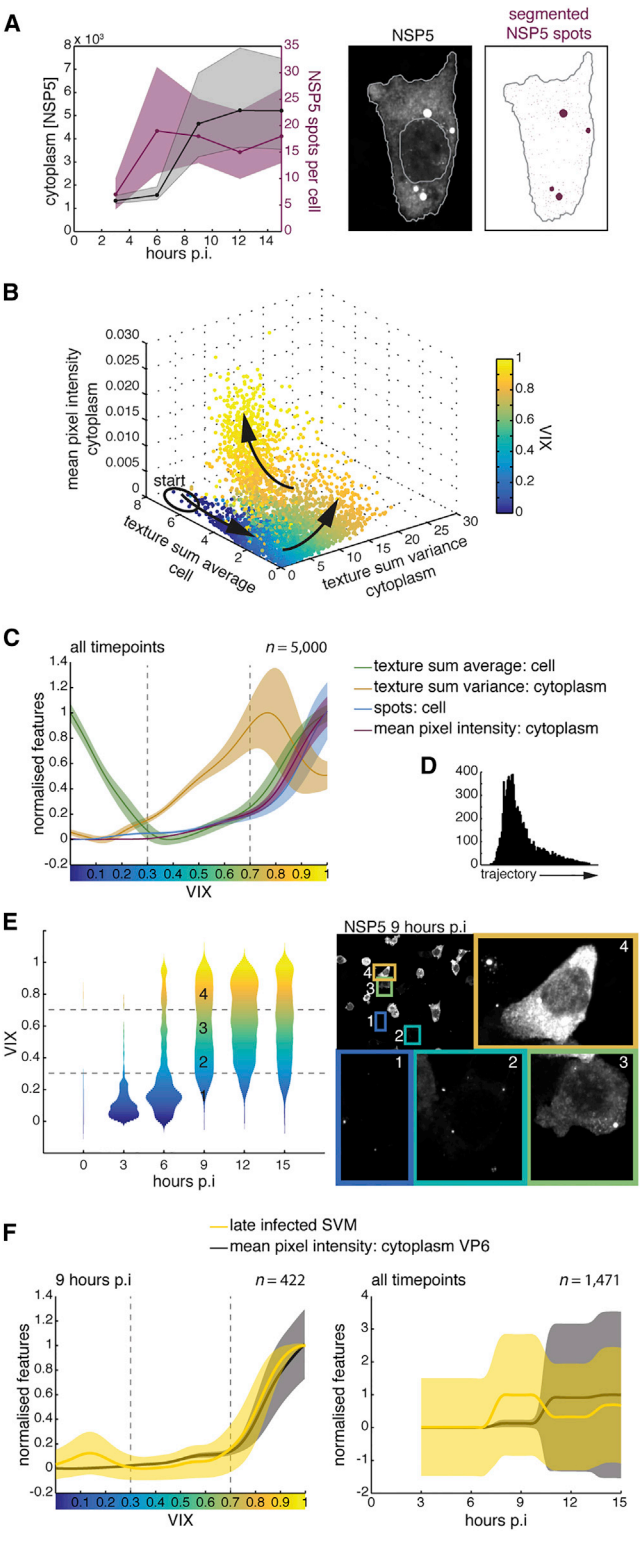


Figure 4. Asynchronous Infected Cells Can Be Ordered along a Virus Infection Progression Trajectory

(A) Rotavirus infection progression is heterogeneous. HCT 116 cells were assayed for NSP5 intensity (black) and spots (purple) at 3 hr intervals post-infection (p.i.). Medians are shown, with single-cell quantiles 37.5 and 62.5 shaded.

monitored with markers for the ER (calreticulin) and Golgi (giantin). This revealed an alteration in the reticular structure of the ER during early- and mid-infection phases, quantified as a reduction in the cytoplasmic contrast of the immunofluorescence signal. Contrast slightly increased during the late phase, reflecting the increase in structure of calreticulin-positive membranes upon their wrapping of viroplasms. Golgi membranes were also altered, specifically in late-infected cells, changing from the typical perinuclear ribbon-like organelle to dispersed fragmented membranes. We also observed an increase along the VIX in the abundance of the GTPase DNM2 (dynamin 2) and, as reported previously, autophagosome marker LC3b (Figure S5A) (Arnoldi et al., 2014; Crawford et al., 2012).

The FA analyses also revealed enrichment for mitochondria organization and lipid homeostasis, perhaps reflecting coordinated alteration of cellular metabolic pathways for efficient viral replication. To examine this, we monitored mitochondrial outer membrane protein TOMM20 and stained lipid droplets with BODIPY (Figures 5A and 5B). Indeed, a strong decrease in perinuclear TOMM20 signal contrast was observed, revealing a mitochondrial accumulation there during mid- and late-stage infection. In addition, lipid droplets, highly variable in the HCT 116 population, undergo rapid depletion upon infection.

Genes functionally annotated to protein translation are strongly enriched among host factors for rotavirus infection (Figure 3). It has been reported that non-polyadenylated viral mRNAs are preferentially translated over poly-adenlylated cellular mRNAs (López and Arias, 2012), primarily as a result of NSP3 displacement of poly(A) binding protein (PABPC1) from translation initiation factor eIF4G1, and nuclear retention of PABPC1 and cellular poly(A) mRNAs (Harb et al., 2008; Rubio et al., 2013). To examine the dynamics of global translation, we quantified de novo protein production, levels of phosphorylated ribosomal protein S6 (a proxy for mTORC1 activity), eIF4G, and PABPC1 along the VIX (Figures 5A–5B and S5A): de novo protein synthesis steadily decreased throughout infection progression, concomitant with

(B) Infection progression is captured by multivariate single-cell features. Data from 5,000 single cells was colored according to position along the virus infection progression axis (VIX), revealing the multidimensional nature of the cell ordering. Trajectory direction from the start population, corresponding to uninfected/early-infected cells, is indicated by arrows.

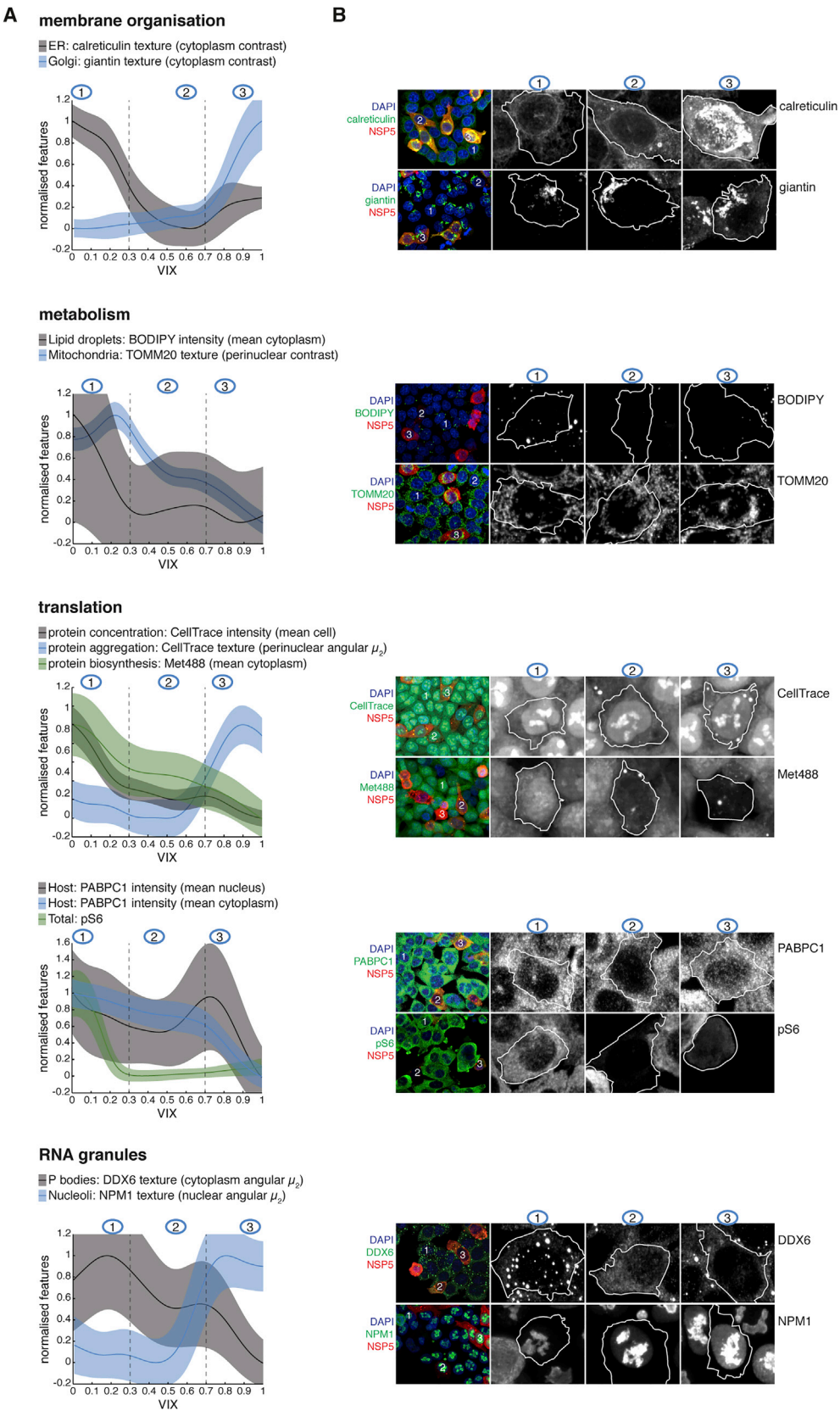
(C) Behavior of features used in trajectory construction. Weighted mean (lines) and SD (shaded) of features along a progression axis constructed from 5,000 infected cells pooled from 3–15 hr p.i., normalized between 0 and 1. Dashed lines represent an empirical separation of early, mid, and late infection stages, based on the robust behavior of these features.

(D) The distribution of cells along the infection trajectory, before ordering into a virus infection progression axis (C).

(E) The distribution of infected cells along the progression axis at 3–15 hr p.i. reflects time. The frequency of cells (violin bin width) along the VIX (colored), for each time point after infection, with early, mid, and late stages separated by dashed lines (left). A representative image of immunofluorescence for NSP5 highlights the heterogeneity in infection progression at 9 hr p.i. (right).

(F) Ordering of infection progression is improved with trajectories compared with time. Cells infected for 3–15 hr were stained for VP6 expression and, based on NSP5 measurements, classified as late infected by support vector machines. Weighted mean and SDs of both features were then overlaid on a trajectory constructed from 422 cells infected for 9 hr only (left) or plotted against time for 1,471 cells from all time points (right).

See also Figure S4 and Table S6.



(legend on next page)

phospho-S6 and, to a lesser extent, eIF4G. What new protein is produced predominantly localizes to viroplasms, accounting for the increase in protein aggregation observed during late-stage infection. Surprisingly, PABPC1 did not show an increased concentration in the nucleus of infected cells until mid-late stages, and host poly(A) mRNAs remained abundant in the cytoplasm of cells 8 hr p.i. (Figure S5B), suggesting that nuclear retention of PABPC1-bound mRNA is not the primary mechanism of early host protein synthesis shut-off in DS-1-infected HCT 116 cells.

The FAs RNP binding ($FAES_{down}$ = 0.0215) and RNA degradation factors ($FAES_{down}$ = 0.0044) were enriched in our screening dataset, along with host factors with compositional bias in their protein sequence (Figure 3). Given that proteins with biased sequences are associated with several RNA processing structures (Mitrea and Kriwacki, 2016), and the import of RNA degradation and translation machinery to rotavirus replication, we examined what happens to P bodies and nucleoli during infection using the markers DDX6 and nucleophosmin (NPM1), respectively (Figure 5). P bodies underwent a depletion during mid- and late-stage infection, as reported recently (Bhowmick et al., 2015). Interestingly, we observed low-level recruitment of DDX6 to late-stage, protein-dense viroplasms, albeit not as strongly as the known viroplasm regulator CSNK1A1 (Figure S5C). The nucleoli signal exhibited a rapid increase in overall brightness and less sub-compartmentalization mid- to late-stage infection. Furthermore, Sam68, an RNA processing factor, was displaced from the nucleus to cytoplasmic viral stress granules (V-SGs) on rotavirus infection (Figure S5A), as has been demonstrated for other viruses.

These results show that rotavirus infection induces widespread morphological and metabolic changes in cells that occur with varying dynamics at different stages of infection progression. Importantly, such changes could only be uncovered when single cells are ordered along the virus infection trajectory, since they are often invisible in averaging approaches, including those capable of separating single infected and uninfected cells (Figure S4D).

Host Factor Perturbations Shift Virus Infection Trajectories According to Gene Function

Host factors influencing rotavirus infection would be expected, on perturbation, to not just alter the fraction of bright infected cells, as detected in the large-scale RNAi screens, but also infection progression in a manner that is dependent on gene function. To avoid complex, multiple stage-specific assays, we examined if functional information could be garnered by ordering one fixed, perturbed population along the VIX.

Iterative computational mapping (see Supplemental Experimental Procedures), which showed minimal error (Figures S6A–S6C), revealed that host factor perturbations can change the infected cell distribution along the VIX, quantified by the Kolmogorov-Smirnov (*KS*) statistic. For example, siRNA knockdown of known host factor *CALM2* strongly altered the distribution of infected cells along the VIX toward early stages (*CALM2 KS* = 0.36; Figure 6A). In contrast, knockdown of CALM2-inhibitor, *CALML3*, had no effect (*CALML3 KS* = 0.007). For a subset of nine host factors, perturbing expression of eight altered infected cell distribution along the VIX; moreover in different manners, reflecting predicted gene function (Figures 6A–6E). For example, perturbation of *CSNK1A1*, required for regulation of viroplasm formation (Campagna et al., 2007), exhibits an increased proportion of cells in early- and mid-infection stages (Figures 6B and 6C), whereas RNAi of *CALM2*, required for calcium-induced signaling on infection (Crawford et al., 2012), or *DNM2*, involved in cell entry (Gutiérrez et al., 2010), showed an increase in the proportion of cells in the early stage of infection (Figures 6A–6C).

We therefore extended the VIX perturbation analysis to the rotavirus host factors, *ABCF1*, *COPG1*, *WDR46*, *DDX52*, and *REEP2* (Figures 6D, S2C, and S5D). The intercepts, between the resulting shifted perturbed populations with control cells, were superimposed on the VIX to generate a prediction of the ordered requirement for these host factors in infection progression (Figure 6E). To test this ordering, we examined early viral RNA production, early viral protein production, and the dynamics of viral RNA amplification (Figure 6F).

With single-molecule RNA FISH to viral RNA segment 5+, we could detect individual, early-stage infected cells before exponential viral RNA increases (Figures S6E and S6F). Only *DNM2* perturbation reduced detection of early viral RNA (Figure 6G), confirming *DNM2* acts before the other host factors, as predicted from computational mapping to the VIX. Early VP6 protein production was more reduced upon *ABCF1*, *WDR46*, and *DDX52* perturbation than upon *COPG1* and *CSNK1A1* perturbation (Figure 6H), in agreement with the VIX ordering. In addition, the exponential increase in viral RNA, typically between 4 and 8 hr p.i., was dramatically inhibited upon *COPG1*, *CSNK1A1*, and, to a lesser extent, *DDX52* perturbation, compared with a mere delay in late-stage viral RNA production upon perturbation of *ABCF1* and *WDR46* (Figures 6I and S6G). This validates the predicted requirement for *COPG1* and *CSNK1A1* later in infection than *ABCF1* and *WDR46* and suggests multiple roles for *DDX52*. Integrating the predicted ordering along the VIX with cellular reorganization events (Figure 6E), the infection phenotype among protein interactors (Figure S6H), and the literature, suggests functional roles for *ABCF1* in viral RNA translation initiation (Paytubi et al., 2009), *WDR46* and *DDX52* in production of the necessary ribosomes (Scherl et al., 2002), the latter also in nucleoli reorganization, and *COPG1* to allow virus-induced membrane reorganization, as well as confirming a role for *CSNK1A1* in viroplasm growth.

RNAi of *REEP2* shifts the distribution of infected cells along the VIX to the later-stage of infection (mean *KS* = 0.32; Figure 6D)

Figure 5. Membrane Organization, Metabolism, Translation, and Endogenous RNA Granules Are Altered with Varying Dynamics during Rotavirus Infection

(A) Dynamics of cellular reorganization along the VIX. HCT 116 cells were infected for 8 hr before fixation and immunostaining for cellular markers and NSP5. Infection trajectories were constructed from all infected cells pooled from duplicate wells and the weighted mean (lines) and SD (shaded) of normalized cellular features overlaid, as indicated. Early, mid, and late stages of infection are discriminated (dashed lines). The location of three example single cells derived from a single image are indicated by numbers and shown with cell segmentation outlines in (B).
See also Figure S5.

