

Single-cell analysis of population context advances RNAi screening at multiple levels

This article has been corrected since Online Publication. Extra information giving access to an external website has been removed.

Berend Snijder¹, Raphael Sacher², Pauli Rämö^{2,11}, Prisca Liberali¹, Karin Mench¹, Nina Wolfrum¹, Laura Burleigh³, Cameron C Scott⁴, Monique H Verheije^{5,12}, Jason Mercer⁶, Stefan Moese⁶, Thomas Heger⁶, Kristina Theusner^{7,8}, Andreas Jurgeit¹, David Lamparter², Giuseppe Balistreri⁶, Mario Schelhaas^{6,13}, Cornelis AM De Haan⁵, Varpu Marjomäki⁹, Timo Hyypiä¹⁰, Peter JM Rottier⁵, Beate Sodeik⁷, Mark Marsh⁸, Jean Gruenberg⁴, Ali Amara^{3,14}, Urs Greber¹, Ari Helenius⁶ and Lucas Pelkmans^{1,*}

¹ Institute of Molecular Life Sciences, University of Zurich, Zurich, Switzerland, ² Institute of Molecular Systems Biology, ETH Zürich, Zurich, Switzerland, ³ Unité de pathogénie virale moléculaire INSERM 819, Institut Pasteur, Paris, France, ⁴ Biochemistry Department, University of Geneva, Geneva, Switzerland, ⁵ Virology Division, Department of Infectious Diseases and Immunology, Utrecht University, Utrecht, The Netherlands, ⁶ Institute of Biochemistry, ETH Zürich, Zürich, Switzerland, ⁷ Institute of Virology, Hannover Medical School, Hannover, Germany, ⁸ Cell Biology Unit, MRC-Laboratory for Molecular Cell Biology, University College London, London, UK, ⁹ Department of Biological and Environmental Science, University of Jyväskylä, Jyväskylä, Finland and ¹⁰ Department of Virology, University of Turku, Turku, Finland

¹¹Present address: Focal Area Infection Biology, Biozentrum, University of Basel, Klingelbergstrasse 70, CH-4056 Basel, Switzerland

¹²Present address: Pathology Division, Department of Pathobiology, Faculty of Veterinary Medicine, Utrecht University, Yalelaan 1, 3584 CL, Utrecht, The Netherlands

¹³Present address: Emmy-Noether Group 'Virus Endocytosis', Institutes of Molecular Virology and Medical Biochemistry, University of Münster, Von-Esmarch-Strasse 56, D-48149 Münster, Germany

¹⁴Present address: Laboratory of Pathology and Molecular Virology, Team Biology of Emerging Viruses INSERM U944, Institut Universitaire d'Hématologie, Hôpital Saint-Louis, 1 avenue Vellefaux, 75010 Paris, France

* Corresponding author. Institute of Molecular Life Sciences, University of Zurich, Winterthurerstrasse 190, CH-8057 Zurich, Switzerland. Tel.: +41 44 635 31 23; Fax: +41 44 63 56 830; E-mail: lucas.pelkmans@imls.uzh.ch

Received 2.8.11; accepted 9.3.12

Isogenic cells in culture show strong variability, which arises from dynamic adaptations to the microenvironment of individual cells. Here we study the influence of the cell population context, which determines a single cell's microenvironment, in image-based RNAi screens. We developed a comprehensive computational approach that employs Bayesian and multivariate methods at the single-cell level. We applied these methods to 45 RNA interference screens of various sizes, including 7 druggable genome and 2 genome-wide screens, analysing 17 different mammalian virus infections and four related cell physiological processes. Analysing cell-based screens at this depth reveals widespread RNAi-induced changes in the population context of individual cells leading to indirect RNAi effects, as well as perturbations of cell-to-cell variability regulators. We find that accounting for indirect effects improves the consistency between siRNAs targeted against the same gene, and between replicate RNAi screens performed in different cell lines, in different labs, and with different siRNA libraries. In an era where large-scale RNAi screens are increasingly performed to reach a systems-level understanding of cellular processes, we show that this is often improved by analyses that account for and incorporate the single-cell microenvironment.

Molecular Systems Biology 8: 579; published online 24 April 2012; doi:10.1038/msb.2012.9

Subject Categories: computational methods; microbiology & pathogens

Keywords: cell-to-cell variability; image analysis; population context; RNAi; virus infection

Introduction

Cells display inherent variability of both a stochastic and deterministic nature. Stochastic cell-to-cell variability arises owing to the probabilistic and discrete nature of biochemical interactions and of the partitioning of cellular components during division (Elowitz *et al*, 2002; Ozbudak *et al*, 2002; Raj and van Oudenaarden, 2009; Eldar and Elowitz, 2010; Huh and Paulsson, 2010). Deterministic or regulated cell-to-cell variability however makes up the larger part of phenotypic variation in a cell population, and consists of predictable responses of cells to both intrinsic and extrinsic cues, such as

cell-cycle dependent variability (Sigal *et al*, 2006) and adaptations of individual cells to changes in their microenvironment (Shapiro, 1998; Raser and O'Shea, 2005; Snijder and Pelkmans, 2011). In cell culture conditions, growing adherent mammalian cell populations create a continuously changing microenvironment for individual cells by increasing the local cell density and reducing the number of cells that constitute the edges of cell islets, factors that shape the cell population context (Snijder *et al*, 2009; Snijder and Pelkmans, 2011). This leads among others, to the population context-determined remodelling of cell-cell and cell-matrix adhesions (Gumbiner, 1996; Parsons *et al*, 2010), polarization (Takeichi, 1991; Ohno,

2001), and the differential regulation of endocytic pathways (Doherty and McMahon, 2009; Scita and Di Fiore, 2010). Such contextual cellular state changes lead to strong differences in the capacity of viruses to infect individual cells (Snijder *et al*, 2009; Snijder and Pelkmans, 2011).

We previously demonstrated that population-averaged measurements of unperturbed virus infection or endocytic activity can be highly sensitive to differences in the population context (Snijder *et al*, 2009). However, these differences in population-averaged measurements can be predicted by mathematical models that describe the cellular activity as a function of the population context of individual cells (Snijder *et al*, 2009). It is however currently unclear if and how population context effects contribute to the challenges present in RNA interference (RNAi) screens (Echeverri and Perrimon, 2006; Moffat and Sabatini, 2006; Cherry, 2009; Mohr *et al*, 2010). RNAi screens in mammalian tissue culture cells have been very successful in revealing novel molecular regulators (Kittler *et al*, 2004; Pelkmans *et al*, 2005; Collinet *et al*, 2010; Neumann *et al*, 2010), but they have suffered from poor comparability (Bushman *et al*, 2009; Cherry, 2009; Mohr *et al*, 2010). Moreover, the large variability in perturbation phenotypes obtained with different siRNAs targeting the same gene (Collinet *et al*, 2010) have suggested widespread off-target effects. We here present a broad and systematic study of virus infection as well as of other readouts of cellular activities, in which we investigate the interplay between population context-regulated cell-to-cell variability and the establishment of RNAi perturbation phenotypes in cell populations. We find that accounting for population context-mediated effects in RNAi phenotypes strongly changes screening results, improving both the comparability of RNAi screens between different cell lines and the consistency of phenotypes obtained with different siRNAs targeting the same gene. We show that a computational analysis of single-cell phenotypes, which does not average out cell-to-cell variability and is aware of population context, greatly enhances the interpretation of RNAi screens for systems-level approaches and a global understanding of cellular processes. We share both our novel computational methods for measurement and correction of cell population context effects as well as screening results used in this work on <http://www.infectome.org>.

Results

We first studied the infection of 17 different mammalian viruses in 34 small-scale and 7 large-scale image-based RNAi screens, and we analysed the RNAi phenotypes at the single-

cell level across populations of cells. In each screen, we targeted each host gene with three independent siRNAs and repeated each screen three times, every time imaging most of the surface of the well to accurately measure the full cell population context (Figure 1A). For the 34 small-scale screens, we selected a set of 49 human kinases that were previously implicated in virus infection and endocytosis (Supplementary Table 1) (Pelkmans *et al*, 2005), while the large-scale screens were performed on 6978 human genes constituting the druggable genome (Hopkins and Groom, 2002). For an overview of the assays and the workflow, and a full description of methods, we refer to Figure 1 and Materials and Methods. The 17 viruses are from 13 different virus families and include most known combinations of enveloped/non-enveloped, single/double-stranded, and DNA/RNA viruses (Figure 1E and Supplementary Table 2). We used four strains of human HeLa cell lines obtained from different sources (see Supplementary Methods), which despite their similar origin demonstrated striking differences in their population-determined phenotypic properties (see Supplementary Figure 8), as well as in their ability to be infected by the different viruses. Transfection controls confirmed that all four cell lines were capable of achieving high and reproducible RNAi-mediated silencing efficiencies, which is also evident from their widespread use in published RNAi screens (Kittler *et al*, 2004; Pelkmans *et al*, 2005; Collinet *et al*, 2010; Neumann *et al*, 2010). In total, we measured over 200 quantitative features of each of 2.4×10^9 individual cells, based on which we classified each cell into multiple phenotypic classes using supervised machine learning (Figure 1). For the 34 small-scale screens, replicate measurements of infection showed an average correlation coefficient of 0.8, demonstrating that the screens were of high technical quality (Supplementary Figure 9).

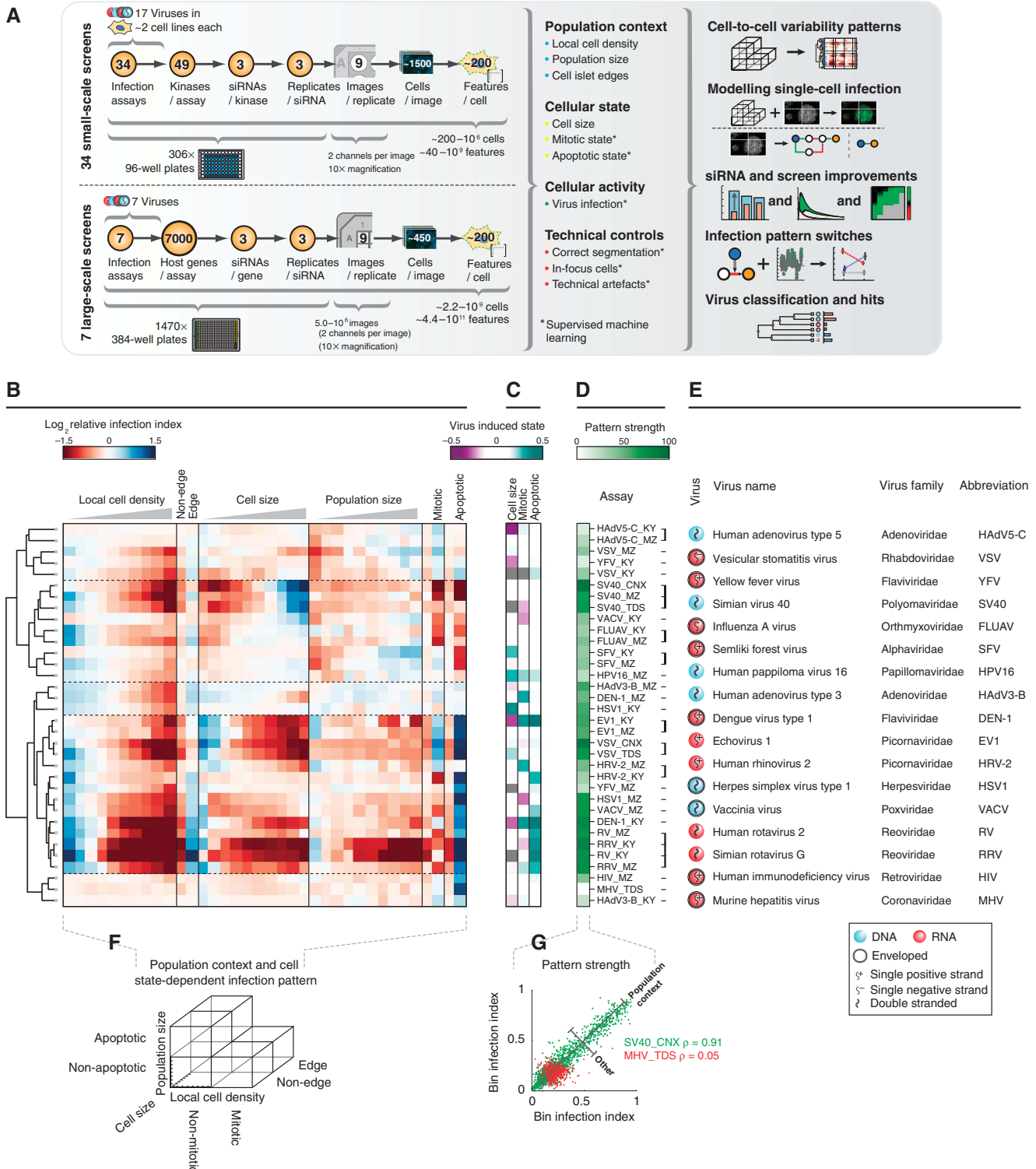
Population context determines the cell-to-cell variability of virus infection in RNAi screens

We previously reported that the cell-to-cell variability observed in the infection efficiency of four mammalian viruses is to a large extent determined by the cell population context, rather than the consequence of stochasticity (Snijder *et al*, 2009). To explore the generality of these observations, we investigated the presence of population context-determined cell-to-cell variability of infection independent of RNAi for all 17 viruses in all cell lines tested (Figure 1B, and Supplementary Figure 5). We calculated the typical cell

Figure 1 Cell-to-cell variability in comparative RNAi screens of virus infection. **(A)** Experimental and computational overview of the 34 small-scale and 7 druggable genome screens analysed in this study. Individual screens are denoted as a combination of the virus and cell line abbreviation (for example, DEN-1_KY for DEN-1 infection in HeLa Kyoto cells). **(B)** Bootstrapped hierarchical clustering of the average population context-determined patterns of virus infection for each screen (Supplementary information). Infection efficiencies per quantile bin of the multidimensional population context and cellular state space were calculated, and z-score \log_2 transformed. Median values along all dimensions per screen were clustered (Supplementary information). **(C)** Heat map of virus infection-induced phenotypes (Supplementary information, and Supplementary Figure 4). Grey boxes indicate insufficient data points. **(D)** Heat map depicting average cell-to-cell variability infection pattern strength and reproducibility. The strength and reproducibility was calculated as the average of bootstrapped cell-to-cell variability pattern correlations observed between pooled populations of cells selected from different sets of perturbations per screen (Supplementary information). **(E)** Table of viruses with characteristic virus properties, virus names, families, and abbreviations (see for more information Supplementary Table 2). **(F)** A schematic outline of the cell population context and cellular state space model used throughout this study. Non-mitotic and non-apoptotic cells denote interphase cells. **(G)** Example scatter plots of the data used to calculate the pattern strength and reproducibility. SV40_CNX (green dots, correlation coefficient of 0.91) shows a very strong and reproducible pattern of cell-to-cell variability, whereas MHV_TDS (red dots, correlation coefficient of 0.05) shows no measurable pattern of infection. Individual dots represent infection indices for 2 bins of equal cellular state and population context and with at least 30 cells in each bin, but sampled from different sets of perturbations.

population context-determined infection pattern for each virus and cell line combination (Supplementary information), and clustered the infection patterns to assess similarities between them (Figure 1B and Supplementary information). Interestingly, of the 14 viruses screened in two or more cell lines, 5 viruses (VACV, YFV, DEN-1, HAdV3-C, and HSV1) displayed

different population context-determined infection patterns depending on the cell line strain. For 9 viruses, the infection pattern was similar among the sampled cell lines. We further controlled for virus infection-induced cellular state changes (see Figure 1C, and Supplementary information), and calculated the reproducibility and strength of the cell-to-cell



variability patterns of infection (see Figure 1D and G and Supplementary information). These results confirm and broaden the observation that the infection efficiency of an individual cell is strongly and reproducibly determined by its adaptation to the population context prior to the viral challenge.

RNAi perturbs virus infection via changes in the cell population context

RNAi in human tissue culture cells can strongly affect population context parameters, such as cell size and local cell density (Figure 2A–C). This can only partially be inferred from knowing the total number of cells in a population (~50%, see Supplementary Figure 14). For instance, two perturbed cell populations can have an identical number of cells, but strongly different distributions of local cell density or fractions of cells located on cell islet edges (see Figure 2D). As population context strongly determines virus infection, RNAi perturbations that alter the population context of individual cells may thereby indirectly change the infection efficiency. To evaluate this, we used the non-parametric models that take all quantified parameters of the population context into account to predict the expected unperturbed infection efficiency of each individual cell (see Figure 2D for DEN-1 virus infection and Supplementary Figure 16 for more prediction examples). The averages of these predicted single-cell values represent the expected unperturbed level of infection of a population of cells. We next compared these predictions with the actual measured fractions of infected cells in each cell population (see Figure 2E for VSV_MZ and see Supplementary information). Strikingly, the correlation coefficients between predicted and measured levels of infection in each population of an RNAi screen were as high as 0.89 (e.g. for VACV in HeLa MZ, Figure 2F). The average correlation coefficient for all 34 screens was 0.64 (Figure 2F). Controls confirmed that this was not the result of model over-fitting on potentially virus-induced phenotypic properties of individual cells (Supplementary Figure 10). This demonstrates that a large part of the changes in infection in these RNAi screens are determined by siRNA-induced changes in the cell population context, which propagate into virus infection in a predictable manner.

Single-cell modelling reveals causality in perturbation phenotypes

To separate perturbations that act via changes in the cell population context from direct perturbations of the infection process itself, we developed two complementary statistical analysis methods. To quantify the siRNA-induced change in virus infection independent of population context changes (direct effects), we normalized the measured level of infection for each siRNA-transfected cell population (total effects, including both direct effects and effects through the population context) with its model-predicted level of infection (see Figure 3 bar graphs on the right for examples, and Supplementary information). As a complementary analysis, we developed a data-driven approach based on Bayesian network inference to understand the causal interactions leading to the observed changes in virus infection (see Supplementary information, and Supplementary Figures 7 and 11). In total, we calculated 4998 bootstrapped Bayesian networks, one for each siRNA in each of the 34 small-scale screens. These networks (Figure 3 and Supplementary Figure 25) are inferred from single-cell measurements and display causal interactions between population context parameters and virus infection, as well as the effects of gene silencing on these parameters. This allowed us to infer how the loss of a gene lead to changes of parameters in this network, and which parameter was the most upstream effector. These analyses revealed direct, indirect and masked RNAi phenotypes (see Figure 3 and Supplementary Figure 12), which we observed both in the small-scale and in the large-scale screens of virus infection (Supplementary Figure 29).

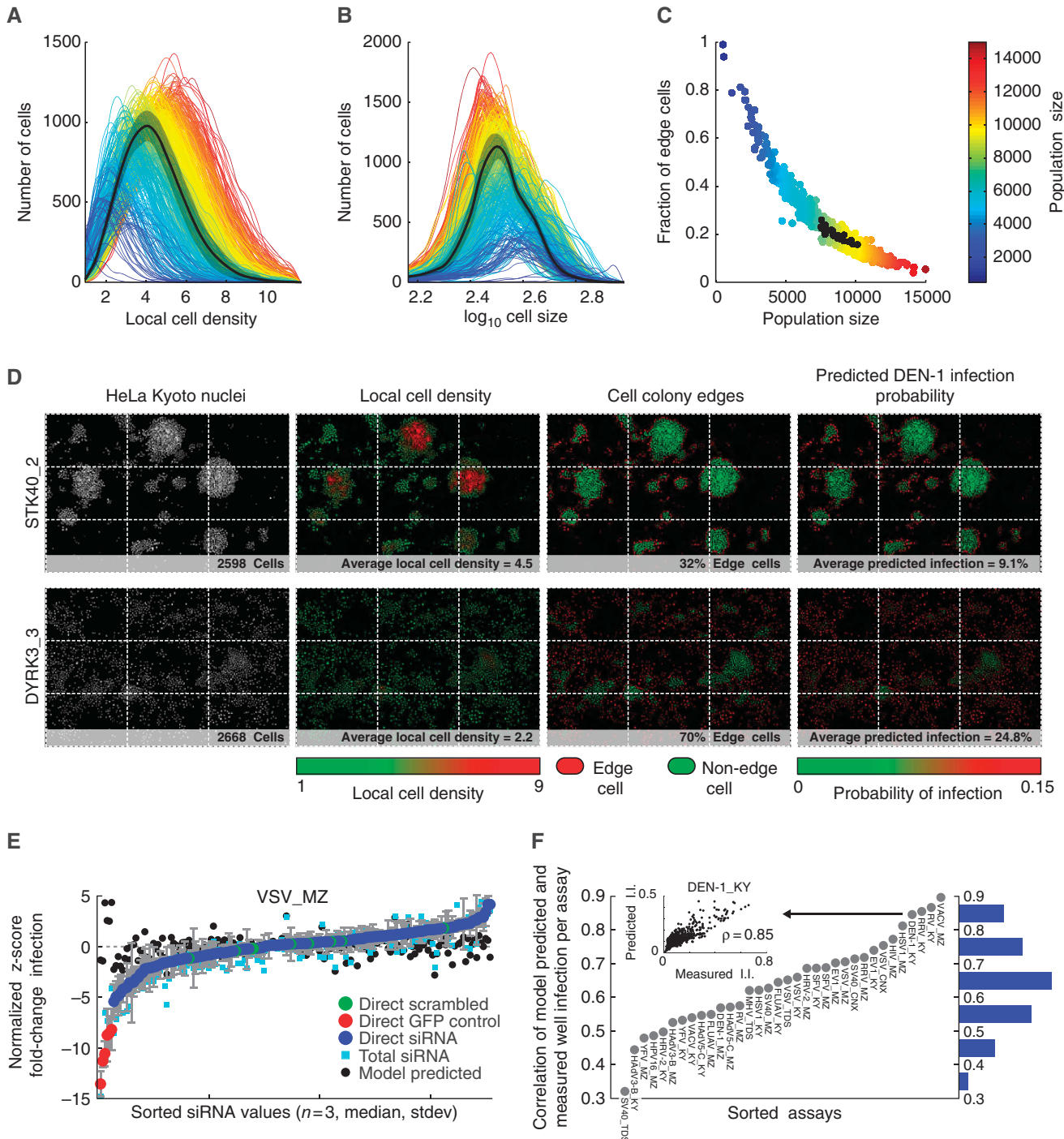
Direct RNAi phenotypes consist of perturbations that act directly on the infection process inside single cells, besides having possible effects on the population context. These changes in virus infection efficiency are traditionally the kind of perturbations one is after in RNAi screens, and cannot be predicted based on the population context. Examples of such direct effects of perturbations are the siRNA-mediated silencing of FK506-binding protein 12-*rapamycin*-associated protein 1 (FRAP1, also known as mechanistic target of rapamycin, mTOR) for the adenoviruses, and the second and third siRNAs silencing the dual-specificity tyrosine-(Y)-phosphorylation-regulated kinase 3 (DYRK3) for VSV (where the first siRNA did not have a direct effect on VSV) (see Figure 3A and