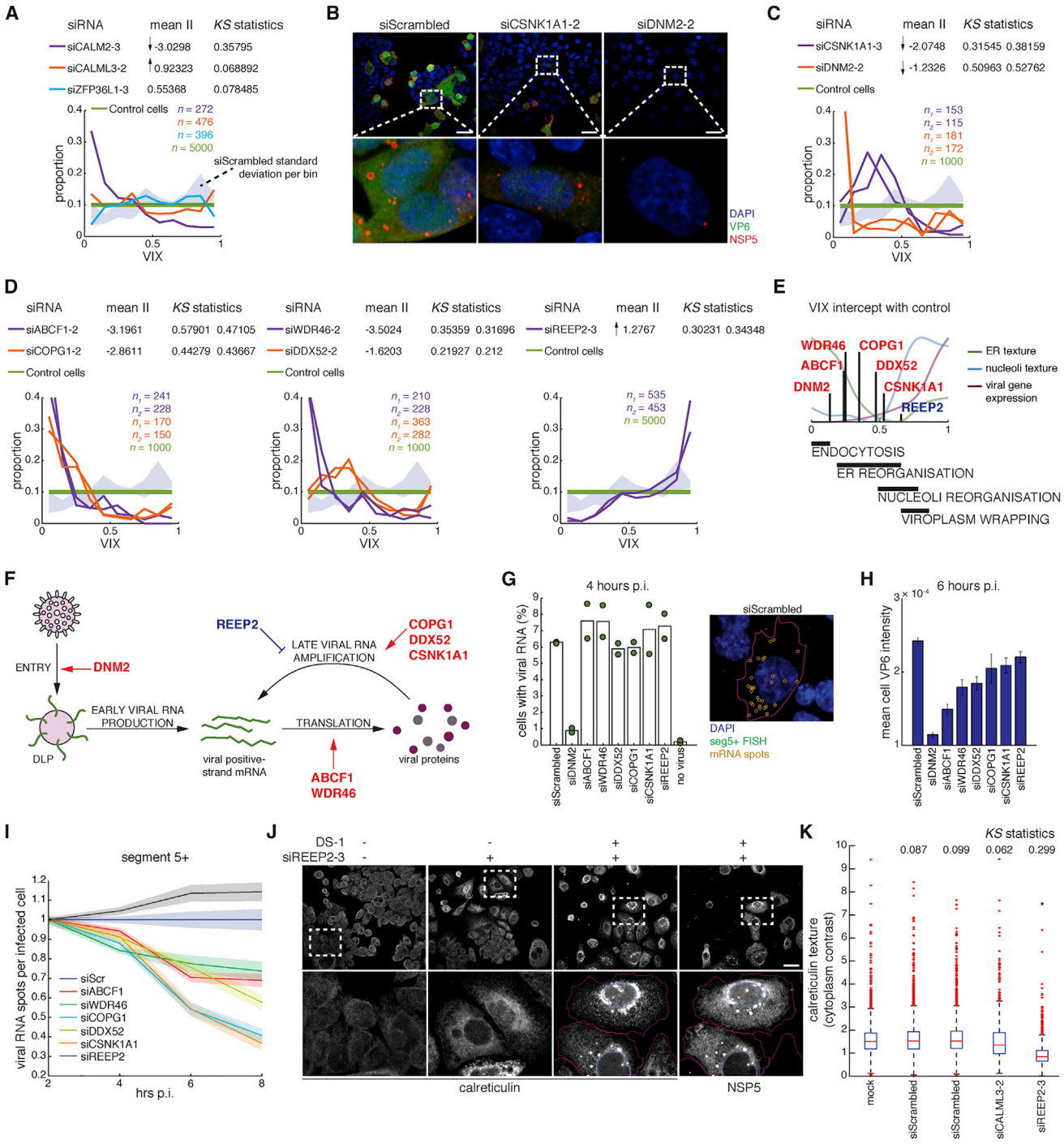


Figure 6. Mapping Gene Perturbations to Trajectories Reveals Infection Stage-Specific Host Factor Functions

(A) RNAi of host factors can change the infection index (II) and/or infection progression. 5,000 mock-transfected control cells infected for 8 hr were used in trajectory construction, along with 40 randomly selected siRNA-perturbed infected cells, for 25 bootstraps. The frequency of the resulting 1,000 perturbed cell trajectory positions in 10 VIX bins were normalized to the control population (green). Down-hit *CALM2* (purple) shifts the distribution of cells along the VIX on perturbation with siRNA-3, whereas up-hit *CALML3* (orange) with siRNA-2 and non-hit *ZFP36L1* with siRNA-3 do not. Mean II are from the secondary screen triplicates of the indicated siRNA. Extent of distribution shifts are indicated with the Kolmogorov-Smirnov (*KS*) statistic. The number of perturbed infected cells are indicated (*n*). The SD per bin of four replicate wells transfected with a non-targeting siRNA, siScrambled (shaded blue), is provided as an additional reference to control cells.

(B) Representative images of HCT 116 cells transfected with individual siRNAs and infected with rotavirus for 8 hr, corresponding to the trajectories in (C).

(C) Quantification of infected cell distribution shifts along the VIX on host factor perturbation. The proportion of infected cells in ten VIX bins are given for duplicate wells transfected with siRNA-3 targeting *CSNK1A1* (purple) or siRNA-2 targeting *DNM2* (orange), with infected cell numbers per well (*n*).

(legend continued on next page)

and, as predicted from the VIX intercept, accelerated viral RNA amplification (Figures 6I and S6G). The *REEP2* VIX intercept aligns with the point of least reticulated ER structure and the onset of extensive viroplasm membrane wrapping (Figures 5 and 6E). Furthermore, it has been reported that *REEP2* regulates ER morphology and selective autophagy (Orvedahl et al., 2011). This suggested that *REEP2* plays a role in facilitating the ER reorganization required for rotavirus infection progression.

We validated this hypothesis by demonstrating that RNAi of *REEP2* reduced visible ER structure in uninfected cells (Figures 6J and 6K), an alteration resembling that of mid-stage rotavirus infection (Figure 5), whereas knockdown of *CALML3* had no effect. Moreover, infected cells with *REEP2* perturbation exhibited ER morphology typical of late-stage infection, with the majority of calreticulin-positive membranes associated with viroplasms (Figure 5B, cell 3; Figure 6J). Together, this indicates that rotavirus infection induces an increase in the proportion of ER sheets to tubules mid-infection, which can be simulated by *REEP2* RNAi to accelerate infection progression, perhaps through enhanced ER-derived membrane delivery to viroplasms. This highlights that computational mapping of perturbed populations to multivariate trajectories can reveal at which stage of infection a host gene functions.

Multiple Aspects of Cellular Reorganization during Rotavirus Infection Can Be Linked to AMPK Activation

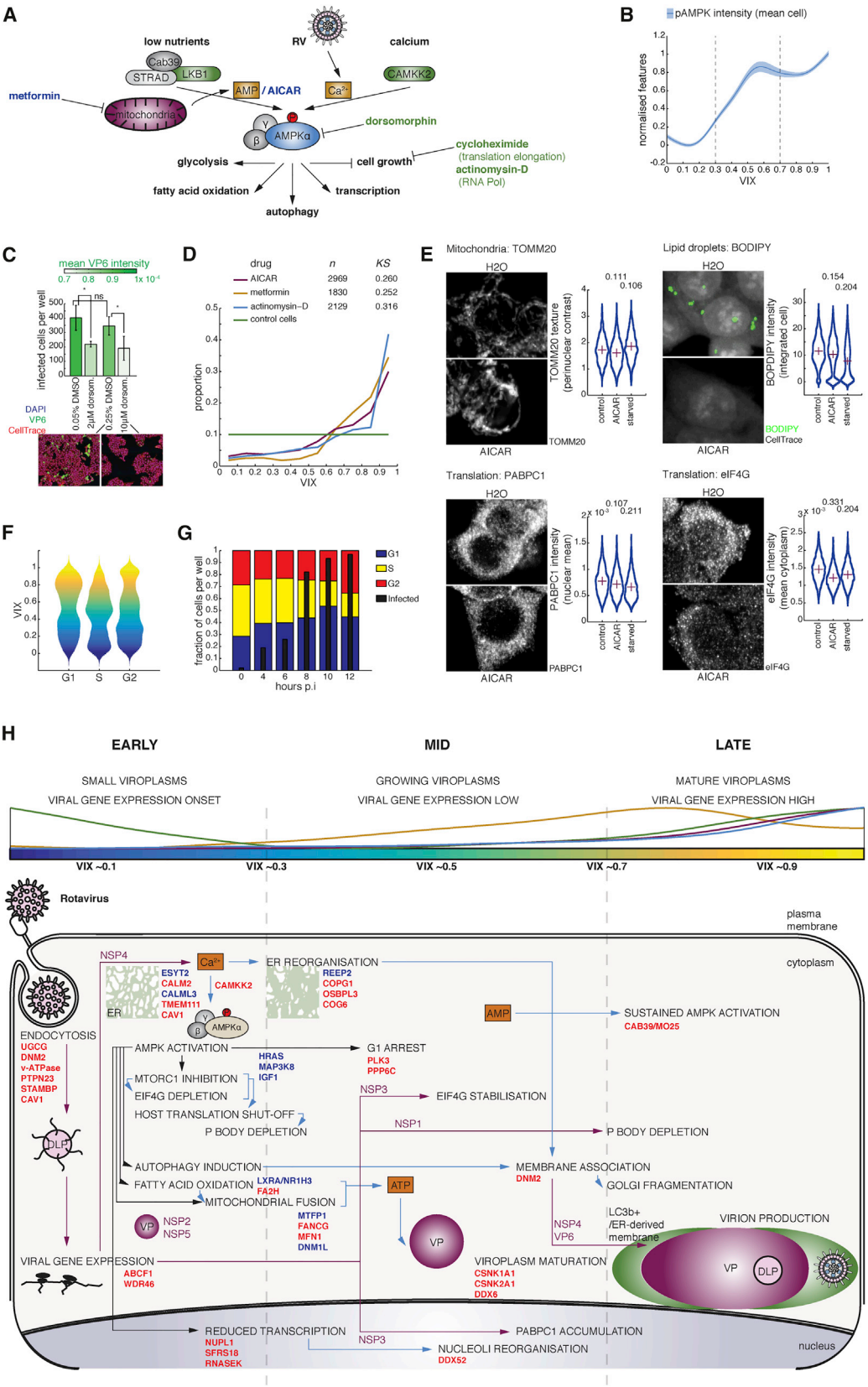
The ordered host cellular reorganization during rotavirus infection suggests that in early-stages the cellular state is altered such that late-stage infection events can happen. Many of the cellular changes commencing early in infection progression, including mTOR and translation suppression, decreased lipid droplet abundance, increased abundance of autophagy marker LC3b, and altered mitochondrial morphology (Figure 5), are associated with AMPK activation (Figure 7A) (Mihaylova and Shaw, 2011). Furthermore, AMPK may be activated by early increases in intracellular calcium on rotavirus infection (Crawford et al., 2012). Thus, we hypothesized that the AMPK signaling axis could be responsible for triggering the required cellular state.

Indeed, levels of phosphorylated (Thr172) AMPK rise from early-stage infection and remain high (Figure 7B). Since RNAi of AMPK reduced cell viability (Figure 2B), we used pharmacological perturbations to probe the requirement for AMPK activity in infection progression. Treatment with the AMPK inhibitor dorsomorphin significantly reduced the number of VP6-expressing cells and their intensity (Figures 7C and S7B). Furthermore, we observed that both a direct activation of AMPK, using the AMP analog AICAR, and an indirect activation with metformin, which inhibits mitochondrial ATP production (Figure 7A), strikingly increased the proportion of late-stage infected cells (Figures 7D, S7A, and S7B), suggesting that AMPK activation accelerates infection progression. This was supported by observations that many of the metabolic and morphological changes typically observed during early-stage rotavirus infection (e.g., mitochondrial alterations, and lipid droplet, PABPC1, and eIF4G depletion) were recapitulated in uninfected cells on AMPK activation, through treatment with AICAR or, to a lesser extent, serum starvation (Figures 7E and S7D). These effects included decreased nucleolar structure, indicative of reduced host gene transcription (Figure S7D). This can be mimicked by RNA polymerase inhibition with actinomysin-D (Figure S7C), which also accelerates rotavirus infection progression (Figure 7D) and suggests one proviral effect of AMPK may be an inhibition of host gene expression, perhaps by increasing the availability of RNA processing and translation machinery to viral transcripts.

Not all virus-induced cellular reorganization events occur downstream of pharmacological AMPK activation. P bodies did not disperse, in contrast to cycloheximide treatment (Figures S7C and S7D). Nor did we observe strong alterations in ER or Golgi morphology upon AMPK activation (*KS* < 0.2; Figure S7D). These changes may require longer AMPK activation, occur via a pathway independent of AMPK activation, or be a direct result of viral protein activity, as recently suggested for P body depletion (Bhowmick et al., 2015).

Activation of AMPK induces cell cycle arrest in the G1 phase (Jones et al., 2005) and, through synergistic effects with *EEF2K*, leads to inhibition of translation elongation specifically in G2 cells (Kruiswijk et al., 2012). Therefore, the induction of the AMPK signaling axis during early-stage infection should (a) arrest cells in G1 and (b) prevent efficient progression of infected G2 cells. To experimentally test these two predictions, we assigned cells into cell cycle phases G1, S, and G2 (excluding M) (Figure S7E). As predicted, we observed an increased proportion of infected cells in G1 compared to non-infected cells at the expense of cells in S (Figure S7F). Both S and G2 cells were depleted from late-stage infected cells compared with G1, and the fraction of G1 cells within a population exposed to rotavirus increased over time (Figures 7F and 7G). Furthermore, our

(D) Host factor perturbations have differential effects on infection progression. As for (C) but for the host factors *ABCF1*, *COPG1*, *WDR46*, *DDX52*, and *REEP2*.
(E) Summary of host factor perturbation impact on infection progression. The mean intercept of the distribution of cells along the VIX from the perturbed population with that of the control population is indicated, with regions of the trajectory where particular cellular changes were observed superimposed (see also Figure 5).
(F) Model for ordered host factor function in rotavirus infection, based on VIX intercepts in (E).
(G) *DNM2* knockdown prevents early viral RNA production. Perturbed populations were analyzed by single-molecule RNA FISH to viral RNA segment 5+ (seg5+) 4 hr p.i. Bars depicting mean of duplicates (green circles) are shown (left), with a representative image of seg5+ spot detection (right).
(H) Host factor perturbation differentially effects early viral protein production. Immunofluorescence for VP6 on siRNA-treated cells infected for 6 hr. Mean ± SEM of single-cells from duplicate wells.
(I) Host factor perturbation differentially effects late viral RNA amplification. As for (G) but at all time points indicated, and with number of spots per infected cell normalized to siScrambled (mean ± SEM), shown.
(J) REEP2 perturbation alters ER morphology and accelerates infection progression. Representative images of calreticulin IF in uninfected cells ± siREEP2-3, and both calreticulin and NSP5 IF in infected cells with siREEP2-3.
(K) REEP2 perturbation reduces ER texture. Quartiles of single cell data from uninfected wells. * indicates significant distribution shifts (*KS* test).
See also Figure S6.



(legend on next page)

screens identified FAs related to cell-cycle progression (e.g., MAPK signaling and Influence of Ras and Rho proteins on G1 to S Transition) among suppressors of rotavirus infection, and the FA associated with AMPK/p53-induced G1 arrest (Hypoxia and p53 in the Cardiovascular system) among enhancers of rotavirus infection, in addition to G1/S checkpoint-related genes (e.g., *PPP6C*, *PLK3*; Table S2 and S4).

We propose a model in which rotavirus replication proceeds through different phases, utilizing different host factors at the various stages of infection, accompanied by progressive changes in cellular morphology and metabolism to enhance virus replication efficiency (Figures 7H and S7G). Following the onset of viral gene expression and insertion of viroporin NSP4 into the ER (Hyser et al., 2013), Ca^{2+}-induced AMPK activation initiates a rotavirus replication-permissive cellular state through multiple downstream effects that is not limited to the induction of autophagy, but includes a whole range of cellular changes, including rearrangement of organelles and consumption of lipid stores, a reduction in host gene transcription and protein translation, and interference with the cell cycle. This cellular environment enhances virus protein-mediated mechanisms that reinforce viral gene expression, including NSP3-mediated PABPC1 nuclear retention and displacement from eIF4G1, and NSP1-mediated disruption of P bodies (Bhowmick et al., 2015; Harb et al., 2008). Alterations to P bodies and nucleoli may enable the virus to hijack RNA processing components, such as DDX6 to late-stage viroplasms, further facilitating viral replication. The resulting high levels of viral protein, which typifies late-stage infection, create mature, protein-dense, membrane-associated viroplasms whose formation is facilitated by ER sheets, as demonstrated by *REEP2* perturbation. Thus, both virus protein-mediated modulation of the host and hijacking of the AMPK signaling axis to enhance the cellular environment for viral gene expression give rise to stage-specific cellular changes along the trajectory of rotavirus infection progression.

DISCUSSION

In this study, we generate a high-confidence resource for the identification of host factors involved in rotavirus infection. All RNAi screens performed were reproducible and consistent. Through a probabilistic method for the aggregation of multiple datasets, we obtained a gene score indicating likelihood of an on-target infection phenotype upon RNAi, within which known rotavirus host factors were retrieved, validating our approach. Using a rank-based analysis for FA enrichment, we provide an unbiased, systems-wide view of the cellular processes important for rotavirus replication, an approach that would enhance the analysis of perturbation screens from any field. Combined with our multivariate, single-cell infection progression trajectory, we uncover a complex yet ordered host cell reorganization program during rotavirus infection that provides a replication-permissive cellular environment for the virus.

Many aspects of this ordered reorganization appear to occur downstream of AMPK activation, a key signaling axis for the coordination of cellular growth and metabolism (Mihaylova and Shaw, 2011). While AMPK activation on rotavirus infection has previously been reported (Crawford et al., 2012), we show that multiple downstream consequences of its activation are required. Thus, generally, AMPK activation facilitates cell survival during nutrient deprivation, but in the context of rotavirus infection, it fuels viral replication. Consequently, MAPK signaling, which activates a number of processes abrogated by AMPK, suppresses infection progression, and the virus preferentially replicates in the G1 cell-cycle phase (mirroring that of infected mature enterocytes in vivo). How AMPK activation and consequent mTORC1 inhibition favors viral protein synthesis remains unclear; whether reduced eIF4G levels is sufficient requires further investigation.