Figure 2 RNAi causes widespread changes in cell population context leading to predictable changes in virus infection. (A–C) siRNA-induced changes in the cell population context ((A) Local cell density distributions; (B) Cell size distributions; (C) Fraction of cells located in cell islet edges) are shown from an example small-scale RNAi screen. Individual coloured lines and dots represent values from individual populations of an example RNAi screen in HeLa MZ cells, colour coded according to the number of cells in each perturbed population (see colour scale on the right of panel C). Black lines (and transparent grey regions) and dots represent the median values (and s.d.) of non-targeting siRNA controls. (D) Two examples of wells with equal cell number (~2600 cells) but different cell population contexts. Segmented nuclei are pseudo-coloured according to the local cell density, cell islet edges, and predicted DEN-1 infection probabilities. The upper panels show HeLa Kyoto cells transfected with the second siRNA targeting STK40, resulting in an average local cell density of 4.5 (a.u.), 32% of cells located on cell islet edges, and a predicted infection index of 0.091. The lower panels show HeLa Kyoto cells transfected with the third siRNA targeting DYRK3, which has an average local cell density of 2.2 (a.u.), of cells located on cell islet edges, and a predicted infection index of 0.248. (E) RNAi results for VSV infection expressing GFP in HeLa MZ cells. Normalized z-scores of fold-change in infection (see Supplementary information) are presented per siRNA perturbation. Values are ranked according to the strength in direct perturbations of the infection process inside single cells. Blue circles, direct effects (i.e. population context-independent) for each siRNA; Green circles, direct effects of non-targeting (scrambled) siRNAs; Red circles, direct effects of positive controls (siRNAs silencing GFP); Light blue squares, total siRNA effects (direct and population context-mediated effects); Black circles, population context model-expected infection indices. Median values of triplicates are shown. Error bars depict standard deviations of direct siRNA effects ($n = 3$). (F) Correlation between all measured infection indices per siRNA and their model predicted infection indices for each small-scale screen (see Supplementary information and Supplementary Figure 13). Inset shows an example scatter plot of these values for DEN-1 infection in HeLa Kyoto cells (correlation = 0.85). The bars on the right axis show the relative histogram of all correlation values.

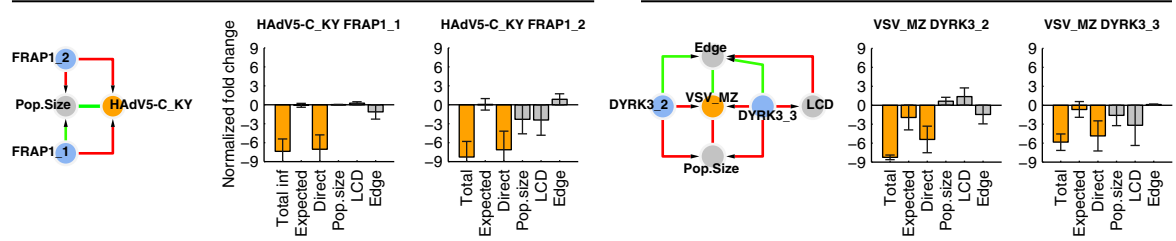
Supplementary Figure 12). These direct effects were confirmed by validation experiments (Supplementary Figure 32) as well as by the Bayesian approach, which inferred direct causal effects between the siRNAs and virus infection, despite the presence of additional siRNA-induced effects on cell growth and downstream cell population context parameters (Figure 3A).

Indirect phenotypes consist of changes in infection that can be fully attributed to changes in cell population context, and

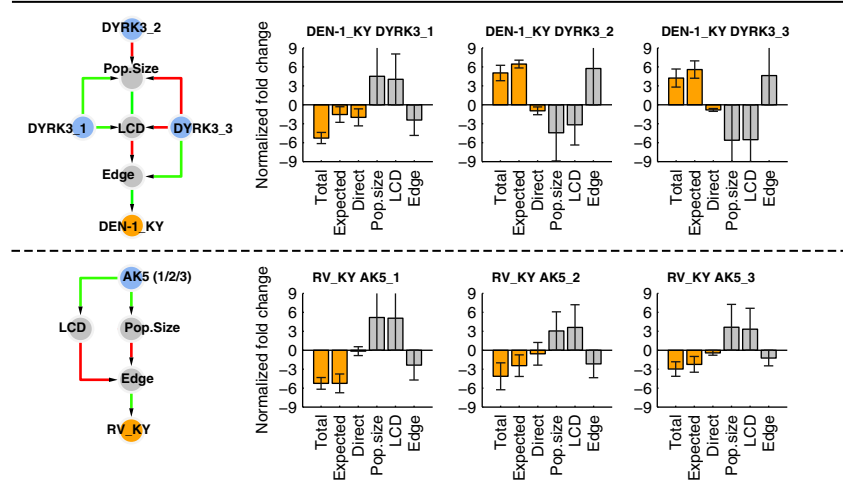
can therefore be predicted. An example of such a phenotype can be observed in the silencing of *DYRK3*, where the first siRNA (*DYRK3_1*) increased cellular growth, whereas the second and third siRNA (*DYRK3_2* and *DYRK3_3*) decreased cellular growth. This leads to a decrease in DEN-1 infection for the first siRNA, and an increase in infection for the second and third siRNA (Figure 3B). Causal modelling revealed that the first *DYRK3* siRNA increased population size and local cell density, resulting in a reduction of the number of cells located



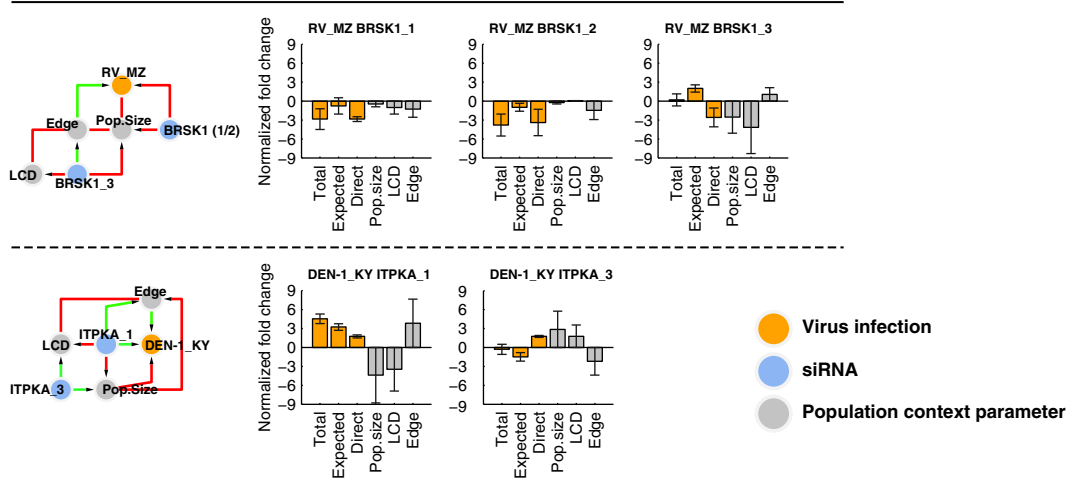
A Direct effects



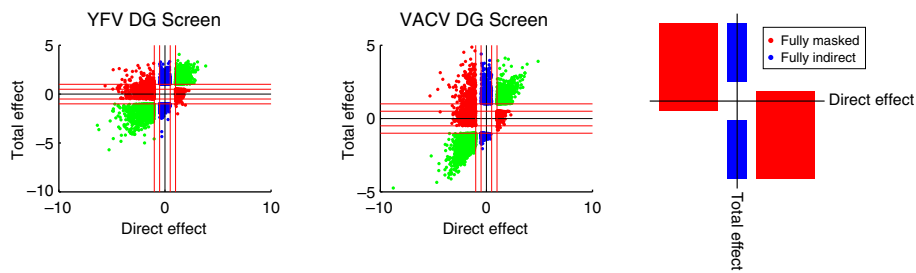
B Indirect (population context-mediated) effects



C Masked effects



D



on cell islet edges. The two other siRNAs showed decreased population sizes, leading to a lower local cell density and a relative increase in cells located on cell islet edges (Figure 3B). Because DEN-1 prefers to infect cells on the edges of cell islets, DEN-1 infection is significantly changed as a result of all three DYRK3-targeting siRNAs, but in opposite directions for the individual siRNAs. A second example is the decrease in RV infection upon silencing of adenylate kinase 5 (AK5) (Figure 3B). This decrease, observed for all three siRNAs targeting AK5 (AK5_1-3), is explained by an siRNA-induced increase in the number of cells and the local cell density, which resulted in fewer cells located on cell islet edges (Figure 3B). Because RV infection occurs preferentially in these cells, RV infection is strongly decreased.

We also observed masked phenotypes, where direct perturbations of infection were counteracted by opposite effects on infection through changes in the cell population context (Figure 3C). For instance, the direct effect of BR serine/threonine kinase 1 (BRSK1, also known as SAD1A) silencing on RV infection in HeLa MZ cells (a reduction) was masked in the total effect on infection of the third siRNA (BRSK1_3), due to a relative increase in the number of cells located on cell islet edges for this siRNA (Figure 3C). Also, the direct effect of inositol 1,4,5-trisphosphate 3-kinase A (ITPKA) silencing on DEN-1 infection in HeLa KY cells (a moderate increase) was masked in the total effect on infection of the third siRNA (ITPKA_3) by a decrease in the number of cells located on cell islet edges.

These examples illustrate the complexity that can underlie the effects of gene silencing over 3 days in a growing cell population. A cell population-averaged readout would overlook this complexity, leading in the case of indirect and masked effects to a wrong interpretation of the loss-of-function phenotype.

Accounting for population context strongly changes screening results and improves cell line comparability

We next asked to what extent the RNAi screening results changed when just the direct perturbation effects were considered instead of total perturbation effects. To account for inconsistent siRNA phenotypes, we calculated the median value of all three siRNAs, which effectively selects away any outlier siRNA phenotype. On average over all 34 small-scale screens, we observed a 40% change of the top 3 genes for which 2/3 or 3/3 siRNAs gave the same phenotype (Figure 4A

and Supplementary Figure 18). In the large-scale RNAi screens of virus infection, both masked and indirect effects occurred frequently. For three of the seven large-scale screens, we observed abundant (25–50%) fully indirect and fully masked siRNA phenotypes (Figure 3D and Supplementary Figure 29). Furthermore, over all seven screens, the identity of the 300 genes for which at least two out of three RNAi phenotypes most strongly and reproducibly reduced infection (host factors required for infection) changed on average by 35% after accounting for indirect and masked effects (Figure 4A). Thus, indirect and masked effects have a strong impact on the choice of genes for follow-up investigation. We further noticed that direct effects showed more significant outliers from the mean (a Kurtosis of 4.73 compared with 3.88 in the 34 small-scale screens) (Supplementary Figure 19), and a higher similarity between different negative controls ($P < 3 \times 10^{-9}$, $n = 306$) compared with total effects. Thus, by eliminating indirect effects and revealing masked effects, hits stand out more clearly from non-hits, improving the statistical window for hit scoring.

Given the strong changes in RNAi screening results after accounting for population context-propagated effects, we next asked how this would impact the reproducibility and comparability of RNAi screens. In general, comparative RNAi strengthens the confidence in reproduced screening hits and, for virus infection, has the potential to reveal structure in the complexity of the underlying host factors (Pelkmans *et al*, 2005; Pelkmans, 2009). However, published RNAi screens of the same virus infection have had poor hit-list overlap (between 3 to 6% overlap in hits from several HIV screens) (Bushman *et al*, 2009; Cherry, 2009; Mohr *et al*, 2010). Based on our findings so far, we reasoned that poor overlap in cell population-averaged RNAi screens could, in part, be caused by different patterns of cell-to-cell variability (see Figure 1B). To investigate this, we compared the two small-scale RNAi screens of each virus infection performed in multiple cell lines. VACV infection preferentially occurred in large HeLa KY cells with intermediate local cell densities, whereas in HeLa MZ cells, it occurred in sparsely growing and small cells (Figure 4B). The total effects of these two screens correlated poorly ($r = 0.12$, $P < 0.16$) (Figure 4C, upper panel), whereas the direct perturbation effects correlated significantly better ($r = 0.42$, $P < 10^{-6}$) (Figure 4C, lower panel). Similar significant improvements in comparability were observed for the HSV1 and DEN-1 RNAi screens performed in different cell line strains (Figure 4D). Like VACV, these virus infections displayed strong patterns of cell-to-cell variability, but with different characteristics for different HeLa strains (see Figure 1B). Notably, accounting for population context-propagated effects

Figure 3 Single-cell modelling of causality in population context-determined cell-to-cell variability of infection in RNAi screens. (A–C) Examples of the causal modelling (see Supplementary information) of direct (A), population context-mediated (B) and masked (C) perturbation effects, as well as their corresponding infection and population context measurements in various RNAi screens. Black arrowheads in the causal graphs denote inferred directionality of effects, whereas their absence indicates effects where both directions had the same likelihood. Red and green edges denote the sign of the pairwise single-cell bootstrapped correlations between nodes of the network (red = negative, green = positive). Orange nodes in the causal graphs represent virus infection; Light blue nodes represent siRNAs indicated by the gene name and their siRNA number (grouped by parentheses if effects for different siRNAs are identical). Grey nodes represent population context parameters (Pop. size, population size; Edge, cell islet edges; LCD, Local cell density). Population context and cellular state nodes not causally linked to infection have been omitted for clarity. The bar plots show normalized z-score values (Supplementary information) for infection (orange bars) and $3 \times$ z-score values (to put them on the same scale as infection) for population context measurements (grey bars). Error bars denote s.d. ($n = 3$) around median values. (D) Quantification of fully indirect (blue) and fully masked (red) siRNA effects for the YFV and VACV large-scale screens. siRNA phenotypes were classified based on their direct and total measurements, as indicated in the legend.

improved the poorest comparing small-scale RNAi screens, without adversely affecting the comparability of screens that displayed already good correlation ($P < 0.001$) of RNAi phenotypes between cell line strains before accounting for population context-propagated effects (Figure 4D). More importantly, when we compared hit-list overlap (and not overall correlation, which is strongly determined by non-hit data points) between small-scale screens of the same virus infection in different HeLa strains, we observed that hit-list overlap improved for 9 of the 15 virus infections, whereas that of four remained constant and two decreased (Figure 4D). Overlap between down-hits (host factors required for infection) increased from 26.7 to 40%, whereas overlap between up-hits (host factors suppressing infection) increased from 11.1 to 20% (Figure 4A and D).

In conclusion, hit lists from RNAi screens are confounded by indirect and masked effects due to changes in the cell population context. Correcting for these unwanted effects changes such lists, improves the statistical window for hit scoring, and increases reproducibility between screens performed in different cell line strains.

Cell population context confounds large-scale and genome-wide RNAi screens of various cell physiological activities

To investigate the generality of these improvements, we used the size of interphase nuclei, a proxy for cell size (Snijder *et al*, 2009), as a readout common in the seven druggable genome RNAi screens (Supplementary Table 6). Also for this readout, we observed significant changes in hits after accounting for population context-mediated effects. The top-300 hit lists for regulators of cell size (in which at least two out of three siRNAs gave consistent reductions of cell size) changed between 50 and 80% (Figure 4A). Furthermore, an average increase of 36% (from 11.8 to 16%) in overlap was observed between druggable genome screens performed in different HeLa strains (Figure 4A).

To investigate to which extent such improvements can still be seen when screens from different laboratories are

compared, using different experimental protocols, siRNA libraries and cell lines, we additionally analysed two genome-wide RNAi screens in HeLa MZ and A431 cells, performed in Lausanne (Scott *et al*, in preparation). First, accounting for population context-mediated effects in the identification of cell size regulators improved the hit-list overlap between these two screens with 35.7% (from 5.6 to 7.6% hit overlap, Figure 4A). We next analysed the impact between comparisons of the seven druggable genome screens performed in Zurich with the two genome-wide screens performed in Lausanne. Average hit-list overlap of all druggable genome screens and the genome-wide screen performed in the same cell line (HeLa MZ) improved by 23.9% (from 7.1 to 8.8%). The average overlap between all seven druggable genome screens performed in three different HeLa strains and the genome-wide screen performed in A431 cells furthermore improved by 13.1% (from 6.1 to 6.9%) (Figure 4A).

The two genome-wide RNAi screens from Lausanne were intended to identify regulators of cellular cholesterol levels in both HeLa MZ and A431 cells. Also for this important cell physiological property, we observed that accounting for indirect and masked effects through changes in the population context changed the top-500 hit list with 26.6% and 31.8%, respectively (Figure 4A). Furthermore, the overlap between these two hit lists improved by 20.8% (from 10.6 to 12.8% overlap). Similar observations were made in comparing two medium-scale RNAi screens (targeting 1183 genes using siRNA pools) of early endosome organelle abundance (EEA1 stain) performed in HeLa MZ cells and in A431 cells. The hit lists of these screens changed between 25 and 50%, whereas hit-list overlap increased by 17% (from 15.9 to 18.6%, Figure 4A).

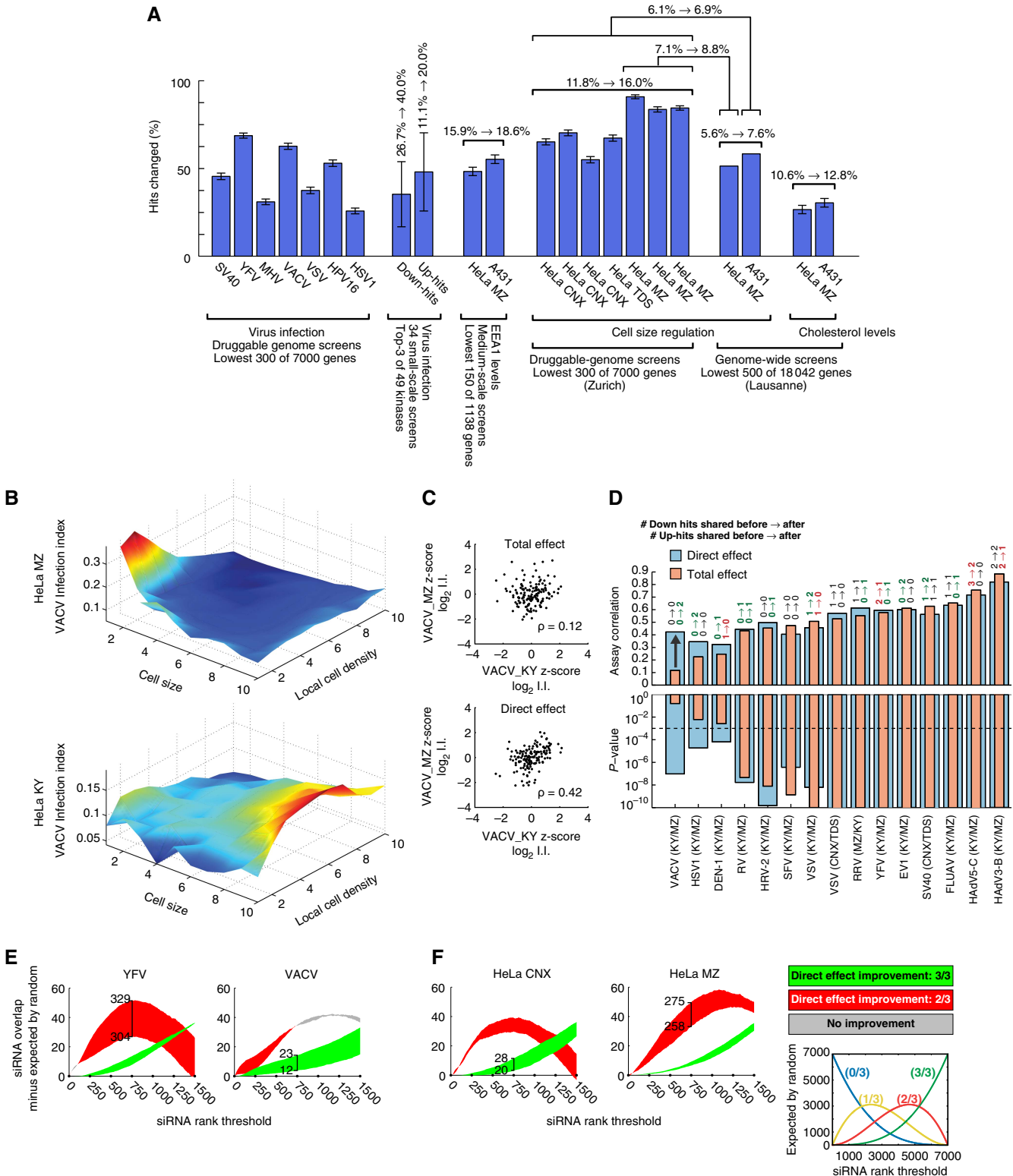
A previous study comparing genome-wide HIV infection screens revealed poor reproducibility in actual hits between these screens, but showed that the functional gene annotations of the hits overlap somewhat better (Bushman *et al*, 2009). As accounting for cell population context-mediated effects improves actual hit-list overlap of cell size regulators (Figure 4A), we also expected to see an improvement in gene