Our approach to harness the variability in virus infection progression to order cells along an infection trajectory enhances our understanding of pathogen-host interaction. It abrogates

Figure 7. AMPK Enhances Rotavirus Infection by Inducing a Replication-Permissive Cellular Environment

(A) Schematic of the pleiotropic coordinator AMPK, activated by low ATP or high calcium. Points of action of pharmacological activators (blue) and inhibitors (green) are indicated.

(B) Levels of phosphorylated AMPK increase along infection progression. 209 cells infected for 8 hr, stained for pAMPK(Thr172) and NSP5, were used in trajectory construction. The weighted mean (line) and SD (shaded) of normalized mean pixel intensity per single cell is plotted. Early, mid, and late infection stages are indicated (dashed lines).

(C) Inhibition of AMPK reduces rotavirus infection and progression. Cells were treated with dorsomorphin or DMSO for 1 hr before addition of virus for a further 7 hr and IF for VP6. Mean ± SD of triplicate wells. * indicates $p < 0.05$ (t test). Representative images of the higher dose of dorsomorphin and control are shown.

(D) Activators of AMPK and inhibitors of cellular growth accelerate infection progression. Cells were treated with drugs for 1 hr before addition of virus for a further 7 hr. Infected cells were analyzed for shifts along the infection trajectory relative to the appropriate solvent-only control (water or DMSO), by spiking small numbers of perturbed cells into control cell populations, as for Figure 6.

(E) AMPK activation triggers metabolic changes that mimic those of mid-stage virus infection. Uninfected cells were serum-starved, treated with water (control) or AICAR for 5 hr, before IF for the cellular marker indicated. Maximum violin width is normalized across conditions. Extent of significant distribution shifts are indicated with the Kolmogorov-Smirnov (*KS*) statistic, and mean with red crosses. Representative images are shown.

(F) Infection progression is enhanced in G1 cells. The relative distribution of 5,000 cells along the VIX (colored), after 8 hr of virus infection, in the cell cycle phases G1, S, and G2.

(G) Rotavirus infection arrests cells in G1. The fraction of G1 (blue), S (yellow), G2 (red), and infected (black) cells in duplicate wells exposed to rotavirus at 0–12 hr p.i.

(H) A model for the dynamics of cellular changes along a rotavirus infection progression axis. Cellular reorganization during infection requires different host factors (down-hits in red; up-hits in blue) at different stages of progression, many downstream of early-stage AMPK activity (black arrows), that enhance virus protein-mediated effects (purple arrows). Host factors and cell reorganization events are positioned based on known functional information and our VIX (top) analysis. Weighted mean of trajectory features are indicated: NSP5 intensity (purple), NSP5 spots (blue), NSP5 texture sum variance (brown), and NSP5 texture sum average (green).

See also Figure S7.

the need for live cell sensors and outperforms laborious time course assays, allowing a systematic mapping of cellular reorganization dynamics from one fixed cell population. Furthermore, this can be combined with genetic or pharmacological perturbation studies, allowing the placement of host factors at specific positions along the infection progression axis corresponding to their function. While here we have used trajectories spanning most stages of rotavirus infection, the approach can be adapted to obtain trajectories on other pathogens or more focused infection stages, such as pathogen entry, replication, assembly, and spread, by choosing appropriate markers and readouts. Also, it will be feasible to obtain trajectories from high-throughput perturbations in parallel, allowing large-scale mapping of host factors along the infection progression axis. Such data would permit inference of causal effects and functional genetic interactions, providing systems-level insights into host-pathogen interaction that are currently unavailable.

EXPERIMENTAL PROCEDURES

Details of all experimental and computational procedures are described in the Supplemental Experimental Procedures, as well as all methods related to supplemental figures and tables. Complete gene and FA data can be browsed at http://rotavirus.infectome.org. Also at this site, the computational code developed for this work is available for download.

RNAi Screens

HCT 116 cells were reverse transfected with RNAi effectors and cultured for 72 hr before rotavirus infection for 8 hr. All cells were detected using a nuclear DAPI stain, and infected cells by immunofluorescence for viral capsid protein VP6, and imaged using a CellVoyager 7000 (Yokogawa) in epifluorescence mode, a 10x Olympus objective of 0.4 N.A, and a Neo sCMOS camera (Andor, 2,560 × 2,160 pixels). Following nuclei detection and single-cell feature extraction, infected cells were classified using support vector machines (SVMs). Infection indices (number infected cells/total cells) per well were corrected for plate, population context, and cell viability effects.

Probabilistic Data Aggregation

To test the significance of RNAi consistency, we derived two probability aggregation scores per gene, PAS_{down} and PAS_{up}, indicating the likelihood of the gene being required for, or inhibiting, infection, respectively. Each of 20,606 FAs were tested for significant enrichment in the PAS-ranked dataset for iteratively increasing numbers of ranked genes using a hypergeometric probability distribution function. The minimum p values from the resulting vectors for each FA were stored as the FA enrichment scores ($FAES_{down}$ and $FAES_{up}$).

Virus Infection Axis (VIX)

HCT 116 cells were infected with DS-1 for 3–15 hr, subject to immunofluorescence for viroplasm component NSP5 after fixation and imaged with the same microscope as before but utilizing the enhanced CSU-X1 spinning disk (Microlens-enhanced dual Nipkow disk confocal scanner, wide view type) and a 60x water immersion Olympus objective of 1.2 N.A. Infected cells were used as input in Cycler (Gut et al., 2015), which uses multivariate single cell features to perform *k*-nearest neighbor graph-based embedding of cells into a single dimension, a virus infection progression trajectory. Combining NSP5 immunofluorescence with readouts of various cellular pathways enabled infected cell activities to be quantified along the VIX. To assess the impact of gene perturbations on infection progression, infected cells from perturbed populations were spiked into controls over multiple bootstraps.

Selected Host Factor Perturbations

HCT 116 cells were reverse transfected using validation siRNA library sets 2 and 3 and, unless otherwise stated, grown 8 hr p.i. Fixed cells were assayed either by branched DNA single molecule RNA FISH (Affymetrix) to probe for viral segment 5+, or immunofluorescence for levels of the viral antigens VP6 and NSP5, with the exception of *REEP2* knockdown, which was assayed for NSP5 and the ER marker calreticulin.

Drug Treatments

HCT 116 cells were treated with the indicated drugs for 1 hr before infection with DS-1 and incubated a further 7 hr before fixation. In cells not infected with virus, drug treatments or serum-starvation was performed for 5 hr.

Cell-Cycle Phase Classification

Mitotic (M) phase cells were identified by training an SVM and excluded. S phase cells were identified based on DNA incorporation of EdU added to cells 20 min before fixation. Together with thresholds for integrated nuclear DAPI intensity and mean cytoplasmic Cyclin B1 intensity, cells could thus be discriminated between G1, S, and G2 phases.

SUPPLEMENTAL INFORMATION

Supplemental Information includes seven figures, six tables, and Supplemental Experimental Procedures and can be found with this article online at http://dx.doi.org/10.1016/j.chom.2016.06.005.

AUTHOR CONTRIBUTIONS

L.P. initiated the study. V.A.G. and L.P. designed the experiments and wrote the manuscript. V.A.G. performed all experiments and all image and data analysis.

ACKNOWLEDGMENTS

We thank K. Mench, R. Holtackers, E. Escher, M. Moser, and N. Battich for experimental support; F. Schmich (ETH Zurich), M. Tadmor (Columbia University), and G. Gut for help with data analysis; Y. Yakimovich for help with computational infrastructure and website development; O. Burrone (ICGEB Trieste), A. Metzler (Virology Institute, UZH), and C. Eichwald (Virology Institute, UZH) for reagents; and all members of the lab for useful comments on the manuscript. V.A.G. was supported by a postdoctoral fellowship from the University of Zurich. L.P. acknowledges financial support from SystemsX.ch, the Swiss National Science Foundation, the University of Zurich, and the University of Zurich Research Priority Program in Systems Biology and Functional Genomics.

Received: January 15, 2016
Revised: April 12, 2016
Accepted: May 24, 2016
Published: July 13, 2016

REFERENCES

Arnoldi, F., De Lorenzo, G., Mano, M., Schraner, E.M., Wild, P., Eichwald, C., and Burrone, O.R. (2014). Rotavirus increases levels of lipidated LC3 supporting accumulation of infectious progeny virus without inducing autophagosome formation. PLoS ONE *9*, e95197.

Bendall, S.C., Davis, K.L., Amir, A.D., Tadmor, M.D., Simonds, E.F., Chen, T.J., Shenfeld, D.K., Nolan, G.P., and Pe'er, D. (2014). Single-cell trajectory detection uncovers progression and regulatory coordination in human B cell development. Cell *157*, 714–725.

Bhowmick, R., Mukherjee, A., Patra, U., and Chawla-Sarkar, M. (2015). Rotavirus disrupts cytoplasmic P bodies during infection. Virus Res. *210*, 1–26.

Campagna, M., Budini, M., Arnoldi, F., Desselberger, U., Allende, J.E., and Burrone, O.R. (2007). Impaired hyperphosphorylation of rotavirus NSP5 in cells depleted of casein kinase 1alpha is associated with the formation of viroplasms with altered morphology and a moderate decrease in virus replication. J. Gen. Virol. *88*, 2800–2810.

Chemello, M.E., Aristimuño, O.C., Michelangeli, F., and Ruiz, M.C. (2002). Requirement for vacuolar H+ -ATPase activity and Ca2+ gradient during entry of rotavirus into MA104 cells. J. Virol. *76*, 13083–13087.

Cherry, S., Doukas, T., Armknecht, S., Whelan, S., Wang, H., Sarnow, P., and Perrimon, N. (2005). Genome-wide RNAi screen reveals a specific sensitivity of IRES-containing RNA viruses to host translation inhibition. Genes Dev. *19*, 445–452.

Crawford, S.E., Hyser, J.M., Utama, B., and Estes, M.K. (2012). Autophagy hijacked through viroporin-activated calcium/calmodulin-dependent kinase kinase-β signaling is required for rotavirus replication. Proc. Natl. Acad. Sci. USA *109*, E3405–E3413.

Eichwald, C., Vascotto, F., Fabbretti, E., and Burrone, O.R. (2002). Rotavirus NSP5: mapping phosphorylation sites and kinase activation and viroplasm localization domains. J. Virol. *76*, 3461–3470.

Eichwald, C., Jacob, G., Muszynski, B., Allende, J.E., and Burrone, O.R. (2004). Uncoupling substrate and activation functions of rotavirus NSP5: phosphorylation of Ser-67 by casein kinase 1 is essential for hyperphosphorylation. Proc. Natl. Acad. Sci. USA *101*, 16304–16309.

Gut, G., Tadmor, M.D., Pe'er, D., Pelkmans, L., and Liberali, P. (2015). Trajectories of cell-cycle progression from fixed cell populations. Nat. Methods *12*, 951–954.

Gutiérrez, M., Isa, P., Sánchez-San Martin, C., Pérez-Vargas, J., Espinosa, R., Arias, C.F., and López, S. (2010). Different rotavirus strains enter MA104 cells through different endocytic pathways: the role of clathrin-mediated endocytosis. J. Virol. *84*, 9161–9169.

Hao, L., He, Q., Wang, Z., Craven, M., Newton, M.A., and Ahlquist, P. (2013). Limited agreement of independent RNAi screens for virus-required host genes owes more to false-negative than false-positive factors. PLoS Comput. Biol. *9*, e1003235.

Harb, M., Becker, M.M., Vitour, D., Baron, C.H., Vende, P., Brown, S.C., Bolte, S., Arold, S.T., and Poncet, D. (2008). Nuclear localization of cytoplasmic poly(A)-binding protein upon rotavirus infection involves the interaction of NSP3 with eIF4G and RoXaN. J. Virol. *82*, 11283–11293.

Hyser, J.M., Utama, B., Crawford, S.E., Broughman, J.R., and Estes, M.K. (2013). Activation of the endoplasmic reticulum calcium sensor STIM1 and store-operated calcium entry by rotavirus requires NSP4 viroporin activity. J. Virol. *87*, 13579–13588.

Jones, R.G., Plas, D.R., Kubek, S., Buzzai, M., Mu, J., Xu, Y., Birnbaum, M.J., and Thompson, C.B. (2005). AMP-activated protein kinase induces a p53-dependent metabolic checkpoint. Mol. Cell *18*, 283–293.

Kruiswijk, F., Yuniati, L., Magliozzi, R., Low, T.Y., Lim, R., Bolder, R., Mohammed, S., Proud, C.G., Heck, A.J.R., Pagano, M., and Guardavaccaro, D. (2012). Coupled activation and degradation of eEF2K regulates protein synthesis in response to genotoxic stress. Sci. Signal. *5*, ra40.

López, S., and Arias, C.F. (2012). Rotavirus-host cell interactions: an arms race. Curr. Opin. Virol. *2*, 389–398.

Martínez, M.A., López, S., Arias, C.F., and Isa, P. (2013). Gangliosides have a functional role during rotavirus cell entry. J. Virol. *87*, 1115–1122.

Mihaylova, M.M., and Shaw, R.J. (2011). The AMPK signalling pathway coordinates cell growth, autophagy and metabolism. Nat. Cell Biol. *13*, 1016–1023.

Mitrea, D.M., and Kriwacki, R.W. (2016). Phase separation in biology; functional organization of a higher order. Cell Commun. Signal. *14*, 1.

Orvedahl, A., Sumpter, R., Jr., Xiao, G., Ng, A., Zou, Z., Tang, Y., Narimatsu, M., Gilpin, C., Sun, Q., Roth, M., et al. (2011). Image-based genome-wide siRNA screen identifies selective autophagy factors. Nature *480*, 113–117.

Paytubi, S., Wang, X., Lam, Y.W., Izquierdo, L., Hunter, M.J., Jan, E., Hundal, H.S., and Proud, C.G. (2009). ABC50 promotes translation initiation in mammalian cells. J. Biol. Chem. *284*, 24061–24073.

Pelkmans, L., Fava, E., Grabner, H., Hannus, M., Habermann, B., Krausz, E., and Zerial, M. (2005). Genome-wide analysis of human kinases in clathrin- and caveolae/raft-mediated endocytosis. Nature *436*, 78–86.

Rämö, P., Drewek, A., Arrieumerlou, C., Beerenwinkel, N., Ben-Tekaya, H., Cardel, B., Casanova, A., Conde-Alvarez, R., Cossart, P., Csúcs, G., et al. (2014). Simultaneous analysis of large-scale RNAi screens for pathogen entry. BMC Genomics *15*, 1162.

Rubio, R.M., Mora, S.I., Romero, P., Arias, C.F., and López, S. (2013). Rotavirus prevents the expression of host responses by blocking the nucleocytoplasmic transport of polyadenylated mRNAs. J. Virol. *87*, 6336–6345.

Ryan, C.J., Cimermančič, P., Szpiech, Z.A., Sali, A., Hernandez, R.D., and Krogan, N.J. (2013). High-resolution network biology: connecting sequence with function. Nat. Rev. Genet. *14*, 865–879.

Scherl, A., Couté, Y., Déon, C., Callé, A., Kindbeiter, K., Sanchez, J.-C., Greco, A., Hochstrasser, D., and Diaz, J.-J. (2002). Functional proteomic analysis of human nucleolus. Mol. Biol. Cell *13*, 4100–4109.

Silva-Ayala, D., López, T., Gutiérrez, M., Perrimon, N., López, S., and Arias, C.F. (2013). Genome-wide RNAi screen reveals a role for the ESCRT complex in rotavirus cell entry. Proc. Natl. Acad. Sci. USA *110*, 10270–10275.

Cell Host & Microbe, Volume *20*

Supplemental Information

A Systems Survey of Progressive Host-Cell Reorganization during Rotavirus Infection

Victoria A. Green and Lucas Pelkmans

Supplemental Figures

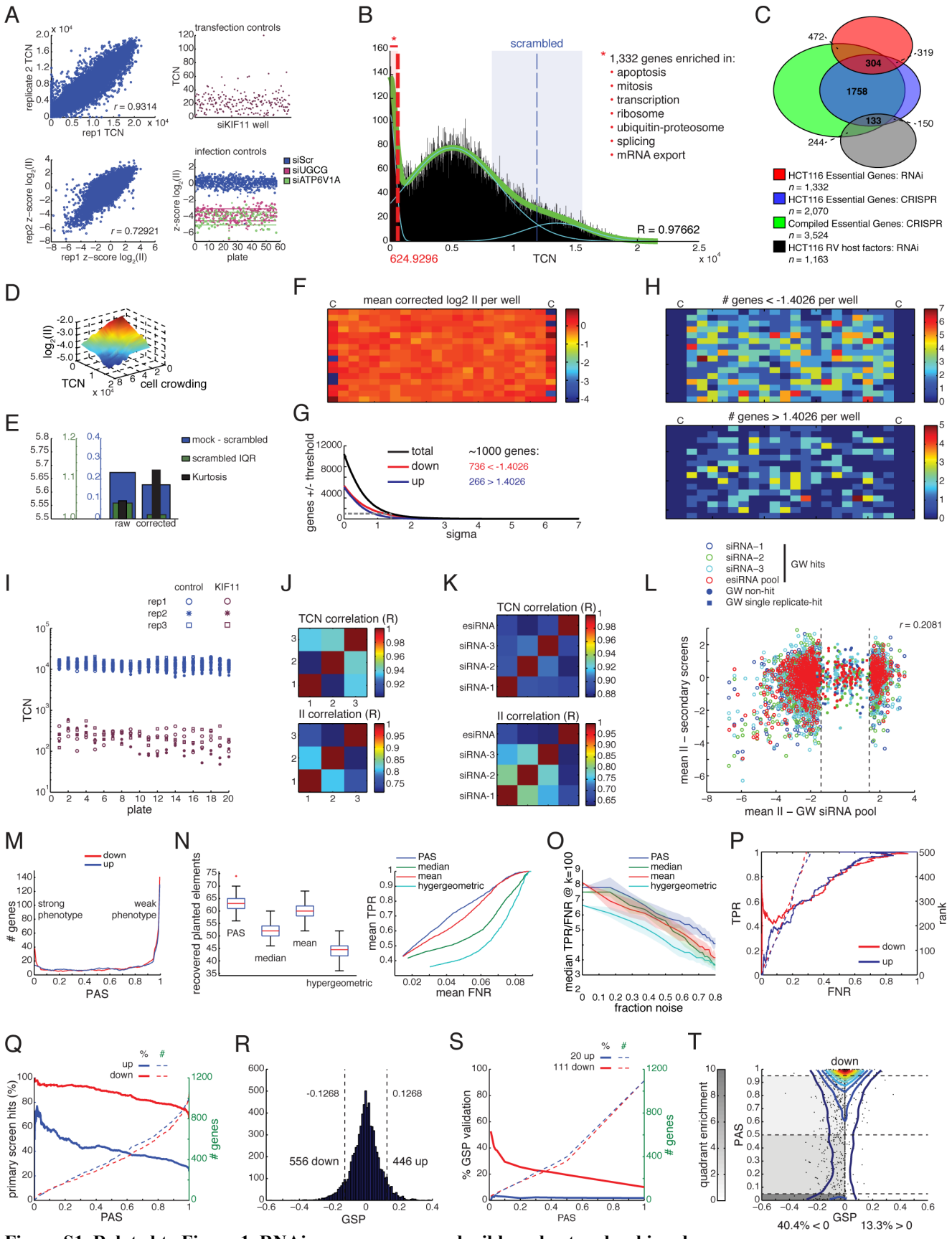


Figure S1. Related to Figure 1. RNAi screens are reproducible, robust and unbiased

(A) Reproducibility of GW screen. Pearson's correlation coefficient (*r*) between replicates for total cell numbers (TCN) and corrected log2 infection index (II) for each replicate (left panels). Results for every well across all 120 plates transfected with cell killer control siKIF11 (upper-right), and infection negative control siScrambled (blue) and positive controls siUGCG (pink) and siATP6V1A (green; lower-right).

(B) Total cell number (TCN) per well for GW screen. A three-Gaussian (cyan) mixture model (green; goodness of fit (R)) were fitted to the data to determine the cell killer TCN threshold (red) of 625 cells. The standard deviation of TCN values

for siScrambled control wells are shaded around the mean (dashed blue). Enriched functional annotations in the 1,332 genes whose TCN was <625 on knockdown, as determined from the DAVID resource.
(C) Overlap of cell killers with published lists of genes essential for cell viability. The Venn diagram indicates the degree of overlap between the 1,332 genes whose TCN was <625 on knockdown (red), with the 2,070 genes identified as essential for cell viability in a genome-wide CRISPR/Cas9 knockout screen in HCT 116 cells (blue), and the 3,524 genes identified from several CRISPR/Cas9 screens in multiple cell lines (green). For comparison, overlap with the 1,163 genes (black) found to most strongly reduce infection (unique from top-ranked 2,000 genes in each replicate) is also shown.
(D) Influence of cell crowding and TCN per well on rotavirus infection indices in unperturbed cells.
(E) Impact of cellular context correction on the GW dataset. The difference between the mean of all mock and the mean of all scrambled wells, along with the interquartile range (IQR) of all scrambled wells, and the kurtosis of the entire GW dataset, are given before (raw) and after correction.
(F) Mean II per well over all GW plates show no well bias, except for control wells.
(G) Threshold determination for selection of the 1,000 genes with the strongest infection phenotype for validation screening.
(H) The frequency of those 1,000 GW hits is evenly distributed across plates.
(I) Robustness of secondary screens. Results for every well across all 60 plates (20 x 3 replicates) transfected with cell killer control siKIF11 (left).
(J) Reproducibility of secondary screens. Pearson's correlation coefficient (*R*) between replicates for total cell numbers and corrected log2 II (right panels).
(K) Consistency of secondary screens. Pearson's correlation coefficient (*R*) between secondary screens for total cell numbers (left) and corrected log2 II (right).
(L) Consistency between primary and secondary screens. A comparison of corrected log2 infection indices (II) for the 1,100 validation genes, averaged over biological replicates, between the primary genome-wide (GW) screen (in duplicate) and four secondary screens (in triplicate): 3 x siRNA (blue, green and cyan) and 1 x esiRNA (red). GW hits (±1.4σ) are indicated with open circles, genes beyond this threshold in a single replicate only with filled squares, and non-hits with filled circles. Pearson's correlation coefficient (*r*).
(M) Number of validation genes with a given PAS_{down} and PAS_{up} value.
(N) PAS outperforms other data aggregation methods in recovering hits from simulated screen data. Percentage of recovered planted elements, i.e. simulated positive controls (data + exponential noise), from simulated screen data (actual data + normal noise) was compared over 100 bootstraps (left) between four data aggregation methods: PAS, median, mean and a method based on hypergeometric probabilities. True positive rate (TPR) versus false negative rate (FNR) were averaged over the 100 bootstraps (right).
(O) PAS outperforms other data aggregation methods in recovering hits from noisy data. Increasing numbers of randomly shuffled screen datasets were added to the original five to increase the fraction of added noise. The mean and standard deviation TPR/FNR ratios were calculated for the top 100 genes (k=100) over 100 bootstraps.
(P) TPR versus FNR ratios for the aggregated screening dataset was calculated iteratively over increasing ranks.
(Q) Aggregated RNAi data validates the genome-wide screen. The percentage of genes below a given probability aggregation score (PAS, dashed lines) that were GW hits (±1.4σ, solid lines), for both PAS_{down} (red) and PAS_{up} (blue).
(R) Histogram of gene specific phenotype (GSP) values predicted for rotavirus infection genome-wide. Only genes with values that deviate from zero (and thus predicted to have a phenotype) are included. Thresholds discriminating the strongest 1,000 predicted phenotypes (±0.1268σ; dashed lines).
(S) Aggregated RNAi data validates few GSP predictions. The precentage of genes below a given probability aggregation score (PAS, dashed lines) that were predicted to have a strong GSP (±0.1268σ, solid lines), for both PAS_{down} (red) and PAS_{up} (blue).
(T) Enrichment for GSPs in validated down-hits. PAS_{down} for validation genes and predicted GSP values, of which 40.4% were <0, and 13.3% >0. The enrichment for genes with a GSP<0 increases for smaller PAS values, or a more stringent definition of hit validation.

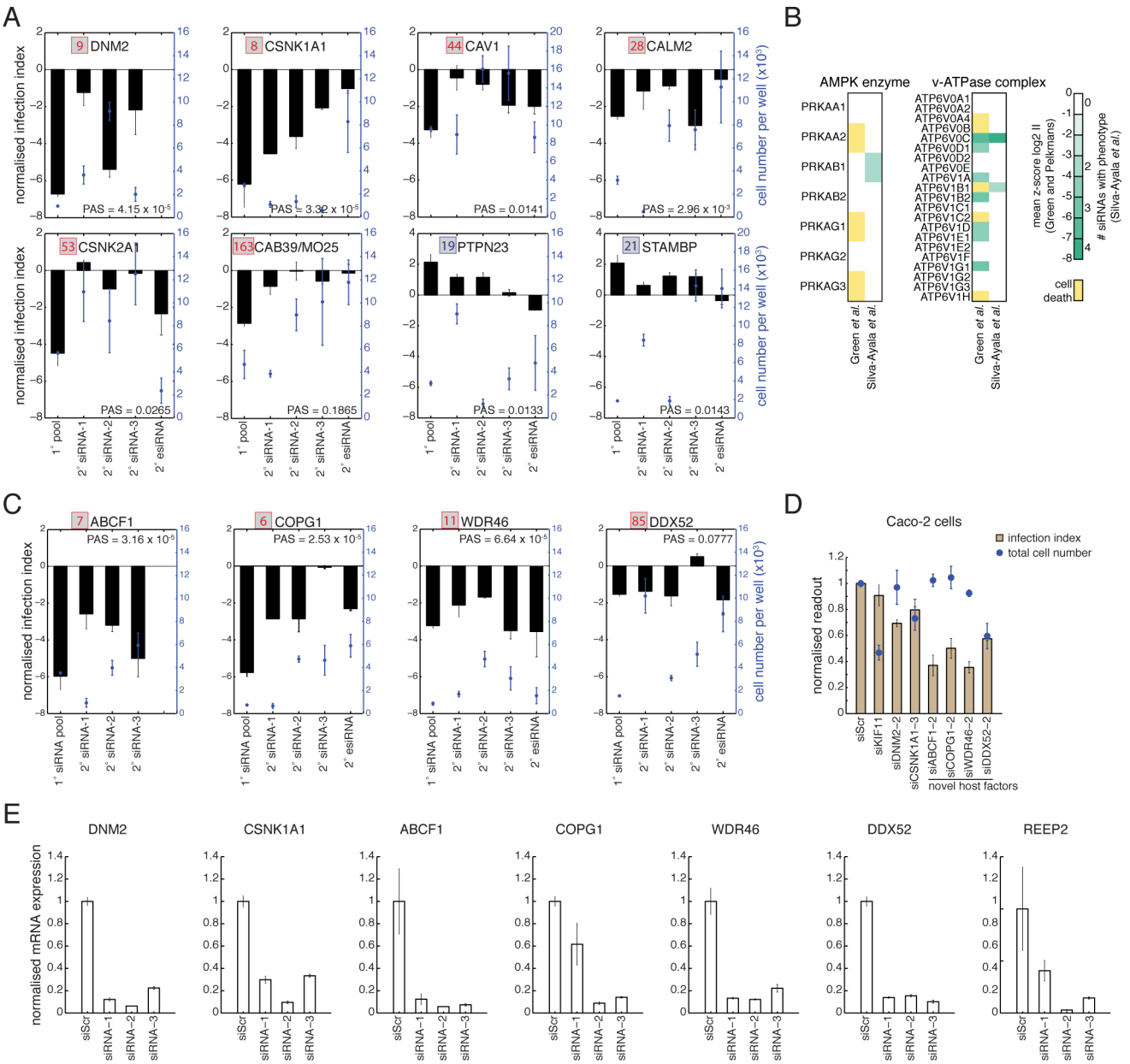


Figure S2. Related to Figure 2. RNAi screens reproduce known, and reveal novel, rotavirus host factors

(A) Summary of TCN and II data from all RNAi screens for known rotavirus host factors. Gene ranks for down- (red) and up-hits (blue) are boxed in grey.

(B) Summary of GW data for each isoform of AMPK enzyme and v-ATPase complex. Mean (of duplicates) corrected II are shaded green, with yellow indicating perturbations that resulted in less than 625 cells per well in both replicates.

(C) Summary of TCN and II data from all RNAi screens for novel rotavirus host factors. As for (A).

(D) Novel rotavirus host factors also inhibit infection in Caco-2 cells. Perturbation, infection and infection index quantification was performed in Caco-2 cells as for HCT116 cells in the RNAi screens. Infection index (II, bars) and total cell number (TCN, dots) are normalised to siScrambled control. Mean ± SD of triplicate wells.

(E) RNAi mediates on-target gene expression knockdown. Host factor mRNA expression, isolated from siRNA-treated HCT 116 cells in triplicate, was quantified by qRT-PCR. Average target-specific mRNA levels were determined relative to three housekeeping genes and normalised to siScrambled control. Mean ± SD.

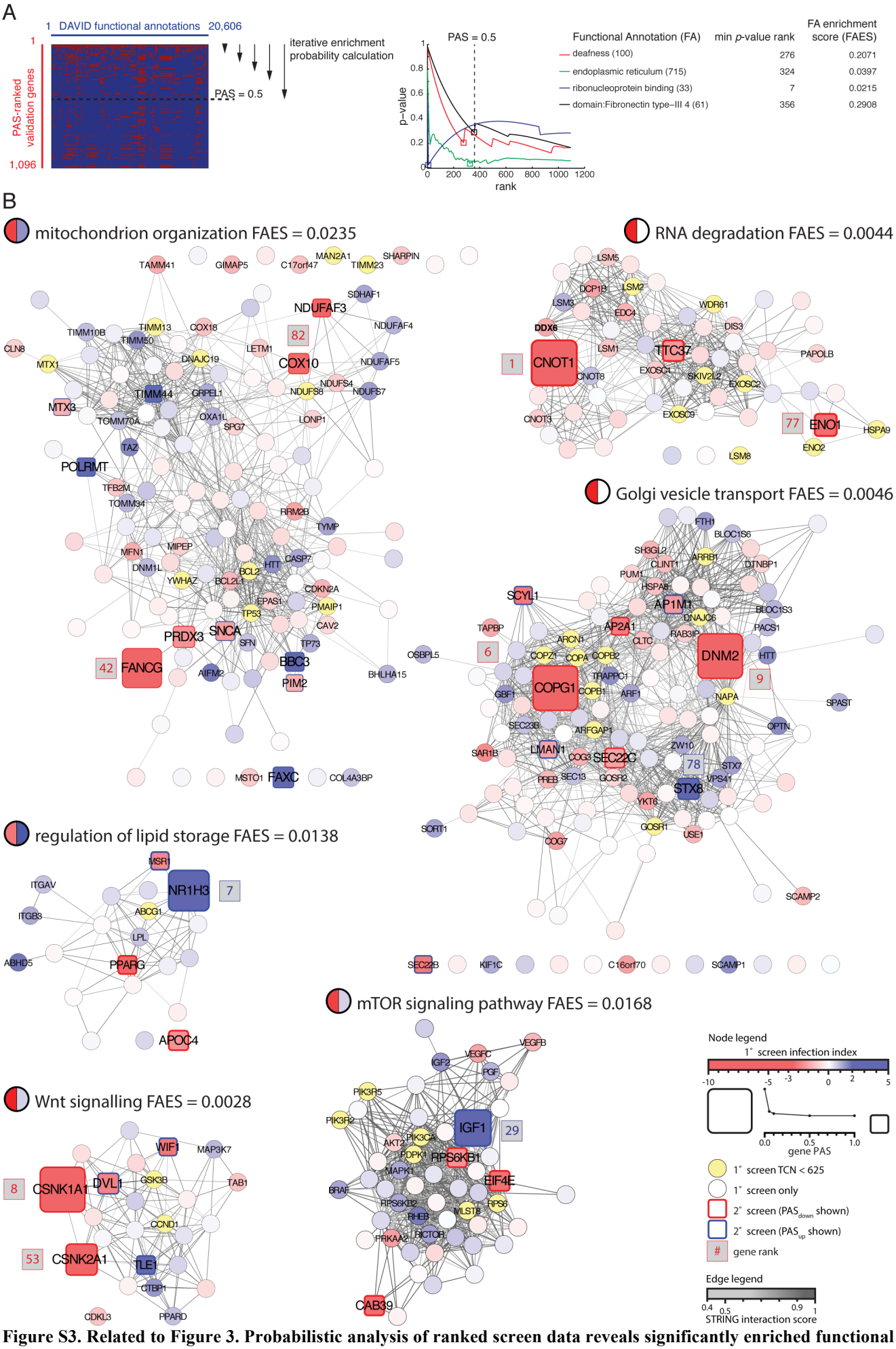


Figure S3. Related to Figure 3. Probabilistic analysis of ranked screen data reveals significantly enriched functional annotations

(A) Principle of the functional annotation enrichment score (FAES). The probability of significant enrichment was determined from a hypergeometric probability distribution function on PAS-ranked genes for iteratively increasing ranks (left). The minimum *p*-value from ranks 1 to *n*, where PAS <0.5, was taken as the FAES (right).
(B) Significantly enriched FAs contain genes that confer an infection phenotype. All genes annotated to the selected enriched FAs are represented as nodes, with protein-protein interaction confidence scores >0.4, from STRING, defining edges. Nodes are coloured according to the mean infection index from the primary, genome-wide screen, with perturbations reducing infection in red, and those increasing infection in blue. Yellow nodes resulted in a cell number per well <625 in both replicates. White nodes were not present in the screen. The PAS of genes also in secondary screens (square), with node size indicating PAS_{down} (red) or PAS_{up} (blue). Genes with a PAS-based ranking <100 are indicated with grey boxes.