Figure 4 Accounting for population context-propagated effects changes hit lists and improves reproducibility, cell line comparability, and siRNA phenotype consistency. **(A)** Percentage of gene hits (observed with at least two independent siRNAs in triplicate) that changed after accounting for cell population context in: (left) the top-300 genes of the druggable genome screens, which silencing reduced virus infection; (second from left) the average of the top-3 genes in the 34 small-scale screens, which silencing reduced or increased infection; (third from left) the top-300 genes of the druggable genome screens, which silencing reduced interphase nuclei size; (right) the top-150 genes of the medium-scale RNAi screens on early endosome abundance (stained for EEA1). Brackets on top of the bar graphs indicate the improvement of hit-list overlap between screens of the same cellular activity performed in different cell lines and/or laboratories, with % overlap before and after the arrow representing results before and after accounting for population context-mediated effects. **(B)** Infection indices at intersections of the multidimensional population context and cellular state space (as in Figure 1F) for interphase cells, cells not located on cell islet edges, at intermediate population sizes. Examples for Vaccinia virus (VACV) infection in HeLa MZ cells (upper) and HeLa Kyoto cells (lower) are shown (see Supplementary Figure 20 for more examples). VACV infection preferentially occurs in different subpopulations of cells in the two cell line strains. **(C)** Scatter plots for measurements of total effects (upper) and direct effects (lower) of siRNAs on VACV infection in HeLa MZ and HeLa Kyoto cells. Black dots depict median values for individual siRNAs ($n = 3$). Values have been z-score \log_2 normalized over all siRNA phenotypes. Total siRNA effects do not significantly correlate between cell lines ($\rho = 0.12$, $P < 0.16$), whereas direct siRNA effects significantly correlate between the two RNAi screens ($\rho = 0.42$, $P < 10^{-6}$). **(D)** Comparison of siRNA effect correlations (upward bars) and their significance (downward bars) between virus infections performed in different cell lines (see also Supplementary Figure 17). Correlations between direct siRNA effects are indicated in blue; Correlations between total siRNA effects are depicted in orange. Dashed black line in lower bar graph indicates the $P < 0.001$ significance threshold. Additionally, absolute numbers of shared hits (up and down) are shown for total (before arrow) and direct (after arrow) effects. **(E, F)** Absolute increases of siRNA consistency (for calculation, see Supplementary information) are shown for the large-scale RNAi screens of YFV and VACV infection (E) and of interphase nuclei size in HeLa CNX and HeLa MZ cells (F). X-axes indicate the siRNA rank threshold at which the perturbation is considered significant. The number of genes for which 2/3 or 3/3 siRNAs give consistent phenotypes are shown, minus the number of corresponding cases expected by random (the random model is depicted on the right of F), and absolute numbers are indicated for certain examples. Numbers of genes with increased consistency for 2/3 or 3/3 siRNAs, after accounting for population context-propagated effects, are color coded as red and green, respectively (see legend). See Supplementary Figure 31 for the results of all seven large-scale infection screens.

annotation overlap. Indeed, overlap in the functional annotations enriched with a P -value below 0.01 and present in at least three of the seven druggable genome screens went up from 4 to 15. Such improvements were consistently found at varying

P -values and varying numbers of screens (Supplementary Figure 30, Supplementary Tables 5 and 6).

Taken together, our analysis of this unprecedented large number of RNAi screens shows that cell population context is a



confounding factor in RNAi screens of virus infection as well as of intrinsic cell physiological properties, such as cell size, cellular cholesterol levels, and organelle abundance. Accounting for this improves the reproducibility of such screens. This improvement is seen between screens performed at different laboratories, with different cell lines, different high-throughput cell culture and RNAi transfection protocols, as well as with different siRNA libraries (Supplementary Table 7).

Population context-propagated effects partially account for inconsistent siRNA phenotypes

We next investigated to what extent population context-mediated effects influence differences in phenotype observed for siRNAs targeting the same gene. Such different phenotypes can occur due to, for instance, dissimilar gene knockdown efficiencies and off-target effects caused by micro-RNA-like behaviour of siRNAs (Jackson *et al*, 2003; Luo and Chang, 2004), although the design of newer siRNA libraries attempts to avoid such micro-RNA-like effects. Interestingly, the direct siRNA effects on virus infection discussed above (see Figure 3) were more consistent among the individual siRNAs targeting the same gene than their total effects (see also Supplementary Figure 12). We therefore measured siRNA rank overlap (see Supplementary Methods) in the seven druggable genome RNAi screens of virus infection before and after accounting for population context-mediated effects (Figure 4E). Strikingly, for five of the seven screens, accounting for population context-propagated effects led to considerable improvements in siRNA consistency, whereas the other two screens were not negatively impacted (Supplementary Figure 31). The strongest relative increases were observed for genes where 3/3 siRNAs gave consistent phenotypes, with up to 80% more 3/3 phenotypes in the top 750 siRNAs of the three siRNA-sets in the VACV infection screen (Supplementary Figure 31). The strongest absolute increases were observed for 2/3 phenotypes, with up to 25 more 2/3 phenotypes observed for the top 750 siRNAs of three siRNA sets in the YFV infection screen (Figure 4E). In the large-scale RNAi screen on VACV infection, 11 more consistent 3/3 phenotypes were found (Figure 4E).

Finally, we also observed improved siRNA consistency for six of the seven druggable genome screens when analysed for cell size regulation (see Figure 4F). For instance, in HeLa CNX cells, 8 more 3/3 consistent siRNA phenotypes were found, whereas in the HeLa MZ cells, 17 more 2/3 consistent siRNA phenotypes were found. These improvements were independent of potential virus-induced cell size changes in these screens, as they remained when infected cells were excluded from the analysis. Thus, differences in population context contribute to inconsistencies between phenotypes of siRNAs that target the same gene.

Accounting for population context improves the functional annotation of large-scale RNAi screens

To understand the impact of scoring direct perturbation effects beyond the single-gene results of a large-scale RNAi screen, we compared the functional annotations (Dennis Jr *et al*, 2003) of SV40 infection hit lists before and after accounting

for population context-propagated effects (Figure 5, Supplementary Figure 28 and Supplementary Table 8). Comparing these two lists of annotations revealed that direct perturbation effects were more enriched in cellular processes known to be required for SV40 infectious entry (Pelkmans *et al*, 2001, 2005; Schelhaas *et al*, 2007), including membrane association, involvement in signalling, vesicle and calcium ion transport, association with the endoplasmic reticulum, and involvement in cell adhesion (Figure 5, Supplementary Figure 28). Accounting for population context-mediated effects, most strongly reduced annotations with expected pleiotropic roles in numerous aspects of cellular physiology, such as alternative splicing (Figure 5, and Supplementary Figure 28). Accounting for population context-propagated effects, therefore, not only improves the reproducibility and consistency of RNAi screening results at the siRNA, gene, and functional annotation level but also increases the focus of screening results on genes that are directly involved in the intracellular activity screened for.

RNAi perturbations of infection patterns reveal regulators of cell-to-cell variability

Apart from direct, indirect, and masked phenotypes, a global inspection of the causal networks derived from perturbed single-cell measurements (as shown in Figure 3) further revealed that certain perturbations invert the sign of the causal interactions by which population context determines virus infection (Supplementary Figure 24). These changes occurred irrespective of changes in cell population context or in the number of infected cells. Most notably, 9 of the 34 small-scale screens showed perturbations that resulted in a sign switch of causal links between cell islet edges and virus infection (Supplementary Figure 24). To further visualize and quantify these, we plotted bootstrapped single-cell correlations between infection and location on cell islet edges for each triplicate perturbation (Figure 6A). The screen of SFV infection in HeLa KY cells showed most opposing RNAi phenotypes, which either led to a positive or a negative correlation. The strongest increase of infection in edge cells was observed for the silencing of TRIO, whereas the strongest decrease was observed for the silencing of ABL1 (see Figure 6A and B, and Supplementary Figure 25). Population-averaged results did not reveal TRIO as a significant regulator in any of the RNAi screens, or ABL1 as a significant regulator of SFV infection. Because both protein expression levels and activity have been reported to correlate with cell population context parameters, such as cell size (Colman-Lerner *et al*, 2005; Newman *et al*, 2006) and cell density (Fukuhara *et al*, 2008; Snijder *et al*, 2009), we reasoned that this could also be the case for Trio and Abl1. Indeed, the single-cell protein levels of Trio and Abl1 showed population context-dependent expression levels, which were highest in those subpopulations where infection was most increased upon their gene silencing (Figure 6C). In addition, the silencing of TRIO and ABL1 increased clathrin-mediated endocytosis (CME, as measured by transferrin uptake, hijacked by SFV for infection (Helenius *et al*, 1980)) in a population context-dependent manner consistent with the upregulation of SFV infection (Figure 6D). Trio and Abl1 are

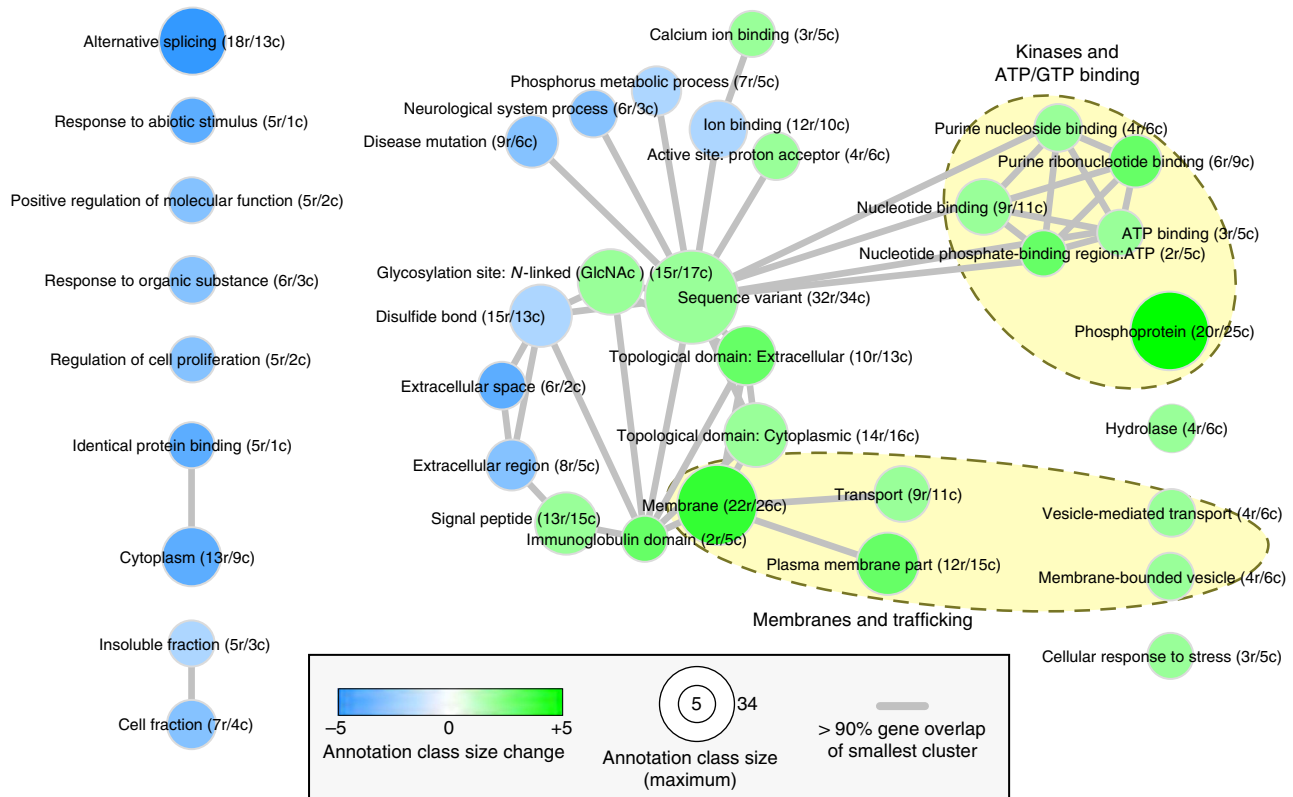


Figure 5 Accounting for cell population context leads to changes in the enrichments of functional annotation classes of hits in SV40 infection. Network view of functional annotation classes (based on DAVID (Dennis Jr *et al*, 2003) annotation clustering), which become depleted (blue) or enriched (green) after accounting for population context-propagated effects. Values between brackets behind annotation class names indicate the number of genes in these classes before (r) and after accounting for cell population context (c). Classes are linked in the network if they share more than 90% of the genes of the smaller class, for either total or direct effects. Two groups of classes, related to membrane trafficking and to kinase-mediated signalling, are highlighted with yellow-filled ellipses. See Supplementary Figure 28 for the complete overview.

direct and indirect activators of the Rho GTPase Rac1 (Scita *et al*, 1999; Sini *et al*, 2004; Zandy and Pendergast, 2008), a negative regulator of CME (Lamaze *et al*, 1996). Consistently, silencing of RAC1 increased SFV infection throughout the entire cell population (Supplementary Figure 26). Together, these results suggest a cell population context-dependent opposite regulation of Rac1 by Trio and Abl1, which has so far only been reported in several developmental studies (Liebl *et al*, 2000; Forsthoefel *et al*, 2005; Hurwitz *et al*, 2009). Multi-cellular organization with relevance for developmental biology might therefore be revealed by the analysis of perturbations of regulated cell population context-determined cell-to-cell variability in monoclonal tissue culture cells.