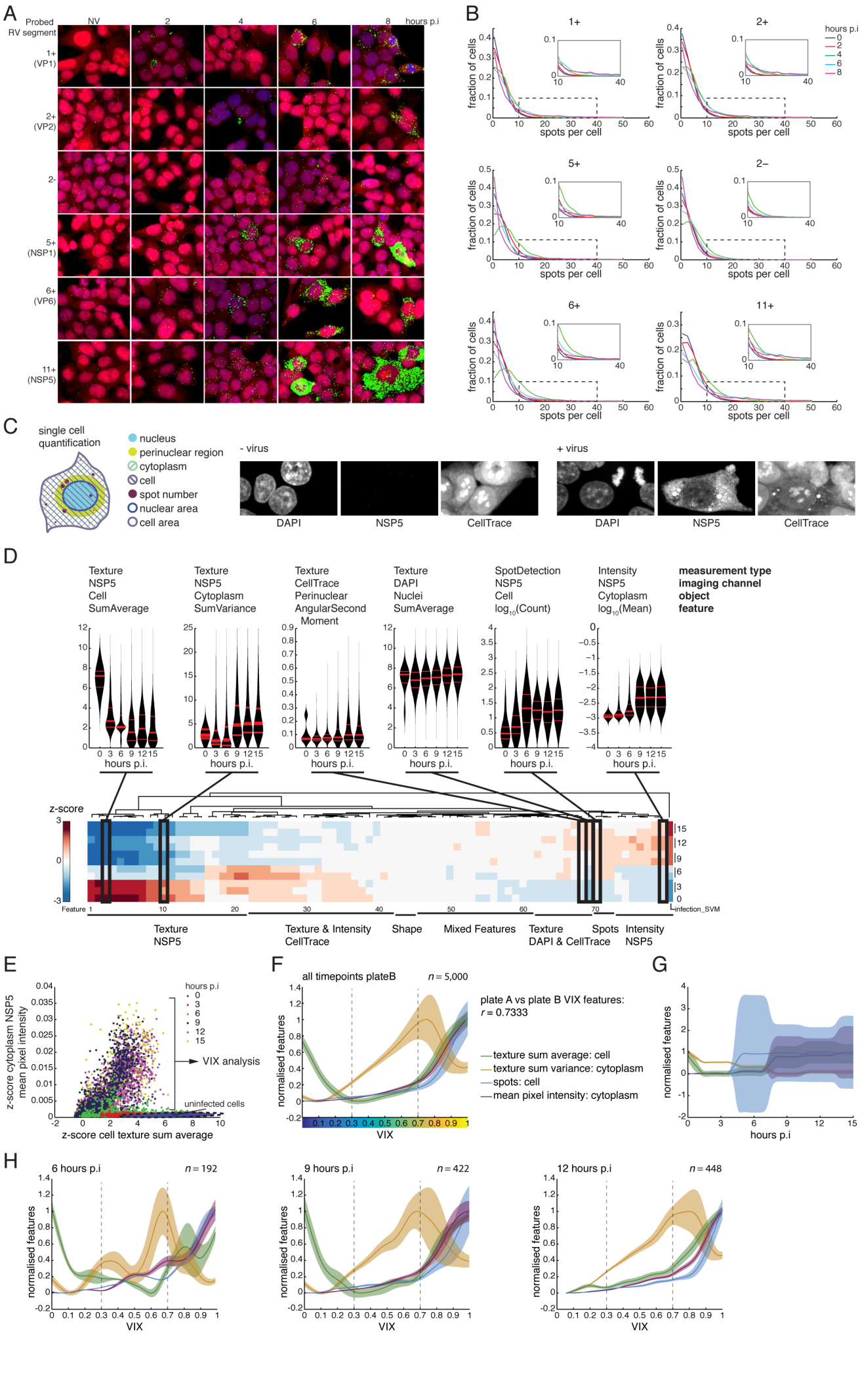

A
Probed RV segment
NV
2
4
6
8
hours p.i
1+ (VP1)
2+ (VP2)
2-
5+ (NSP1)
6+ (VP6)
11+ (NSP5)
B
1+
2+
5+
2–
6+
11+
hours p.i
0
2
4
6
8
fraction of cells
spots per cell
C
single cell quantification
nucleus
perinuclear region
cytoplasm
cell
spot number
nuclear area
cell area
- virus
+ virus
DAPI
NSP5
CellTrace
D
Texture NSP5 Cell SumAverage
Texture NSP5 Cytoplasm SumVariance
Texture CellTrace Perinuclear AngularSecond Moment
Texture DAPI Nuclei SumAverage
SpotDetection NSP5 Cell log10(Count)
Intensity NSP5 Cytoplasm log10(Mean)
measurement type
imaging channel
object
feature
hours p.i.
z-score
Feature
infection_SVM
Texture NSP5
Texture & Intensity CellTrace
Shape
Mixed Features
Texture DAPI & CellTrace
Spots
Intensity NSP5
E
z-score cytoplasm NSP5 mean pixel intensity
hours p.i
VIX analysis
uninfected cells
z-score cell texture sum average
F
all timepoints plateB
n = 5,000
normalised features
VIX
plate A vs plate B VIX features:
r = 0.7333
texture sum average: cell
texture sum variance: cytoplasm
spots: cell
mean pixel intensity: cytoplasm
G
normalised features
hours p.i
H
6 hours p.i
n = 192
9 hours p.i
n = 422
12 hours p.i
n = 448
normalised features
VIX

Figure S4. Related to Figure 4. Multivariate features of infection define progression along a virus infection axis

(A) Viral gene expression is heterogeneous. HCT 116 cells infected with rotavirus for 2-8 hours were probed for viral RNA segments in quadruplicate by branched-DNA single-molecule RNA FISH. Targeted rotavirus segments are listed with the respective viral protein encoded, where appropriate, given in brackets underneath. Representative images are shown.

(B) Quantification of viral RNA segments. At least 2,000 cells, infected at an MOI of 0.2, were anlaysed per probe per replicate for image-based spot detection.

(C) Schematic of single cell features quantified (left) along with representative images of the impact of rotavirus infection on nuclear DAPI stain, NSP5 expression and unspecific protein stain, CellTrace.

(D) Semi-supervised feature selection for rotavirus infection trajectory construction. Z-scored single cell features were clustered according to their behaviour over an infection time course of 0-15 hours (lower). The violins indicate the single cell distributions of six features that varied over the time course, with violin width normalised across time points, and mean and standard deviations shown (red lines).

(E) Gating strategy for selecting infected cell-trajectory input. Single cell data from wells infected with rotavirus were combined with those from uninfected control wells. Z-scored NSP5 intensity and texture features were used to identify where cells from the uninfected well were enriched and exclude cells within this region from the trajectory input.

(F) Trajectory features behave reproducibly between experiments. As for Figure 4C, 5,000 infected cells from multiple time-points were used in trajectory construction. The mean Pearson's correlation coefficient (*r*) between feature behaviour along the VIX from this experiment and the one depicted in Figure 4C, which were conducted several weeks apart.

(G) Trajectory features are highly variable over time. The weighted mean and standard deviations of the four features used in trajectory construction.

(H) Trajectories can be constructed from single time points. A comparison of the trajectory feature behaviour when constructing trajectories from cells infected for 6, 9 or 12 hours.

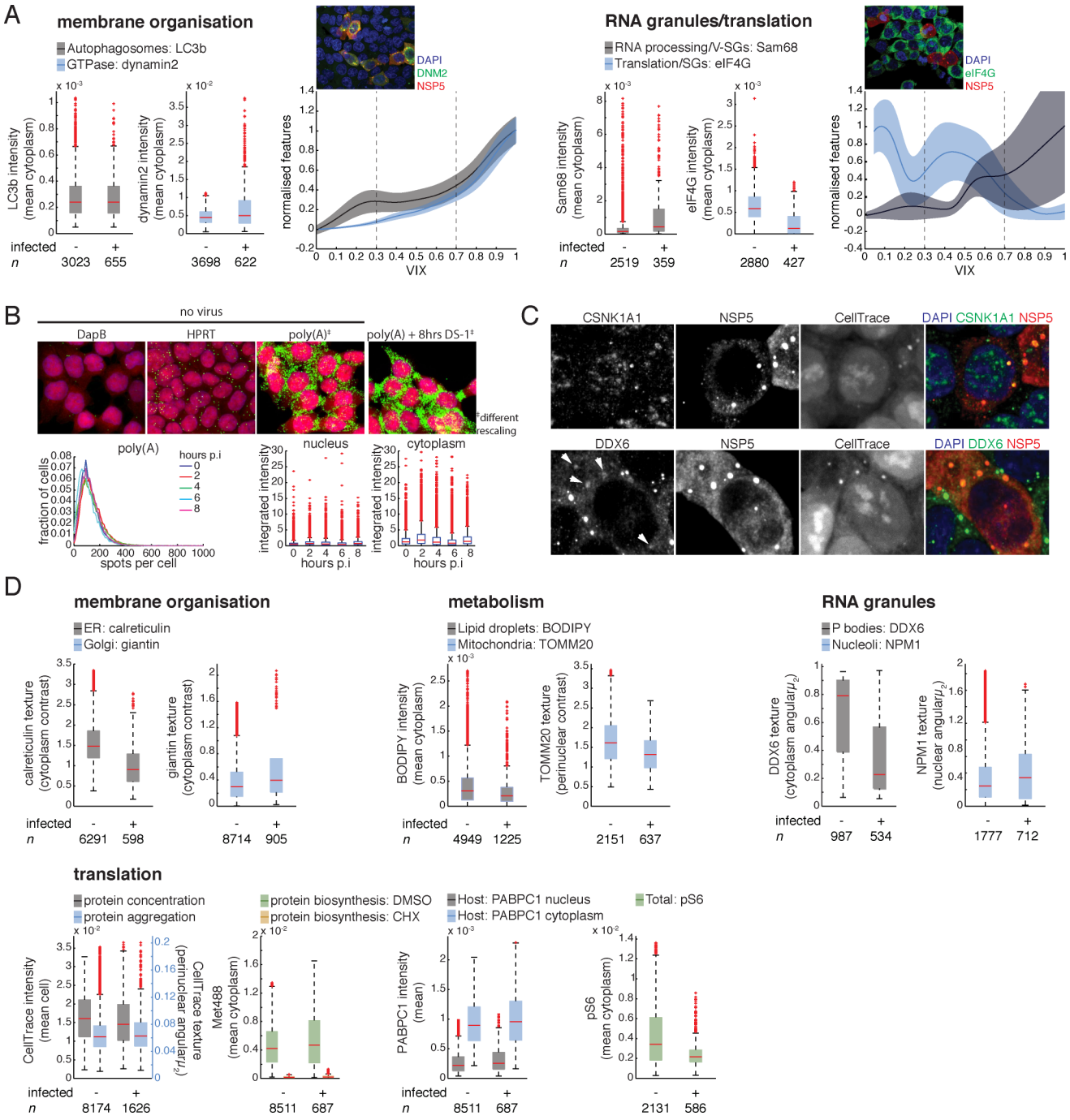


Figure S5. Related to Figure 5. Cellular changes accompanying rotavirus infection

(A) Dynamics of cellular reorganisation along the VIX. HCT 116 cells were infected for 8 hours before fixation and immunostaining for cellular markers and NSP5. Box plots of single cell population quartiles for uninfected and infected cells (left). Cells from duplicate wells were pooled and classified as either infected or uninfected, based on NSP5 features (see also Figure S4E). These same infected cells were used in trajectory construction with the weighted mean (lines) and standard deviation (shaded) of normalised cellular features overlaid, as indicated. Early, mid and late stages of infection are discriminated (dashed lines). Representative images of DAPI, NSP5 and either immunofluorescence for DNM2 (left) or eIF4G (right), are shown.

(B) Cellular mRNA over an infection time course. HCT 116 cells infected for 2-8 hours were probed by branched DNA single-molecule RNA FISH for levels of cellular poly-adenylated (poly(A)) mRNA. Representative images of FISH (green) and CellTrace (red) for negative control bacterial transcript *DapB*, positive control *HPRT*, and poly(A) before and after 8 hours of infection. Frequency of mRNA spots per cell (lower left) and the integrated intensity in the FISH channel over all nuclear or cytoplasmic pixels (lower right).

(C) Representative images of DDX6 recruitment to late-stage, protein-dense viroplasms. CSNK1A1 is included as a positive control for colocalisation with all viroplasms.

(D) Pooling infected cells masks the dynamics of cellular reorganisation events. Single cell population quartiles of cellular features for uninfected and infected cells. Cells from duplicate wells were pooled and classified as either infected or uninfected, based on NSP5 features (see also Figure S4E). These same infected cells were used in the trajectories displayed in Figure 5A.

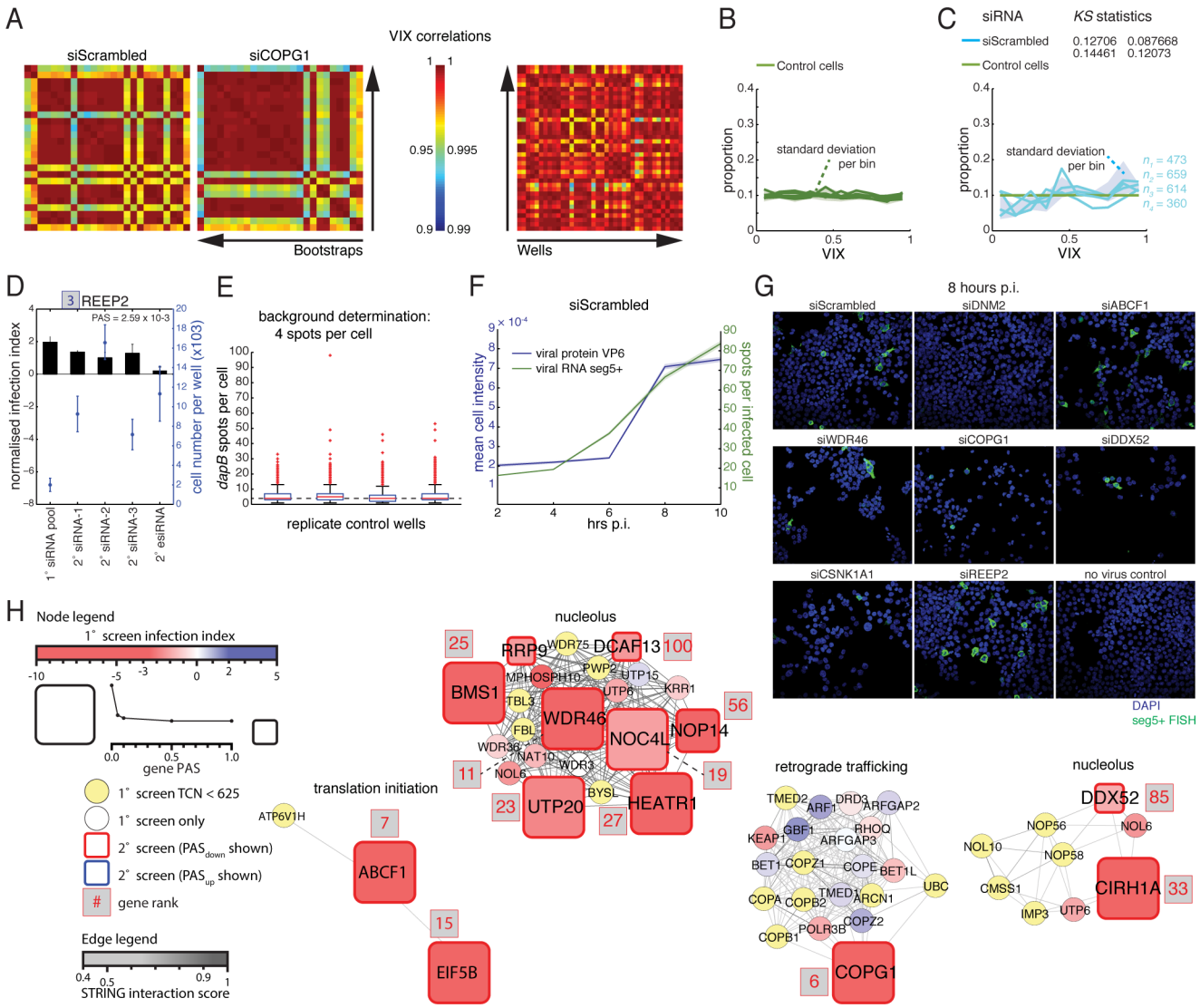


Figure S6. Related to Figure 6. Integrating protein interactions, screen data and the mapping of host factor perturbations to the VIX, suggests gene function in rotavirus infection

(A) Control cell positioning along the infection trajectory is not affected by spiking in cells from perturbed populations. Pearson's correlations across 25 bootstraps, for two different example perturbations (left), and for mean trajectory values (from bootstraps) across wells transfected with different siRNAs (right).

(B) Estimation of mapping error. Cells from four mock-transfected wells (green) not included in the control population were treated as perturbations and mapped back to the trajectory. Standard deviation per bin is shaded.

(C) Scrambled siRNA alters infection progression in a small, reproducible manner. The standard deviation per bin (shaded) of four replicate wells (cyan) transfected with non-targeting, siScrambled, normalised to mock-transfected control (green).

(D) Summary of TCN and II data from all RNAi screens for novel rotavirus host factor REEP2. Gene rank for up-hits (blue) is boxed in grey.

(E) Determination of the threshold of RNA molecule detection. The mean number of spots per cell on performing branched DNA single-molecule RNA (bDNA smRNA) FISH against negative control bacterial transcript *DapB* was calculated as the background level of spot detection (dashed line). The variability within four duplicate wells are shown.

(F) Dynamics of viral RNA and protein production. Cells transfected with siScrambled control were infected with rotavirus for the times shown. In one experiment, fixed cells were subject to immunofluorescence against viral protein VP6, and in another to (bDNA smRNA) FISH against viral RNA segment 5+. For VP6, mean (blue line) ± SEM (shaded) single-cell intensities are plotted. For seg5+, mean (green line) ± SEM (shaded) spots per infected cell (>4 spots per cell, see also (D)) are shown.

(G) Host factor perturbation differentially alters late viral RNA amplification. Representative images of bDNA smRNA FISH to segment 5+ at 8 hours p.i. on host factor knockdown.

(E) Interactors of novel host factors are enriched for infection phenotypes on perturbation. Genes listed as interacting with the featured host factors *ABCF1*, *COPG1*, *WDR46* and *DDX52*, based on experimental evidence and an interaction score of >0.4 in STRING, were used in networks. Nodes are coloured according to the mean infection index from the primary, genome-wide screen, with perturbations reducing infection in red, and those increasing infection in blue. Yellow nodes resulted in a cell number per well <625 in both replicates. White nodes were not present in the screen. The PAS of genes also in secondary screens (square), with node size indicating PAS_{down} (red) or PAS_{up} (blue). Genes with a PAS-based ranking <100 are indicated with grey boxes.