Towards a systems-level analysis of virus infection

Using bootstrapped hierarchical clustering, we next analysed the complex structure underlying the molecular regulation of virus infection by the 49 kinases tested in all 34 small-scale RNAi screens (Figure 7A) (see Supplementary information, and Matlab code available at www.infectome.org). Direct perturbation effects improved the co-clustering of the same virus infection screened in different cell line strains (see

Supplementary Figure 21). This indicates that clusters based on direct perturbation effects on virus infection are less sensitive to specific, virus infection-unrelated properties of the cell line used. Intriguingly, the obtained clustergram revealed a striking separation between enveloped and non-enveloped viruses ($P < 8.6 \times 10^{-9}$) (Figure 7A), suggesting that virus particle structure and type constrain the host cellular mechanisms that can be hijacked for infection.

To reveal the genes determining the clustergram structure, we calculated significant RNAi phenotypes for each branch point (see Supplementary Figure 22). This identified DYRK3 (independently validated for VSV, and HSV1 infection, see Supplementary Figure 32) and FRAP1 (independently validated for HAdV3-B and HAdV5-C infection, see Supplementary Figure 32) as the two most global determinants of this clustering (Figure 7B). FRAP1 and DYRK3 also displayed one of the strongest negative correlations between their RNAi profiles over all 34 small-scale assays ($r = -0.48$, $P < 0.0038$) (see Figure 7B and Supplementary Figure 23). More than half of all enveloped virus infections were more sensitive to DYRK3 silencing than to FRAP1 silencing, while most of the non-enveloped viruses were sensitive to FRAP1 silencing and not to DYRK3 silencing. FRAP1 is well known to regulate protein translation downstream of growth factor signalling and nutrient sensing (Sarbasov *et al*, 2005). DYRK3 has so far

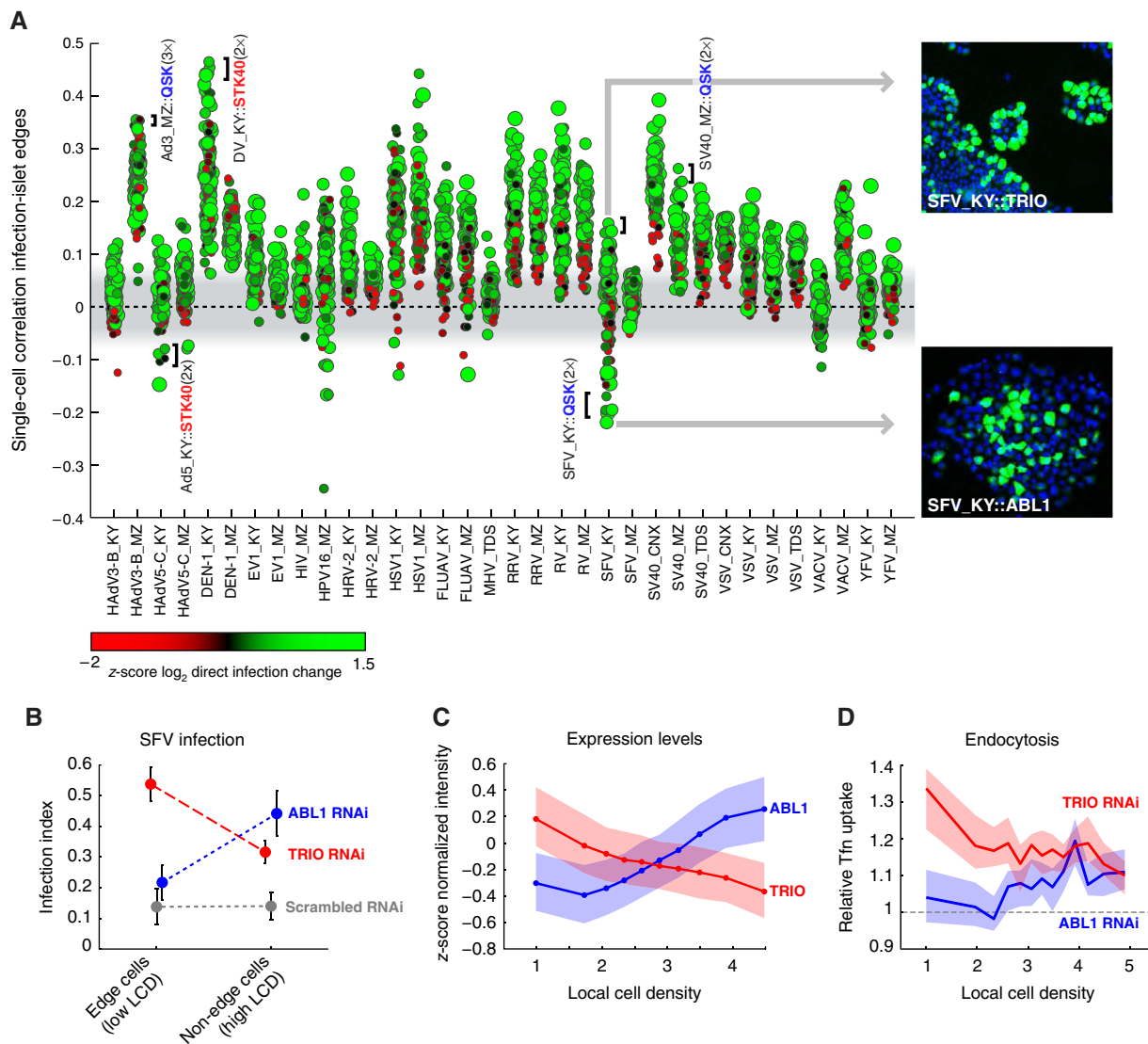
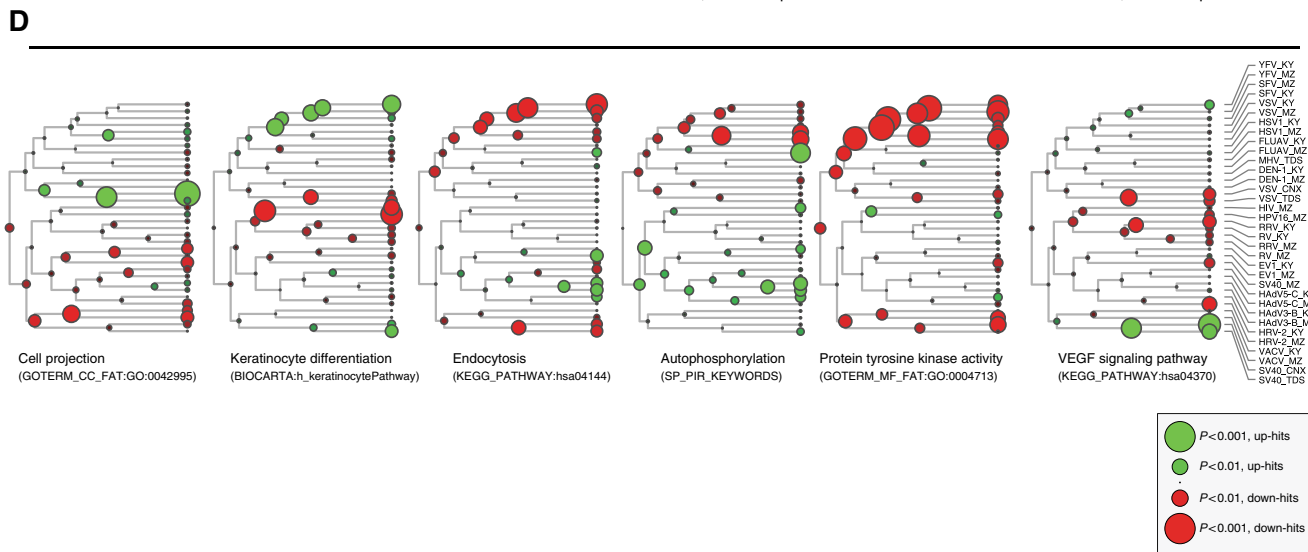
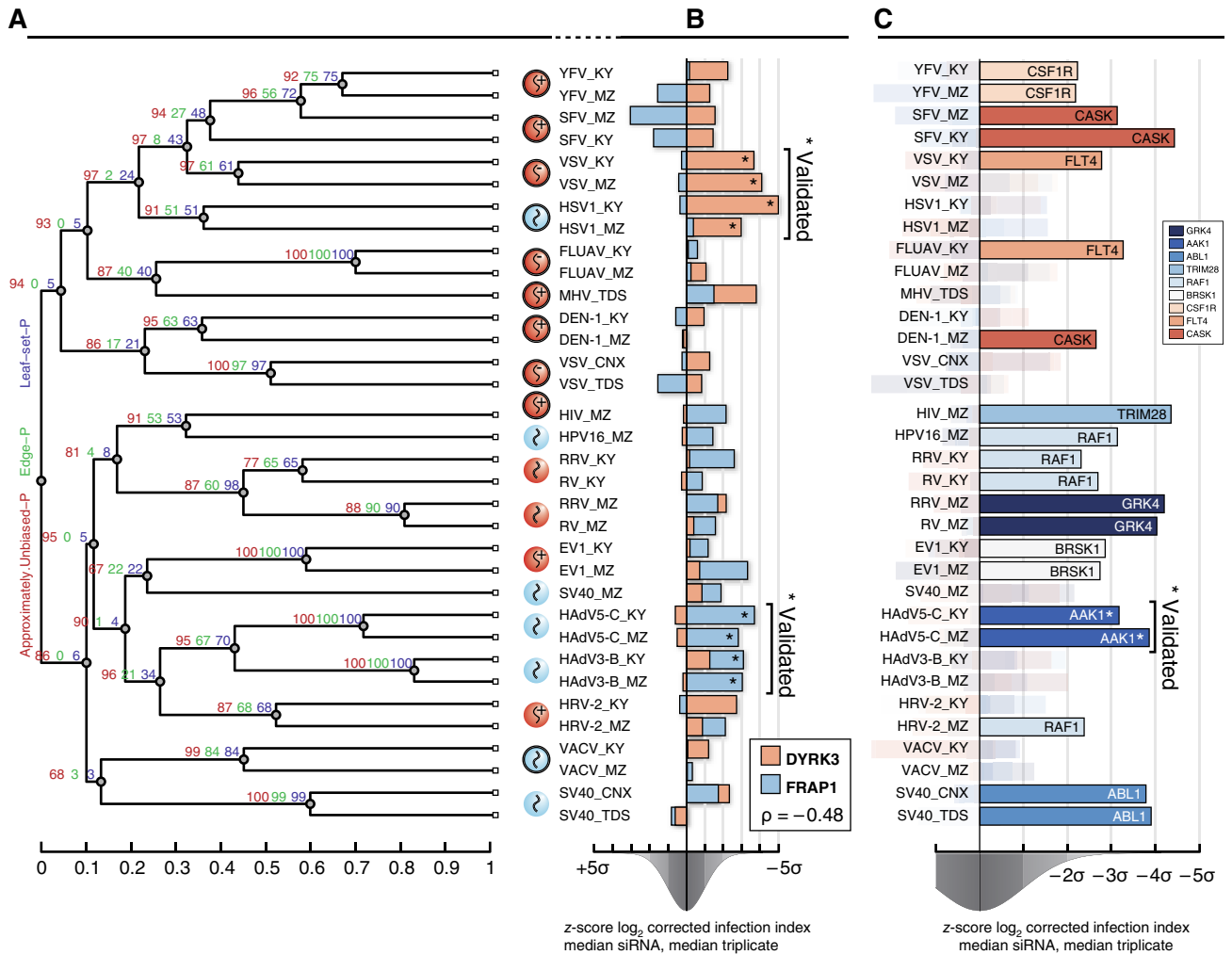