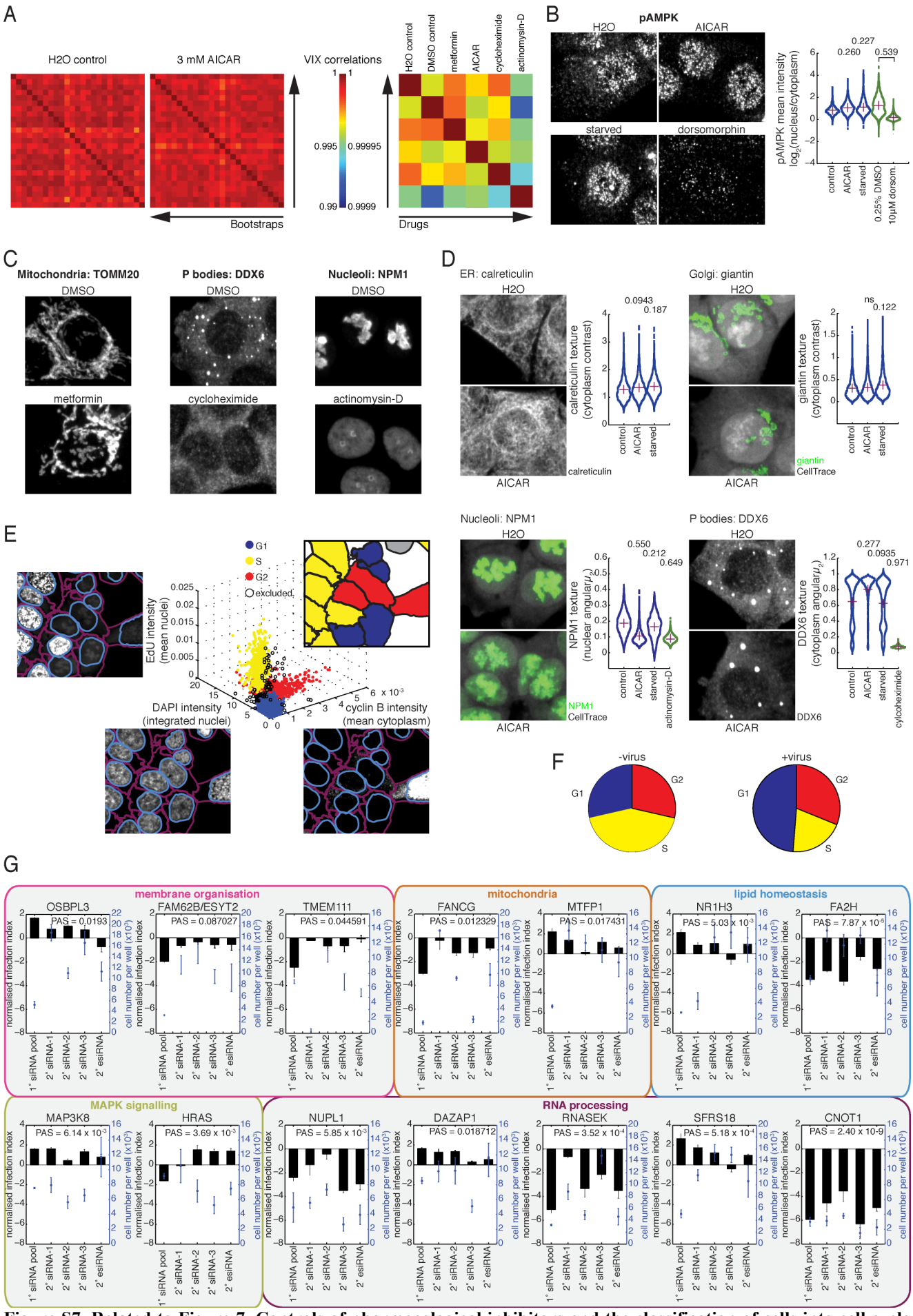


Figure S7. Related to Figure 7. Controls of pharmacological inhibitors and the classification of cells into cell cycle phases

(A) Control cell positioning along the infection trajectory is not affected by spiking in cells from perturbed populations. Pearson's correlations across 25 bootstraps, for two different example perturbations (left), and for mean trajectory values (from bootstraps) across wells treated with different drugs (right).
(B) Representative images of pAMPK immunofluorescence (IF) on treatment with water control, 3 mM AICAR, serum-starved or 10 μM dorsomorphin, for 5 hours (left). Single cell quantification of duplicate wells, with significant shifts indicated by the Kolmogorov-Smirnov (*KS*) statistic (right) and mean values with a red cross.
(C) Immunofluorescence for cellular markers on drug treatment confirms pharmacological action. Representative images of cells treated with DMSO control or 50 mM metformin, 50 μg/ml cyclohexamide and 0.67 μg/ml actinomysin-D for 5 hours were stained by IF for the markers indicated. Metformin alters mitochondria morphology, cycloheximide disperses P bodies, and actinomysin D disperses nucleoli, consistent with their purported action on inhibiting mitochondrial ATP production, translation and transcription, respectively.
(D) AMPK activation does not recapitulate all cellular reorganisation events observed on virus infection. Uninfected cells were serum-starved or treated with water (control), 3 mM AICAR, 50 μg/ml cyclohexamide or 0.67 μg/ml actinomysin-D, as indicated, for 5 hours in duplicate wells. Cells were then fixed and stained for the cellular marker indicated. Where significant shifts in cellular features were observed ($p = 0$), the extent of the shift is indicated with the *KS* statistic above the respective violin, with mean values indicated by red crosses. Representative images are shown in grayscale, except for giantin and NPM1, which are shown in green superimposed on CellTrace (gray).
(E) Image-based, cell cycle phase classification. Integrated DAPI intensity in the nucleus, incorporation of EdU and cyclin B cytoplasmic concentration were used to discriminate G1 (blue), S (yellow) and G2 (red) cell cycle phases. Poorly segmented and mitotic (M) phase cells were excluded (gray in pseudo-colored image; black circles in plot).
(F) Infection increases the proportion of cells in the G1 cell cycle phase. Cells from multiple wells infected with rotavirus for 4-12 hours were gated for infection, as for trajectory input selection (see Figure S4E). The proportion of cells in the G1 (blue), S (yellow) and G2 (red) cell cycle phases were then calculated in the uninfected and infected cells from the same wells.
(G) Summary of total cell number (TCN) and infection index (II) data from all RNAi screens for novel rotavirus host factors, grouped according to the functions these genes are associated with.

Supplemental Tables

Table S1. Related to Figure 1. Genome-wide screen data

Table S2. Related to Figure 2. Ranked aggregated data from all screens for the validation subset of genes

Table S3. Related to Figure 2. Greedy functional annotation assignment to validation genes

Table S4. Related to Figure 3. Rank-based functional annotation enrichment analysis

Table S5. Related to Figure 3. Threshold-based functional annotation enrichment analysis

Table S6. Related to Figure 4. Single cell features included in cluster analysis for trajectory feature selection

Supplemental Experimental Procedures

Cell culture

Cells of human colon carcinoma cell line HCT 116 (ATCC) were cultured in McCoy's 5A Medium Modified (Sigma-Aldrich) supplemented with 10% fetal calf serum (FCS) (complete medium). These cells were used in all experiments, except for when stated otherwise, at passage >2 and <7. We selected HCT 116 cells for this study following the testing of multiple human cell lines for high-throughput screening suitability, namely a) good image-based cell segmentation, b) high infection rates, c) high transfection rates, and d) high gene knockdown rates. For example, HeLa cells exhibited low infection efficiency, whereas human cell lines typically used in rotavirus infection, such as Caco-2 and HT-29, were not transfected efficiently. FHs 74 Int, a non-transformed intestinal cell line, were infected and transfected well, but this did not translate into a high gene knockdown efficiency due to their slow growth. HCT 116 cells, performed the best over all criteria.

For validation of selected novel host factors, we also used Caco-2 cells, which were cultured in Dulbecco's Modified Eagle Medium (Gibco) supplemented with 10% fetal calf serum (FCS).

RNA interference

In all screens, 2,000 HCT 116 cells (ATCC) were seeded in 60 µl of seeding medium, composed of McCoy's 5A Medium Modified supplemented with 13.3% FCS and 1% pen/strep (Gibco), in each well of a 384-well plate containing 20 µl of reverse transfection reagents. The latter comprised RNAi effectors targeting a single gene, 0.05 µl Lipofectamine 2000 (Life Technologies), and OptiMEM (Gibco).

For the primary, genome-wide screen, the RNAi effectors were the Dharmacon ON-TARGETplus® SMART pool® siRNA Library G-105005-02, comprising 18,026 pools of four siRNAs per gene, at a final concentration of 20 nM. Cell plates containing 1.6 pmol of siRNA pool were kindly provided by the LMC, RISC facility, ETH Zurich. The primary screen was conducted in duplicate, with biological replicates of 60 plates each assayed two weeks apart. Three secondary screens employed individual Ambion Silencer® Select siRNAs per well at a final concentration of 5 nM, targeting 1,100 genes, and the fourth secondary screen used 867 MISSION® esiRNA pools (Sigma-Aldrich) at a final concentration of 20 nM per well. All secondary screens were performed in triplicate. Cell plates for secondary screens were prepared from master plates through a series of dilutions with water using a BioMek® FXP liquid handling workstation (Beckman Coulter).

Following addition of cells to the reverse transfection mixes, plates were shaken at 700 rpm for 2 sec and stood at room temperature (RT) for 1 hour, before 72 hours of cultivation at 37°C, 5% CO_2. All screening plates contained multiple negative controls: for siRNA-based screens these were non-targeting sequences (scrambled) and, for the esiRNA screen, a pool targeting renilla luciferase (RLUC). Each plate also included positive transfection control wells with effectors targeting KIF11 (Entrez ID 3832), whose knockdown would be expected to kill cells, and, for the primary screen, infection control wells targeting both v-ATPase subunit ATP6V1A (Entrez ID 523) and an enzyme catalysing the first glycosylation step of glycosphingolipid synthesis, UGCG (Entrez ID 7357).

For RNAi of Caco-2 cells, to each well of a 384-well plate, 20 µl of reverse transfection reagents, comprising siRNA, 0.2 µl TurboFECT (Thermo Fisher Scientific) and OptiMEM, was added along with 1,750 cells in 60 µl media, giving a final siRNA concentration of 10 nM. Cells were incubated at 37°C, 5% CO_2 for 72 hours prior to assay.

Rotavirus infection assay

Human rotavirus A strain DS-1 (G2-P1B[4]-I2-R2-C2-M2-A2-N2-T2-E2-H2; Taxon ID 10950), kindly provided by A. Metzler, Virologisches Institut, University of Zurich, was propagated in African Green Monkey MA-104 cells, kindly provided by C. Eichwald, Virologisches Institut, University of Zurich, as described previously (Arnold et al., 2009). Sufficient virus for all screens and subsequent experiments was pooled, and aliquots stored at -20°C. Thawed virus was activated with 10 µg/ml trypsin (Sigma-Aldrich) for 120 mins at 37°C, prior to 3.75-fold dilution with serum-free McCoy's medium (SFM) supplemented with 0.05x protease inhibitor cocktail (Roche), and then left to stand for 1 hour to equilibrate to RT. Media from reverse transfected cells was aspirated to 20 µl before addition of 20 µl virus (to a final dilution of 7.5-fold – an MOI of ~0.2), and incubated for 2 hours at 37°C. Cell plates were then washed three times with 80 µl media before a further 6 ½-hour incubation, unless otherwise specified. Cells were fixed with 4% formaldehyde for 30 mins at RT, before washing three times with phosphate buffered saline (PBS) prior to storage in PBS + 1% pen/strep at 4°C until immunofluorescence.

For infectious particle quantification, after 10 hours of virus infection according to the standard infection protocol described above, 30 µl of supernatant was collected from each well and transferred to a fresh 384-well plate containing 1.5 µl of 0.2 mg/ml trypsin per well. This was incubated at 37°C for 90 mins. 20 µl of this activated virus was then used to infect a further 384-well plate seeded with 2,000 HCT 116 cells 48 hours previously. Following addition of activated virus, cells were incubated for 8 hours, without any washing steps, then fixed before immunofluorescence.

Immunofluorescence

All steps were carried out at RT. Fixed cells were permeabilised with 0.1% (v/v) Triton X-100 in PBS for 35 mins, washed three times with PBS, then blocked with 1% (w/v) bovine serum albumin (BSA), 50 mM NH_4Cl in PBS for 45 mins. Primary antibody targeting rotavirus capsid protein VP6 (AbD serotec OBT0882) was incubated on cells at a final concentration of 200 ng/ml in blocking solution for 2 hours. Cells were then washed three times with PBS before incubation with anti-mouse AlexaFluor488 secondary antibody (Life Technologies) at 1:1000 in blocking solution for 1 hour, followed by another three washes with PBS. Nuclei were detected by incubation of cells with 1 μg/ml DAPI in PBS for 30 mins. Following a final three washes with PBS, cells were stored in PBS + 1% pen/strep at 4°C until imaging.

In non-screening experiments, immunofluorescence was performed as above, with the exception that cells were fixed with 4% paraformaldehyde warmed to 37°C, and additional antibodies targeting the following proteins were used: rotavirus NSP5 (guinea-pig (gp) Ab at 1:500), kindly donated by Oscar Burrone, ICGEB Trieste; calreticulin (rabbit (r) Ab at 1:1,000, Abcam ab2907); giantin (rAb 1:1,000, Abcam ab24586); TOMM20 (mouse (m) Ab at 1:200, (Abcam ab56783); PABPC1 (mAb at 1:200, Santa Cruz Biotechnology sc-32318); phospho(Ser235/236)-S6 ribosomal protein (rAb at 1:500, Cell Signalling Technology #4858); DDX6 (rAb at 1:500, Bethyl Laboratories A300-461A); nucleophosmin-1/B23 (mAb at 1:200, Sigma-Aldrich B0556); LC3b (rAb at 1:200, Cell Signalling Technology #2775); dynamin-2 (rAb at 1:500, Abcam ab3457); Sam68 (rAb at 1:200, Santa Cruz Biotechnology sc-333); eIF4G (rAb at 1:200, Cell Signalling Technology #2498); casein kinase 1 alpha (rAb at 1:200, Abcam ab63373); phospho(Thr172)-AMPK (rAb at 1:100, Cell Signalling Technology #7535); and, Cyclin B1 (rAb at 1:200, Cell Signalling Technology #12231). The above primaries were detected with highly cross-adsorbed secondary antibodies anti-guinea pig AlexFluor568 (A11075), anti-mouse AlexFluor488 (A11029) and anti-rabbit AlexFluor488 (A11034), all at 1:500.

For lipid droplet detection, cells were incubated with BODIPY dye conjugated to AlexFluor488 (Life Technologies) at 1:2,000 in PBS during secondary antibody incubations. Cell outlines and protein concentrations were determined by reacting cells with CellTrace, a carboxylic acid, succinimidyl ester conjugated to AlexFluor647 (Life Technologies), at 0.3 ng/μl in carbonate buffer (0.1 M $NaHCO_3$, 25 mM Na_2CO_3) for 5 mins, after DAPI staining for 5 mins.

Liquid handling and batch processing
Experimental procedures for high-throughput RNAi screening of rotavirus infection, namely reverse transfection of cells and rotavirus infection, were conducted using an integrated BioTek EL406 liquid handler, a Twister II microplate handler (Caliper), and Liconic rotating incubator (Caliper), to minimise plate effects and increase assay robustness. Both primary and secondary screens consisted of 60 plates per biological replicate, which were processed within a 24-hour period assayed in five batches of 12, staggered by 80 mins, to permit exactly the same incubation times for every stage of the assay for every plate. Immunofluorescence was performed on batches of 12 plates using a BioTek EL406, which enabled two staggered batches to be processed within a 13-hour period.

High-content imaging
Screen plates were imaged using a CellVoyager 7000 (Yokogawa) in epifluorescence mode, a 10x Olympus objective of 0.4 N.A, and a Neo sCMOS camera (Andor, 2,560 x 2,160 pixels). For other experiments, imaging was conducted with the same microscope but utilising the enhanced CSU-X1 spinning disk (Microlens-enhanced dual Nipkow disk confocal scanner, wide view type) and a 60x water immersion Olympus objective of 1.2 N.A.

Image analysis
All images were analysed with the open source software CellProfiler (Carpenter et al., 2006). Images were subject to illumination correction, as previously described (Stoeger et al., 2015), with the adaptation that for 10x images, empty regions of the image (outside the well border) were masked in a site-specific manner.

In 10x images, nuclei were detected using smoothed DAPI images and an adaptation of a spot detection algorithm (Battich et al., 2013), followed by applying the watershed (Mixture of Gaussian – Adaptive) algorithm on 5-pixel expanded spots. Cells were defined as 10-pixel expansions of nuclei. For 60x images, nuclei and cells were identified based on a watershed of DAPI and CellTrace signals, respectively, using iterative segmentation, as described in detail previously (Stoeger et al., 2015). The cytoplasm was considered the cell area minus the nuclear area. In addition, the perinuclear region of each single cell was defined by a 30-pixel expansion around nuclei, and NSP5 spots were detected using spot detection and 5 steps of deblending based on an expected spot size of 5 pixels (Battich et al., 2013).

Single cell features were extracted using standard CellProfiler modules. Images were subject to background subtraction before measurement of features, including intensity (11 features per object and channel), texture (15 features per object and channel), and shape (10 features per object). CellProfiler texture measurements, using various algorithms, quantifies local pixel intensity variations and, as such, is useful for quantifying changes in the degree of signal structure. Single cell population context measurements, such as cell crowding and whether a cell resided at an islet edge or not, were obtained as described previously (Snijder et al., 2012). Supervised vector machines were trained, using CellClassifier (Ramo et al., 2009), for both data quality control (ie removal of (1) debris, and (2) poorly segmented, (3) apoptotic, and (4) mitotic cells from the dataset), and classification of infected cells. Submission to the ETH high performance computer cluster (Brutus) of all jobs, including image compression (*.tiff to *.png without information loss), CellProfiler pipelines, population context measurements and SVMs, were automated using the image analysis platform iBRAIN (Snijder et al., 2012).