Figure 6 RNAi perturbations of population context-determined infection patterns. **(A)** Single-cell bootstrapped correlations between virus infection and location on cell islet edges are shown for the 147 siRNAs in each of the 34 small-scale screens. Bootstrapped correlations were calculated combining single-cell measurements from triplicate experiments. The grey region around zero indicates weak correlations. Dot colour and dot size depict the median direct effects of siRNAs on infection and the population size, respectively. Insets on the right show the observed patterns for SFV infection in HeLa Kyoto cells, in which TRIO (upper), or ABL1 (lower), is silenced (Blue = DAPI, green = SFV infection) (see also Supplementary Figure 25). Additional RNAi phenotypes of STK40 and QSK, which have multiple siRNAs (number in parentheses) targeting the same gene that cluster together in the dot plots (clustering indicated with black bold lines), are highlighted. **(B)** Measured indices of SFV infection in HeLa Kyoto cells located or not on cell islet edges. Median and s.d. of triplicate measurements are shown for TRIO-silenced cells (red), ABL1-silenced cells (blue) and scrambled (non-targeting) siRNA-silenced cells (grey). **(C)** Protein levels of Trio (red) are higher in sparsely growing cells, whereas protein levels of ABL1 (blue) are higher in cells in densely growing cells. Solid lines and dots indicate median values of mean intensity per cell, for varying local cell densities. Light-coloured regions indicate the inter-quartile range of intensities of individual cells. **(D)** Silencing of TRIO increases CME (as measured by transferrin uptake relative to that of non-perturbed cells) more strongly in single cells at lower local cell densities (red), whereas ABL1 silencing increases CME more strongly in single cells at high local cell densities. The solid lines represent median values, and transparent region $0.5 \times$ inter-quartile ranges of single-cell values.

Figure 7 Towards a systems-level analysis of host factors regulating infection across mammalian viruses. **(A)** Bootstrapped hierarchical clustering of all 147 direct siRNA effects of the 34 small-scale RNAi screens. Per branch-point, three bootstrapped empirical P -values were calculated: Red, approximately unbiased P -value; Green: sub-tree P -value; Blue: leaf-set (i.e. clade or cluster) P -value. The tree with the highest summed edge P -value from 10^6 bootstrapped trees is shown (see Supplementary information). **(B)** Direct effect of DYRK3 (orange) and FRAP1 (blue) silencing for each assay. Per screen, the median value of all three siRNAs, and median of triplicates is shown (i.e. similar to a two out of three criterion). Asterisks indicate gene-silencing phenotypes, which were independently validated (see Supplementary Figure 32). Distribution below the x -axis depicts typical data density of the normal distribution. Note that for visual purposes the x -axis is inverted, with negative values (reduced infection phenotypes) on the right side. The correlation of DYRK3/FRAP1 phenotypes over all assays is -0.48 ($P < 0.0038$). **(C)** Gene scores for selected genes are shown for all small-scale RNAi screens of virus infection. Bar colours correspond to different genes (see legend). **(D)** Functional annotation enrichment P -values are visualized for six functional annotation categories on the clustering of the 34 small-scale RNAi screens. Enrichment was calculated using a rank-based Kolmogorov-Smirnov method (see Supplementary information). Node colour indicates enrichment in down-hits (red) or up-hits (green), and node size indicates P -value of enrichment (see legend).

been poorly characterized. Interestingly however, the yeast homologues of FRAP1 and DYRK3 oppositely regulate ribosomal protein gene expression (Martin *et al*, 2004). This suggests the presence of two distinct and envelope-dependent

strategies by which viruses hijack regulation of protein translation. Analysis of RNAi phenotypes that determine the lower part of the clustergram structure further revealed virus-specific regulators of infection (validated for AAK1 in HAdV-



C5, see Supplementary Figure 32), cell line strain-independent regulators of infection, and additional regulators of multiple virus infection pathways (see Figure 7C, Supplementary Figure 22, Supplementary Table 4, and www.infectome.org).

We further analysed functional annotation enrichments at each branch-point of the clustering of the 34 small-scale screens, using a rank-based Kolmogorov–Smirnov enrichment score based on average gene ranks over all assays below each branch-point in the tree (Figure 7D, and Supplementary information). This revealed several global enrichment patterns that can only be identified by comparative RNAi. For instance, the clustered enveloped viruses YFV, SFV, and HSV1 share a stronger dependency on protein tyrosine kinases than the other viruses. We further observed several opposite roles for host factors with shared functional annotations. For instance, cell projection genes were required for VACV infection in HeLa KY and MZ but suppress VSV infection in HeLa CNX and TDS. Furthermore, VEGF signalling is required for VSV infection in both HeLa CNX and TDS, but suppresses SV40 infection in the same cell lines (see Figure 7D). Such global analyses are a first step towards a systems-level view of the complexity in host factors regulating infection across mammalian viruses.

Discussion

We have presented a systematic study of the interplay between population context-regulated cell-to-cell variability and RNAi phenotypes of virus infection in an unprecedented large set of functional RNAi screens. Two separate statistical methods consistently reveal abundant population context-mediated RNAi effects, as well as regulators of cell-to-cell variability. Accounting for even a limited set of population context measurements helps to address two of the largest technical problems currently faced in RNAi screening (Mohr *et al*, 2010). It leads to improved comparability of RNAi screens performed in different cell lines, and to improved similarity of perturbation phenotype obtained with different siRNAs that target the same gene. Besides virus infection, improvements were also observed for screens investigating cell size, cellular cholesterol levels, and organelle abundance, even when comparing screens performed in different cell lines, screening facilities, and with different siRNA libraries. Discrepancies in screening results caused by true differences in the molecular mechanisms by which cell lines regulate their physiology will naturally remain.

The improved consistency in phenotype of siRNAs targeting the same gene stem in part from different effects on cellular growth, which then propagate via the cell population context into the cellular activity screened for. It will be interesting to understand to which extent these differences are the result of different silencing efficiencies or cytotoxicity due to true off-target effects (Jackson *et al*, 2003; Luo and Chang, 2004). However, regardless of the nature of these differences, our approach partially resolves the problem.

Given the broad range of improvements observed in this study, we expect our methods to also improve other RNAi screens of cellular processes that show cell-to-cell variability determined by the cell population context. For virus infection, which has been a common focus of RNAi screening in the past 5 years (Bushman *et al*, 2009; Cherry, 2009; Mohr *et al*, 2010),

as well as cell size regulation, lipid homeostasis, and endocytosis, second-generation RNAi screens that account for cell-to-cell variability will likely result in significantly altered and improved hit lists. If population context-mediated effects should generally be accounted for is currently not clear, but we expect the cell population context to influence the expression of a large number of genes (see for instance (Fukuhara *et al*, 2008)), metabolic activity, signal transduction, cell polarization, cell proliferation and migration, and the determination of cell shape (Kim *et al*, 2010; Schultz *et al*, 2011; Snijder and Pelkmans, 2011).

Besides RNAi screening, our findings in principle apply to any type of cellular perturbation, although at different levels. When both the perturbation and the assay occur in a time period that is shorter than one cell cycle, for instance in small compound screens, and cytotoxicity is not too high, there may be little effect on the cell population context, and therefore likely less indirect effects. However, one can still expect molecular perturbations to show effects only in a specific subpopulation of cells that reside in a particular microenvironment (for instance, cells on cell islet edges or cells with extensive cell–cell contacts), or to even show different effects in different subpopulations of cells (Singh *et al*, 2010). The fact that Trio and Abl1 show population context-specific expression levels suggests that small-molecule inhibitors of such kinases would also show context-specific effects. Population-averaged readouts may thus also in such screens lead to an incomplete or wrong interpretation of results.

Our findings show how cell population averaging can be misleading in interpreting a perturbation phenotype. This can only be overcome by applying methods that have the depth to measure activities at single-cell resolution, in a large number of single cells, and that can quantify each single cell's population context. At the moment, these requirements are only met in image-based approaches combined with single-cell image analysis and data-driven modelling. Perturbation screens that measure average signals from a whole perturbed population, for instance via fluorescence plate readers, cannot apply such a deep screening approach and will therefore lack the technical advances it offers. However, with the advent of affordable automated microscopes, and the continued development of software and computational capacity, single-cell image-based screens at a depth performed here may now become more mainstream. To facilitate the adoption of the methods presented here, a MatLab-based toolbox that integrates with CellProfiler (Carpenter *et al*, 2006) is available for download from our website (<http://www.infectome.org>).

Single-cell analysis of the population context allows a more sensitive and complete mining of the potential resource that image-based RNAi screens are for systems-level approaches. Highly parallel RNAi screens of virus infection can now reveal the combinatorial molecular modules of the host cell that viruses hijack for different steps of entry and replication (Damm and Pelkmans, 2006; Li *et al*, 2009), reducing the bias of cell line specificities. Furthermore, the ability to identify perturbations of population context-determined cell-to-cell variability, which are fundamentally different from perturbations of overall cellular activity, enables the systematic investigation of how single cells adapt to their microenvironment and shape the cell population context.

Materials and methods

Experimental material and methods

Cell lines and tissue culture

All human cell lines were maintained under standard tissue culture conditions, at 37 °C in a humidified atmosphere of 95% air and 5% carbon dioxide. HeLa Kyoto cells were obtained from Jan Ellenberg (EMBL, Heidelberg), HeLa MZ cells from Marino Zerial (MPI-CBG, Dresden), HeLa CNX from Cenix Biosciences GmbH (Dresden), and HeLa TDS from the High-Throughput Technology Development Studio (MPI-CBG, Dresden).

siRNA transfection

RNAi screens were performed in black 96-well (small-scale screens) or 384-well (medium, druggable genome, and genome-wide screens) μ Clear plates from Greiner. RNA interference against the selected 49 kinases (Supplementary Tables 1 and 3) (small-scale screens) and against the 6986 genes of the druggable genome (siRNA #1 and #2 from Qiagen druggable genome V2, and siRNA #3 from Qiagen druggable genome V3) was performed with three independent siRNAs from Qiagen each, and each siRNA transfection was repeated three times. siRNA transfection was performed in a reverse manner, seeding the cells into a well already containing transfection mix. The transfection mix for a single well of a 96-well contained 0.1 μ l LP2000 (Qiagen), 14.9 μ l Optimem and 5 μ l of 1 μ M siRNA (diluted in RNase-free ddH₂O). The cells were seeded in 80 μ l/well of complete medium (DMEM, Gibco#41965039; 10% FCS, Sigma #F7524 lot.# 085K3396; 1% Glutamax, Gibco#35050038). The number of seeded cells was 3200 for HeLa MZ, and 2500 for HeLa Kyoto, HeLa TDS, and HeLa CNX, except for the YF and DEN-1 screens, where 2700 and 2200 cells were seeded for HeLa MZ and HeLa KY, respectively. Reagent and cell amounts for 384-well plate-based experiments were scaled according to the volume and surface area difference.