Single cell data correction of RNAi screens

For screening datasets, the infection index (II) was defined as the fraction of SVM-classified infected cells per well. On average, perturbed wells contained ~$6x10^3$ cells. Wells with less than 625 cells were excluded from the dataset. This "cell killer" threshold was calculated by fitting three Gaussians to the cell numbers per well, based on the assumption that perturbations would (a) accelerate cell division, (b) decelerate cell division, or (c) kill cells. The threshold of 625 cells was the intercept point between the latter two distributions, at which there is a 50% probability of the perturbation resulting in cell death.

To validate that the genes reducing cell number to <625 per well were not a result of technical problems, we performed functional annotation enrichment analysis on these genes, using the DAVID functional annotation tool (Huang et al., 2009; 2008). We also compared our cell killers (n = 1,332) with those identified by recently published CRISPR/Cas9 screens that sought to identify genes essential to cell viability. Firstly, we pooled the list of essential genes identified in three papers which used several different cell lines ((Blomen et al., 2015; Hart et al., 2015; Wang et al., 2015), which we termed, "Compiled Essential Genes" (n = 3,524). Secondly, we extracted the list of essential genes identified by CRISPR/Cas9 screening in HCT 116 cells (n = 2,070) (Hart et al., 2015). Thirdly, as a comparison control, we prepared a list of 1,163 genes found to most strongly reduce infection in the primary screen i.e. the unique genes from the top-ranked 2,000 genes in each duplicate. Gene overlap between lists were visualised using the Matlab function venn.m.

Given that, under unperturbed conditions, rotavirus displays a preference for infection of less crowded cells, we corrected the data for perturbations that change the fraction of crowded cells and thereby indirectly affect the fraction of infected cells. To exclude such population context-mediated effects of the perturbation on infection, II values were subject to quantile multidimensional binning (QMB) correction, as previously described (Snijder et al., 2012), using the following single cell features: cell crowding (12 bins), edge/non-edge (2 bins), interphase/mitotic (2 bins), and total nuclei area (12 bins). Bin edges were derived from mean bin edges calculated from QMB modelling per plate, and then applied to every well of each biological replicate in the screen. The resulting QMB models reflected the expected II for each well, which were used to correct the observed II per plate using a variation of the z-score in which each well $\log_2$(II) was normalised by the average expected (QMB model) $\log_2$(II) per plate (Snijder et al., 2012). Log_2(II) were also b-score corrected, that is normalised for row or column effects (Boutros et al., 2006). A last z-scoring over all wells for each screen replicate was then performed to give the final corrected $\log_2$(II) per well.

Candidate host factor selection

To increase the likelihood of detecting mild, yet consistent, phenotypes arising from gene perturbation, an inclusive threshold was applied to the primary screen to select genes for validation screening. Therefore, the 1,002 genes with the mean corrected $\log_2$(II) that deviated the most from 0 (+/-1.4σ) without significant cell death (266 with corrected z-scored II >1.4sigma, and 726 with corrected z-scored II <-1.4sigma, in both replicates) were selected for validation. A further 98 genes of various phenotypes, including 13 genes whose II was beyond +/-1.4σ in a single replicate only, were also included in secondary screens to facilitate computational analyses and estimation of primary screen false negative rates.

Gene phenotype scoring

Biological replicates were pooled by calculating the mean corrected $\log_2$(II) per gene, for each of the five RNAi screens. The resulting five-screen datasets were aggregated to rank genes on the basis of the strength of their infection phenotype on knockdown. We employed a Matlab implementation of the Robust Rank Aggregation method to test the significance of RNAi consistency (Kolde et al., 2012). Firstly, the 1,100 genes in all five screens were separately ranked and assigned a normalised ranking (between 0 and 1), based on the number of genes per screen, thus accounting for incomplete gene lists, as in the case of the esiRNA screen. This was done for both gene perturbations that decreased infection the most ranked highest (down-hit ranking), and for gene perturbations that increased infection the most ranked highest (up-hit ranking). Each gene in the resulting normalised rank matrices were assigned a significance score for ranking consistently better than expected under the null hypothesis of uncorrelated inputs, corrected against bias for multiple hypothesis testing (Bonferroni correction). Since the number of ranks that would indicate a role in infection is unknown, the binomial probability of a random distribution was calculated iteratively for increasing ranks ($k_{1\text{-}n}$), and the minimum p-value from each vector was used as the gene's probabilistic aggregation score (PAS). Thus each gene was assigned two probability aggregation scores (PAS), PAS_{down} and PAS_{up}, derived from rankings for both down- and up-hits, respectively. The full RNAi screen datasets and PAS values can be browsed at http://rotavirus.infectome.org.

Aggregation method comparison

The PAS was compared to three other screen aggregation methods: (1) the mean (over screens) of mean (over replicates) corrected $\log_2$(II) per gene; (2) the median (over screens) of mean corrected $\log_2$(II) per gene; and, (3) a phenotype score derived from hypergeometric probabilities, similar to that employed previously (König et al., 2007). For the PAS, phenotype scores are derived from consideration of separate ranked screens. For the hypergeometric-based phenotype score, rather all screens were pooled and ranked. Then the probability of drawing up to x (number of times gene drawn ie 1-4 or 1-5) of a possible K items (number of times gene present in all screens ie 4 or 5) in N drawings (1-n ranks) without replacement from a group of M objects (sum total of number of genes in all screens ie (1,100 x 4) + 867 = 5,267), using a cumulated hypergeometric distribution function was calculated.

Methods were compared with two datasets, both derived from the actual screen data. In the first, we generated a simulated screen dataset with some simulated host factor positive controls, that is elements that should be preferentially ranked higher, termed planted elements. We then compared how well the different methods separated those planted elements from noise. Therefore, we created five simulated lists of 1,100 genes with 100 "hits" or planted elements, defined as the top-100 mean corrected $\log_2$(II) with addition of noise randomly selected from an exponential distribution with lambda of 0.5, and the remaining 1,000 genes with noise added from a standard normal distribution with unit variance. Each of the four aggregation methods, PAS, mean, median and hypergeometric, were then applied to this simulated screen dataset, and the entire procedure was bootstrapped 100 times. Methods were compared by their ability to recover the 100 planted elements, and also how well the degree of recovery, or true positive rate (TPR), compared to the degree of incorrect recovery, or false negative rate (FNR).

In the second method comparison, the TPR/FNR ratio for the top 100 ranked genes was compared with increasing amount of noise, where noise was defined as addition to the original screen data of N lists of randomly shuffled screen data. Thus, for addition of increasing numbers of 0-20 shuffled lists, the fraction of noise added to the actual screen data (5 lists) increased from 0-80% (ie 20 of a total of 25 lists are "noise").

Gene specific phenotype analysis

Gene specific phenotype (GSP) prediction matrices for all genes represented in the Dharmacon genome-wide library were kindly provided by Fabian Schmich and Niko Beerenwinkel, Department of Biosystems science and Engineering, ETH Zurich. The GSP is the predicted on-target infection phenotype for a given gene, derived from subtracting the predicted cumulated off-target effects of each siRNA in the transfected pool from the observed phenotype (Schmich et al., 2015). Only genes with a predicted on-target effect have a GSP that deviates from 0, with those <0 predicted down-hits (required for infection, thus decrease infection indices on knockdown), and those >0 predicted up-hits (inhibit infection, thus increase infection indices on knockdown).

The 1,000 genes with the strongest GSP, at +/- 0.1268σ, were tested for validation by quantifying what fraction of genes with a PAS below a particular threshold were predicted to have a GSP in the top 1,000. GSP values for all 1,100 genes in the validation screens were also compared against the respective PAS. Enrichment for regions between various PAS values (PAS = 0-0.05; PAS = 0.05-0.5; PAS = 0.5-0.95; PAS = 0.95-1.0) were then calculated taking into account the skewed distribution of genes between up (13.3% >0) and down (40.4% <0) predicted GSPs, to generate quandrant enrichment factors, a measure of the relationship between the distribution of predicted GSPs and validated hits, the latter given by the PAS.

Quantitative reverse transcription polymerase chain reaction (qRT-PCR)

To confirm host factor gene knockdown on treatment with siRNAs, 5,250 HCT 116 cells were seeded in a 96-well plate containing 60 µl of transfection reagents, comprising Ambion Silencer® Select siRNA, 0.4 µl lipofectamine RNAiMAX (Life Technologies), and OptiMEM, resulting in a final siRNA concentration of 5 nM, in duplicate wells per siRNA. Cells were then incubated at 37°C, 5% CO_2 for 72 hours, and RNA isolated using the NucleoSpin® 96 RNA isolation kit (Macherey-Nagel). Reverse transcription was carried out using the Transcriptor High Fidelity cDNA Synthesis kit (Roche), along with no reverse transcriptase (RT) controls, using random hexamer primers. The resulting cDNA was combined with SYBR® Green Real-Time PCR Mastermix (Thermo Fischer) for quantitative PCR reactions in technical triplicates, using the following primers: *DNM2* Fw ACTGTCCTGGTACAAGGATGAG and Rev AGACGTGCTTGTTGGACATG; *CSNK1A1* Fw CAGTGGGGAAGAGGAAAAGAAG and Rev TGTGTTGCCTTGTCCTGTTG; *ABCF1* Fw TGCACTCAAGGGCAAAAAGG and Rev TGTTTGGGAGGCTCCTTTTC; *COPG1* Fw AACACGCCGTCCTTATGAAC and Rev AGCGCGCTTATACTCAAAGC; *WDR46* Fw TATGCCAGGCTGACATTGTG and Rev AACTGCCGCAGATTCAAGTC; *DDX52* Fw TGTCATCAGTGTGTCCATTGG and Rev TCTCACGGCCAGAAGTTTTC; *REEP2* Fw CTGGATCGTCTTTGCCTTCTTC and Rev CCATATCACGAAGGCGATCTTC. Expression levels of mRNA were calculated relative to three housekeeping genes, *EEF1A1*, *TFB* and *TFRC*.

Functional annotation enrichment analysis

We employed a novel rank-based functional annotation (FA) enrichment approach, developed from a previously published method based on hypergeometric probability calculations (Liberali et al., 2014), but introducing a rank-based iterative procedure. PAS values were used to rank genes, both from those that most decreased infection on perturbation (down-hits) and from those that most increased infection on perturbation (up-hits). A table of human genome-wide functional gene annotations was downloaded from the DAVID database (Huang et al., 2009; 2008), from which a logical matrix of 18,037 genes x 20,606 annotations was created. This was used to test whether each of the 20,606 FAs was significantly enriched for iteratively increasing numbers of ranked genes using a hypergeometric probability distribution function, as follows:

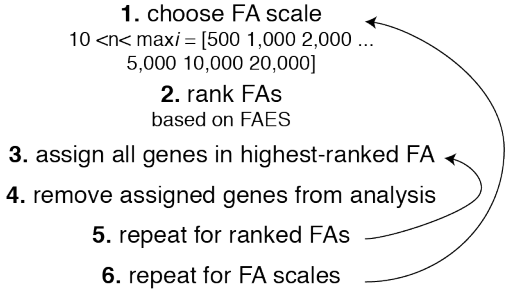
$$y = f(x|M,K,N) = \frac{\binom{K}{x}\binom{M-K}{N-x}}{\binom{M}{N}}$$

where the result, y, is the probability of drawing exactly x (number of genes, above rank n, in annotation) of a possible K items (number of genes in a FA) in N drawings (1-n ranks) without replacement from a group of M objects (18,037 genes). The resulting p-value vectors were smoothed by taking the mean over a window of 10 consecutive values. The final FA

enrichment score (FAES) was stored as the minimum *p*-value from these vectors of length *n*, where $PAS_{1\text{-}n} < 0.5$, since at PAS = 0.5 the probability of a gene conferring an infection phenotype is 50-50, and thus rankings beyond this point are not informative for FA enrichment analysis. These rank limits were $n = 359$ and $n = 399$ for down and up FA enrichment analysis, respectively. The rank from which FAES values were derived were also stored. These attributes were inputs for network construction in Cytoscape 2.8.6, using the Organic layout. In Figure 2, the network depicts only FAs with more than 20 genes, and also with either a $FAES_{down}$ of ≤0.05 or a $FAES_{up}$ of ≤0.02. The entire FA enrichment analysis can be browsed in Table S4. The full functional annotation data can be browsed at http://rotavirus.infectome.org.

Greedy functional annotation gene assignment and enrichment

To facilitate visualisation of aggregated screen data, genes were assigned to an individual FA, using a number of constraints: (1) FAs with complete overlap with another were removed; and, (2) FAs with less than 10 genes were removed. The remaining FAs with 10-500 genes were ranked based on their FAES. Genes annotated to the highest-ranked FA were then assigned this greedy FA and removed from the analysis. This process was repeated for all FAs with between 10 and 500 genes. The upper limit on FA size was then increased to 1,000 genes, and the rank-based greedy gene assignment process repeated. This was then repeated several times for increasing upper limits on FA size, namely 2,000, 5,000, 10,000 and 20,000, as depicted in the schematic below. Since there are two FAES per FA ($FAES_{down}$ and $FAES_{up}$), a gene was greedily assigned to the FA derived from the ranking that complemented the lowest PAS for that gene. Thus, if for gene *x*, $PAS_{down} < PAS_{up}$, gene *x* was more likely required for infection than inhibiting infection, and therefore greedily assigned a FA based on $FAES_{down}$ rankings.



Hierarchical clustering with Euclidean distance metric and average linkage was performed on those FAs assigned at least one validation gene, based on their degree of overlap ie the fraction of genes in common between those FAs. The resulting dendrogram was manually summarised into the 27 greedy FA groups coloured in Figure 1E. Enrichment for greedy FA groups was defined as a proportion of validation genes in those groups higher than that observed in the genome, and therefore represented more than expected in the validation screening datasets.

Threshold-based functional annotation enrichment analysis

Genes with a PAS<0.1 in either down- or up-based rankings ($n = 207$) were used as inputs in the online DAVID functional annotation tool (Huang et al., 2009; 2008), which tested for enrichment against all human genes. The resulting functional annotation clusters identified were considered significant if they had an enrichment score (E) >1.0.

Gene interaction networks

Protein-protein interaction data was obtained from the STRING v10 database (Snel et al., 2000; Szklarczyk et al., 2015) filtering for only experimentally validated interactions with a medium confidence interaction score (>0.4). The mean II between replicates from the primary genome-wide screen was used to colour nodes, with those that resulted in total cell numbers per well <625 in both replicates coloured yellow. Those genes also present in the validation screens were given square nodes sized according to their PAS.

Branched DNA single molecule RNA fluorescence *in situ* hybridisation

Branched DNA (bDNA) single molecule RNA (smRNA) fluorescence *in situ* hybridisation (FISH) was performed as described previously (Battich et al., 2013) on HCT 116 cells seeded in complete medium in 384-well plates 72 hours before rotavirus infection assay for 0-8 hours, fixation and probing. Type 1 probe sets (Affymetrix) were specifically designed to target the following rotavirus A DS-1 segments: segment 1+, region 421-1,933 (Accession HQ650116, cat# VF1-17346-01), segment 2+, region 2-1,418 (Accession HQ650117, cat# VF1-17348-01), segment 2-, region 899-2,278 (Accession HQ650117_N, cat# VF1-17347-01), segment 5+, region 108-1,532 (Accession HQ650120, cat# VF1-17350-01), segment 6+, region 82-1,316 (Accession HQ650121, cat# VF1-17351-01), and segment 11+, region 4-780 (Accession HQ650126, cat# VF1-17345-01). In addition, negative control probe sets targeting bacterial transcript *dapB* (Entrez ID 944762, cat# VA1-10272) and positive control probe sets targeting human gene *HPRT* (Entrez ID 3251, cat# VA1-11124) were used. Cellular poly-adenylated (poly(A) transcripts were detected using probe sets composed of extended regions of poly-thymidines (Affymetrix). Each time point was assayed by each probe set in quadruplicate. The probe sets and hybridisation reagents build large fluorescent trees on each transcript permitting single mRNA detection as an individual spot in microscopy images. Spot detection and quantification was performed by CellProfiler pipelines using the custom modules previously developed (Battich et al., 2013; Stoeger et al., 2015). For poly(A) FISH, integrated cell intensity was also quantified as a surrogate for the number of transcripts. The number of spots per cell detected on probing for *dapB* could be used as a background threshold of mRNA molecule detection.

Virus infection axis

As a result of the observed asynchrony in infection progression over time, we developed a computational method to order cells along infection progression, so as not to rely on experimental synchronisation. Such a computational method requires 1) heterogeneity in infection progression, and 2) a readout that captures as much of this heterogeneity as possible. Therefore, we could have ordered infected cells based on VP6 intensities, which increase over infection progression. However, given the fact that NSP5 is one of two viroplasm scaffold proteins, this provided additional information at early-stage infection not captured by NSP5 or VP6 intensity alone. Therefore, here we harness the properties of the NSP5 signal, which transitions from tiny spots, to diffuse cytoplasmic signal with spots of variable sizes, to very bright cells with large perinuclear inclusions. Capturing these changes with multivariate single-cell features of the NSP5 signal increases the resolving power of our trajectory, particularly for early-stage infection events when intensity features of viral antigens are largely unchanged.

HCT 116 cells grown in complete medium in 384-well plates for 72 hours were infected with activated rotavirus DS-1 for 1 hour, prior to washing, and a further incubation to total 3, 6, 9, 12 or 15 hours post-infection. Control wells were treated with non-virus containing media for 3 or 15 hours. Single cell measurements from CellProfiler pipelines were filtered for those of infected cells, defined as those with a mean NSP5 cytoplasmic intensity >=0.001 and a NSP5 spot count of >=2 per cell.

Given that CellProfiler pipelines output a large number of single-cell features, we needed to select which features of the NSP5 signal to use in construction of the virus infection progression trajectory. As stated above, CellProfiler texture measurements, through the quantification of local pixel intensity variations, are useful for capturing changes in the degree of signal structure. For the NSP5 signal, texture measurements can therefore be very informative in capturing the emergence and growth of viroplasms, providing additional information to intensity measurements. Firstly, z-scored mean values for all measurements related to virus specific or general cellular features were calculated over the infected cells per well (see also Table S6 for a full list of these CellProfiler measurements). Secondly, hierarchical clustering with Euclidean distance metric was used to visualise the pattern of feature changes over the duration of virus infection. Thirdly, those features showing step-wise changes over time in their mean values in infected cells were empirically tested in various combinations for their ability to construct a virus infection progression trajectory. For example, variation in the texture of CellTrace, a dye that non-specifically labels proteins, showed least overlap between 9 and 12 hours p.i., reflecting the late-stage maturation of viroplasms to protein-dense structures in the perinuclear region.

We performed infected cell gating and trajectory construction using Cycler, an adaptation of the Wanderlust (Bendall et al., 2014) algorithm, that was developed for the construction of cell cycle trajectories from fixed images of cells (Gut et al., 2015). Briefly, Cycler uses an input of multivariate single cell features to perform a *k*-nearest neighbour graph-based embedding of cells into a single dimension representing progression. This process is repeated for the same inputs 10 times, with the output reflecting a mean of those 10 iterations. Infected cells were gated from the total infection time course population by plotting single cell NSP5 cell sum average texture (NSP5textureSA) against NSP5 cytoplasm mean pixel intensity (NSP5intensity), and excluding cells that were both $\text{NSP5intensity}_{low}$ and $\text{NSP5textureSA}_{high}$, where cells in no virus control wells were enriched. Cycler requires a start population, which was gated from the remaining cells based on $\text{NSP5intensity}_{low}$, $\text{NSP5textureSA}_{high}$ and low NSP5 cytoplasm sum variance texture ($\text{NSP5textureSV}_{low}$).

A successful trajectory was defined in three ways: (1) ordered infected cells such that the features used in trajectory construction varied in ways observed in the time course experiments; (2) ordered cells such that the cytoplasmic intensity of viral antigen VP6, when overlaid, increased over trajectory progression, despite not being used in trajectory construction; and, (3) was robust to inputs from different experiments performed weeks apart. Tests were conducted and visualised using the Matlab-based Cycler graphical user interface (http://www.cellcycler.org/). The following six features, after z-scoring over all time points, were tested for their ability to construct meaningful trajectories: NSP5textureSA, NSP5textureSV, CellTrace perinuclear angular second moment texture, DAPI nuclei sum average texture, NSP5 spots per cell, NSP5intensity. The combination of the first two and latter two features were deemed to produce the most meaningful and reproducible trajectories. The final algorithm parameters used were: 10 iterations, 100 landmarks/waypoints, Euclidean distance metric, cell neighbourhood of 8, and 15 randomly selected graphs from the cell neighbourhood. Detailed explanations of these parameters have been provided previously (Gut et al., 2015). Once infected cells were assigned a trajectory value, these were given a normalised ranking, ie ordered between 0 and 1, to generate the final virus infection axis (VIX) positioning. Cellular features from infected cells, such as the intensity (to capture changes in concentration) or texture (to capture changes in structure) of a cellular marker normalised between 0 and 1, could then be visualised over the course of the VIX.

Nascent protein synthesis

To quantify nascent protein synthesis rates, HCT 116 cells were grown in complete medium in 384-well plates for 72 hours and assayed using the Click-iT® AHA AlexaFluor488 Protein Synthesis HCS Assay kit, as per manufacturer's instructions (Life Technologies). Briefly, cells were washed four times with methionine (Met)-free medium (Gibco), then fed with an amino acid analogue of Met containing an azido moiety at a final concentration of 50 μM for 30 mins, plus either 0.1% DMSO or 50 μg/ml cyclohexamide in DMSO. Following fixation, incorporated amine acid was detected in a "click" reaction using an AlexaFluor488-modified alkyne.

Drug treatments
To assess the impact of small compounds on virus infection progression, HCT 116 cells were grown in complete medium in two 384-well plates for 48 hours. Media was aspirated from cells to a residual volume of 20 μl per well, to which 20 μl of drugs or solvent in complete media were added to 12 wells each, distributed over two plates. Stock solutions of AICAR (5-Aminoimidazole-4-carboxamide 1-β-D-ribofuranoside, Acadesine, N^1-(β-D-Ribofuranosyl)-5-aminoimidazole-4-carboxamide**,** Sigma-Aldrich) and metformin (Sigma-Aldrich) were prepared in distilled water, whereas solutions of dorsomorphin (Sigma-Aldrich), cyclohexamide (Sigma-Aldrich) and actinomysin-D (Sigma-Aldrich) were prepared in DMSO. Following addition to cells, drug concentrations were 4.5 mM AICAR, 75 mM metformin, 2 or 10 μM dorsomorphin, 75 μg/ml cyclohexamide and 1 μg/ml actinomysin-D. After a one hour incubation, 20 μl of activated virus was added, thus diluting drug concentrations 1.5-fold from those values quoted above. Cells were then incubated a further 7 hours before fixation, immunofluorescence and imaging.

For the analysis of drug effects on markers of various cellular processes, HCT 116 cells were incubated with drugs or solvent controls for five hours only. In addition, some wells were washed three times with serum-free medium (SFM) before addition of solvent controls (water or 0.1% DMSO) in SFM for five hours, to induce acute starvation. Each condition was performed in duplicate wells for each cellular marker immunofluorescence.

Selected host factor siRNA perturbations
To assess the impact of host factor knock down on virus infection progression, HCT 116 cells were grown in complete medium in a 384-well plate for 72 hours following reverse transfection with Lipofectamine RNAiMAX (Life Technologies) and Silencer® Select siRNAs (Ambion) for selected genes, as those used in validation screen library sets 2 and 3 ie individual siRNA-2 and siRNA-3, in separate wells in duplicate. Cells were infected with activated rotavirus DS-1 for one hour, washed three times with media and then incubated for a further 7 hours before fixation and immunofluorescence. Cells were assayed for levels of the viral antigens VP6 and NSP5, with the exception of REEP2 knockdown, which was assayed for NSP5 and the ER marker calreticulin. The siRNA whose mean corrected $\log_2$(II) was closest to -2 in the validation screens was the siRNA perturbation selected for infection progression analysis (see below).

Mapping perturbations to VIX
Typically, staging the role of host factors in pathogen infection is achieved by time course experiments and/or the use of multiple infection stage-specific assays. However, with our method to infer a trajectory of virus infection, we examined if it was possible to obtain such functional information simply from the analysis of alterations to the trajectory in one fixed population of cells at 8-9 hours p.i. upon host gene knockdown.

Given that the VIX is an ordering of cells, mapping the impacts of perturbations to virus infection progression without changing the overall behaviour of the trajectory was achieved by spiking small numbers of infected cells from perturbed populations into a large number of control cells. Since the number of infected cells in perturbed populations could be relatively small, infected cells were gated from pooled populations of cells from all conditions, perturbed and non-perturbed, according to NSP5 features and where cells from no virus control wells were enriched, as described above in "Virus infection axis". Following this gating, 5,000 infected cells, randomly selected from control wells only that were not transfected with siRNAs, were the basis of the trajectory. Start populations of 1-10 cells were gated from this control population. Added to the control population of 5,000 cells, 40 perturbed cells were "spiked", which were randomly selected from the infected cells in a single siRNA-transfected well. Trajectory values were then computed for these 5,040 cells over 10 iterations, using the same parameters as above, and stored as a single bootstrap. This process was repeated with new random samplings of 40 perturbed cells for 25 bootstraps, resulting in 1,000 trajectory values for each perturbed population. Distribution shifts between trajectory values of control and perturbed populations was assessed with the Kolmogorov-Smirnov (KS) test. Given the large numbers of single cells in our typical samples, *p*-values are often uninformative as they are almost always significant. We have elected to display the KS test statistic in figures to give a more informative description of the degree of distance between the two test distributions. In practice, we observe that a significant KS test *p*-value results in a KS statistic >0.09. Where KS statistics are shown in the absence of a significant *p*-value, this is explicitly stated in the corresponding figure legend.

A similar procedure was employed for calculating trajectory values for cells perturbed with drug treatments on infection except that, because the number of infected cells was reduced, the control population comprised 1,000 cells and 20 perturbed cells were spiked per bootstrap.

Control cell trajectory values were normalised to VIX positions between 0 and 1 as before. Perturbed cells were then assigned a VIX position equal to that of a control cell that was the nearest neighbour, as determined from trajectory values. This resulted in perturbed infected cells being mapped along the VIX, an ordering of infection progression given by non-perturbed populations. To facilitate visualisation of VIX distribution shifts, the fraction of cells within 10 linearly-spaced bins along the VIX was calculated for perturbed cell populations. As VIX is an ordering based on the control population, 10% of this group of cells were therefore present in each bin, whereas the fraction of perturbed populations in each bin could vary along the VIX.

To verify that perturbed cell spiking did not disturb the overall behaviour of the trajectory, the trajectory values for control cells were compared between bootstraps and between perturbations, using Pearson's correlation. To examine the reproducibility of perturbed-cell spiking, duplicate wells were analysed separately. Furthermore, the standard deviation in the fraction of cells present in each VIX bin for quadruplicate wells non-transfected or transfected with non-targeting siRNA (siScrambled) was calculated to estimate the technical error and reproducibility, respectively, in VIX distribution shifts.

Cell cycle phase classification

HCT 116 cells grown in 384-well plates were infected with rotavirus DS-1 for 8 hours, as described above. Active DNA replication was detected using the Click-iT® EdU AlexaFluor647 Imaging Assay kit (Life Technologies), such that 20 mins before fixation, cells were incubated with 100 μM EdU solution. Using "click" chemistry, the incorporated EdU was detected as a bright nuclear signal in the FarRed microscopy channel (mean nuclear pixel intensity >0.004 in S phase cells), in addition to a comparatively dim cell staining with CellTrace, allowing simultaneous detection (and discrimination) of S phase cells and cell outlines (Gut et al., 2015). Immunofluorescence for Cyclin B1, which is expressed weakly in S phase cells and strongly in G2 cells, was used to discriminate these cell cycle phases from G1 and M using a threshold of mean cytoplasmic pixel intensity of 0.001. The integrated nuclear intensity of DAPI staining was used to discriminate G1 cells (<4.5) from S and G2 cells, as in the latter categories the DNA is >2n. Using these three thresholds on DAPI, EdU and Cyclin B intensities, cells could be classified into one of G1, S and G2 phases. M phase cells, identified by SVM using DAPI intensity and texture features, were excluded, along with cells that did not meet the cell cycle phase classification constraints outlined here.

The effects of cell cycle phase on infection progression were assessed by only considering infected cells that were included in the VIX analysis. In contrast, to examine whether virus infection induced changes to cell cycle phase distributions, the proportion of cells in each phase were compared between all cells from control wells not exposed to virus and all cells in a virus-infected well.

Supplemental References

Arnold, M., Patton, J.T., McDonald, S.M., 2009. Culturing, Storage, and Quantification of Rotaviruses. Current Protocols in Microbiology CHAPTER, Unit–15C.3.

Battich, N., Stoeger, T., Pelkmans, L., 2013. image-based transcriptomics in thousands of single human cells at single-molecule resolution. Nat Meth 1–10. doi:10.1038/nmeth.2657

Bendall, S.C., Davis, K.L., Amir, E.-A.D., Tadmor, M.D., Simonds, E.F., Chen, T.J., Shenfeld, D.K., Nolan, G.P., Pe'er, D., 2014. Single-Cell Trajectory Detection Uncovers Progression and Regulatory Coordination in Human B cell Development. Cell 157, 714–725.

Blomen, V.A., Majek, P., Jae, L.T., Bigenzahn, J.W., Nieuwenhuis, J., Staring, J., Sacco, R., van Diemen, F.R., Olk, N., Stukalov, A., Marceau, C., Janssen, H., Carette, J.E., Bennett, K.L., Colinge, J., Superti-Furga, G., Brummelkamp, T.R., 2015. Gene essentiality and synthetic lethality in haploid human cells. Science 350, 1092–1096. doi:10.1126/science.aac7557

Boutros, M., Brás, L.P., Huber, W., 2006. Analysis of cell-based RNAi screens. Genome Biol 7, R66–R66.

Carpenter, A.E., Jones, T.R., Lamprecht, M.R., Clarke, C., Kang, I.H., Friman, O., Guertin, D.A., Chang, J.H., Lindquist, R.A., Moffat, J., Golland, P., Sabatini, D.M., 2006. CellProfiler: image analysis software for identifying and quantifying cell phenotypes. Genome Biol 7, R100–R100.

Gut, G., Tadmor, M.D., Pe'er, D., Pelkmans, L., Liberali, P., 2015. Trajectories of cell-cycle progression from fixed cell populations. Nat Meth 1–10. doi:10.1038/nmeth.3545

Hart, T., Chandrashekhar, M., Aregger, M., Steinhart, Z., Brown, K.R., MacLeod, G., Mis, M., Zimmermann, M., Fradet-Turcotte, A., Sun, S., Mero, P., Dirks, P., Sidhu, S., Roth, F.P., Rissland, O.S., Durocher, D., Angers, S., Moffat, J., 2015. High-Resolution CRISPR Screens Reveal Fitness Genes and Genotype-Specific Cancer Liabilities. Cell 163, 1515–1526. doi:10.1016/j.cell.2015.11.015

Huang, D.W., Sherman, B.T., Lempicki, R.A., 2009. Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. Nucleic Acids Res. 37, 1–13.

Huang, D.W., Sherman, B.T., Lempicki, R.A., 2008. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. Nature Protocols 4, 44–57. doi:10.1038/nprot.2008.211

Kolde, R., Laur, S., Adler, P., Vilo, J., 2012. Robust rank aggregation for gene list integration and meta-analysis. Bioinformatics 28, 573–580. doi:10.1093/bioinformatics/btr709

König, R., Chiang, C.-Y., Tu, B.P., Yan, S.F., DeJesus, P.D., Romero, A., Bergauer, T., Orth, A., Krueger, U., Zhou, Y., Chanda, S.K., 2007. A probability-based approach for the analysis of large-scale RNAi screens. Nat Meth 4, 847–849. doi:10.1038/nmeth1089

Liberali, P., Snijder, B., Pelkmans, L., 2014. A Hierarchical Mapof Regulatory Genetic Interactions in Membrane Trafficking. Cell 157, 1473–1487. doi:10.1016/j.cell.2014.04.029

Rämo, P., Sacher, R., Snijder, B., Begemann, B., Pelkmans, L., 2009. CellClassifier: supervised learning of cellular phenotypes. Bioinformatics 25, 3028–3030. doi:10.1093/bioinformatics/btp524

Schmich, F., Szczurek, E., Kreibich, S., Dilling, S., Andritschke, D., Casanova, A., Low, S.H., Eicher, S., Muntwiler, S., Emmenlauer, M., Rämö, P., Conde-Alvarez, R., Mering, von, C., Hardt, W.-D., Dehio, C., Beerenwinkel, N., 2015.

gespeR: a statistical model for deconvoluting off-target-confounded RNA interference screens. Genome Biol 1–12. doi:10.1186/s13059-015-0783-1

Snel, B., Lehmann, G., Bork, P., Huynen, M.A., 2000. STRING: a web-server to retrieve and display the repeatedly occurring neighbourhood of a gene. Nucleic Acids Res. 28, 3442–3444.

Snijder, B., Sacher, R., ouml, P.R.A.M., Liberali, P., Mench, K., Wolfrum, N., Burleigh, L., Scott, C.C., Verheije, M.H., Mercer, J., Moese, S., Heger, T., Theusner, K., Jurgeit, A., Lamparter, D., Balistreri, G., Schelhaas, M., De Haan, C.A.M., ki, V.M.A., auml, T.H., Rottier, P.J.M., Sodeik, B., Marsh, M., Gruenberg, J., Amara, A., Greber, U., Helenius, A., Pelkmans, L., 2012. Single-cell analysis of population context advances RNAi screening at multiple levels. Molecular Systems Biology 8, 1–18. doi:10.1038/msb.2012.9

Stoeger, T., Battich, N., Herrmann, M.D., Yakimovich, Y., Pelkmans, L., 2015. Computer vision for image-based transcriptomics. METHODS 85, 44–53. doi:10.1016/j.ymeth.2015.05.016

Szklarczyk, D., Franceschini, A., Wyder, S., Forslund, K., Heller, D., Huerta-Cepas, J., Simonovic, M., Roth, A., Santos, A., Tsafou, K.P., Kuhn, M., Bork, P., Jensen, L.J., Mering, von, C., 2015. STRING v10: protein-protein interaction networks, integrated over the tree of life. Nucleic Acids Res. 43, D447–D452. doi:10.1093/nar/gku1003

Wang, T., Birsoy, K., Hughes, N.W., Krupczak, K.M., Post, Y., Wei, J.J., Lander, E.S., Sabatini, D.M., 2015. Identification and characterization of essential genes in the human genome. Science 350, 1096–1101. doi:10.1126/science.aac7041