Virus preparation

Virus preparations for the 17 different mammalian viruses (see Supplementary Table 2) were performed as described in the Supplementary Methods and references therein.

Infection assays

We aimed for a mean infection index (fraction/percentage of infected cells in a well) of 10 to 20% for all infection assays. As the infection index is a combined function of the amount of virus added, and of properties of the host cells (Snijder *et al*, 2009), we tested the various virus stocks by seeding a gradient of cells (HeLa MZ: 1000 – 22 000, HeLa Kyoto: 500 – 20 000) in the columns of a 96-well plate and subsequently infecting the rows with a gradient of virus stock (see Supplementary Methods). Optimal virus amounts were selected for each corresponding cell line that resulted in an infection index of ~10% over a wide range of different population sizes and contexts. Most infection assays were typically performed 48 h after seeding and reverse siRNA transfection of cells by adding an infection volume of 10 μ l (HadV3-B, HadV5-C), 20 μ l (RV,RRV), 30 μ l (HRV-2) or 50 μ l (FLUAV, VSV). For SFV (PBS wash), SV40 (R-medium wash) and VACV, the cell culture medium was removed and replaced by the infection volume, 20 μ l (SV40), 50 μ l (VACV) or 90 μ l (SFV). See the Supplementary Methods and Supplementary Table 2 for more details on the individual infection assays.

Detection of virus infection

Successful virus infection was detected either by immunofluorescence staining of a viral protein or by detection of a fluorescent protein integrated into the viral genome. Immunofluorescence was performed for and with the following virus and antibodies: DEN-1 (antibody: monoclonal antibody 2H2 (ATCC collection), 8 μ g/ml, the antibody was labelled with Alexafluor 488, Invitrogen A10235; epitope: prM), EV1 (antibody: rabbit Ab against purified EV1; epitope: capsid proteins VP1–3), RV (antibody: OBT0882, Oxford Biotechnology, 1:500; epitope: VP6), RRV (antibody: OBT0882, Oxford Biotechnology,

1:500; epitope: VP6), HRV-2 (monoclonal antibody 8F5 (from D Blaas, Department of Medical Biochemistry, Max F Perutz Laboratories, Vienna Biocenter, Medical University of Vienna, Austria); epitope: VP219), FLUAV (antibody: anti influenza A nuclear protein mouse hybridoma strain, ATCC HB-65, unpurified supernatant 100 μ l per well; epitope: nuclear protein), SFV (antibody: Lady Di, 1:200; epitope: E1, E2), SV40 (epitope: T-antigen1) and YFV (antibody: monoclonal antibody 2D12 (ATCC collection), 2 μ g/ml, the antibody was labelled with Alexafluor 488, Invitrogen A10235; epitope: E protein). The infection of HSV-1, HadV3-B, HadV5-C, HIV-1, HPV16, MHV, VACV, and VSV was detected by expression of fluorescent reporters integrated in the viral genomes. See Supplementary Table 2 for more information on the specific virus strains used.

Fixation and staining

All assays were fixed by adding PFA to an end concentration of 4%, and incubating the cells for 15 – 30 min, followed by two PBS washes. To assure a homogeneous DAPI stain, cells were permeabilized for 3 min with PBS containing 0.1% Triton X-100. After one PBS wash, the cells were incubated with PBS containing 1 μ g/ml DAPI for 10 min. Before scanning a final wash with ddH₂O was performed, which was found to minimize contaminants when imaging samples with automated microscope. For those assays where virus infection was assessed by indirect immunofluorescence, the cells were incubated between 2 h and overnight with the specific primary antibody (see paragraph above for details), secondary antibody incubation, and by two PBS washes before DAPI staining and imaging.

Imaging

All small-scale and druggable genome screen images were acquired on two automated wide-field cell WoRx™ microscopes (Applied Precision) with a 10 \times objective. Per well, 3 \times 3 directly adjacent images were acquired, covering ~65% of each 96-well and over 90% of each 384-well bottom surface. Image-based auto-focussing was performed on the DAPI signal for each imaged site. The images were recorded with 12-bit CCD cameras, and stored as individual 16-bit uncompressed TIFF files per imaged site and per channel. Combined for all 34 small-assays, 396.018 images (994 GB of image data) were analysed, whereas the seven druggable genome screens amounted to over 5 million individual images (amounting to roughly 5 TB of raw image data).

Transferrin (Tf) uptake assay

HeLa KY cells were grown and siRNA-transfected as described above. The cells were incubated with Alexa488-Tf for 15 min at 37 °C in the incubator. The cells were then washed once with PBS at RT and then fixed for 15 min with PFA 4%. After fixation, the cells were permeabilized with 0.1% Triton for 5 min at RT and the nuclei stained with DAPI (1 μ g/ml) in PBS for 10 min at RT. The cytosol was stained incubating the cells with CellTrace (Molecular Probes, Invitrogen, 1:10 000 from stock solution 1 μ g/ μ l in DMSO) in 100 mM Na₂HCO₃ for 5 min at RT to remove non-internalized Tf. The cells were then finally washed twice with PBS. The image acquisition was performed with a Molecular Device microscope (ImageXpress Micro) with a 20 \times objective. Different z-planes were acquired for the Tf channel (only the maximum intensity projection was saved) and 49 fields per well were acquired.

ABL1 and TRIO immuno-staining

Cells were incubated with blocking solution (1% BSA and 50 mM NH₄Cl) for 45 min at RT, washed with PBS, incubated with specific primary antibody (1:100 ABL-Cell Signaling, rabbit, Cat.num.2862; 1:100 TRIO- Santa Cruz, goat, C-20 and N-20,sc-6061 and sc-6060, respectively) in blocking solution for 1 h and subsequently incubated with the required secondary antibody (Molecular Probes, Invitrogen). Image acquisition was performed with a Molecular Device microscope (ImageXpress Micro) with a 20 \times objective, and 49 fields per well were acquired.

SFV follow-up infection experiments

HeLa Kyoto cells were reversely transfected as previously described with 50 nM siRNA and Lipo2000 (Invitrogen) as a carrier in 96-well plates (Greiner). At 3 days (72 h) after transfection, the cells were washed twice with serum-free RPMI medium, containing 0.2% BSA and 20 mM HEPES (pH 6.8). After 15 min incubation at 37 °C in the incubator, the medium was removed and the virus was added with a 1:900 dilution prepared in RPMI-BSA, pH 6.8. The cells were incubated with the virus for 60 min at 37 °C in the incubator. The medium was then replaced with RPMI-BSA medium without virus (pH 7.2) and the cells were further incubated for 2 or 4 h at 37 °C in the incubator. The cells were washed once with PBS at RT and then fixed for 15 min with PFA 4%. After fixation, the cells were permeabilized with 0.1% Triton for 5 min at RT and the nuclei stained with DAPI (1 µg/ml) in PBS for 10 min at RT. The cells were next incubated with blocking solution (1% BSA and 50 mM NH₄Cl) for 45 min at RT, washed with PBS, incubated with specific primary antibody (Mouse monoclonal anti dsRNA (J2)-use at dilution 1:1000; Rabbit polyclonal antibody against E1/E2 glycoproteins (Lady-Dye)-use at dilution 1:1000) in blocking solution for 1 h at 4 °C and finally incubated with the appropriate secondary antibody (Molecular Probes, Invitrogen). Image acquisition was performed with a Molecular Device microscope (ImageXpress Micro) with a 10 × objective.

DYRK3, FRAP1, and AAK1 follow-up experiments

Validation experiments were conducted with siRNAs different from the original screens, but in otherwise identical setups (Supplementary Figure 32). Adenovirus experiments were analysed with a Tecan Safire plate reader and normalized for the total amount of DAPI signal, whereas all other experiments were analysed the same as the original screens, and normalized to the respective controls. Cells were incubated with small compounds (1 µM GSK626616, a DYRK3-specific inhibitor, or 1 µM Harmine, a general DYRK family kinase inhibitor) for 12 h prior to virus infection. Cell number effects were measured (~55% and ~71% of cells remaining after GSK626616 and Harmine treatment, respectively) and did not invalidate the interpretation of the direct effect of the small compounds. Validation of AAK1 for HAdV5-C infection was done by transient transfection of a human AAK1-construct combined with the expression of a validated Sigma MISSION shRNA (5'-CCGGCTCACTGTATCACTCCAAATCTCGAGATTTGAGTGATACAAGTGAAGTTTTTGG-3') targeting the 3'-UTR of AAK1. Qiagen DYRK3_2 and DYRK3_3 knockdown efficiency was measured using qRT-PCR to be, respectively, 83% and 99% efficient.

Computational image analysis

The images were analysed with open-source image analysis software (Carpenter *et al*, 2006) (<http://www.cellprofiler.org>), combined with software published previously for supervised classification of cellular phenotypes (Ramo *et al*, 2009), and image analysis methods specifically designed for this study and implemented in MATLAB (<http://www.mathworks.com>) (see below). All computational image analysis was run on the high-performance computing cluster Brutus from the ETH Zurich. Novel methods can be downloaded from <http://www.infectome.org>.

CellProfiler Image analysis

The CellProfiler image analysis pipeline was as follows: First, nuclei were identified based on the DAPI stain. Next, cell boundaries were estimated using nuclear expansion. Standard CellProfiler texture, intensity, size, and shape features were extracted from nucleus and cell regions, as well as from complete images, for all available channels. In total, 224 raw and derived features were measured per individual cell.

Supervised classification of cellular phenotypes

We applied supervised machine learning using the open-source support vector machine (SVM) learning tool CellClassifier (Ramo *et al*, 2009) (<http://www.cellclassifier.ethz.ch>), to identify biologically and technically relevant cellular phenotypes. We classified all cells in

each screen into five distinct binary phenotype classes (infected/non-infected, interphase/non-interphase, mitotic/non-mitotic, apoptotic/non-apoptotic, blob/non-blob). See the Supplementary Methods for more information.

Population context measurements

The population context parameters (local cell density, population size, cells residing at cell islet edges, cell size, distance from cell-colony edge) were measured using custom software written in Matlab specifically for this study. The algorithm works on top of CellProfiler measurements of the weighted centroids of individual objects (nuclei), and can be downloaded from <http://www.infectome.org> as CellProfiler module (MeasurePopulationContext). See the Supplementary Methods for more information.

Image analysis automation via iBRAIN

Image analysis and data content management was largely automated using iBRAIN (image-based RNA interference) (see <http://www.infectome.org>), a custom-made software package that manages job submission, job tracking, result visualization and error recovery on the ETH high-performance computing cluster (Brutus), of the full analysis pipeline. The pipeline includes, amongst others steps, lossless image compression, merging, stitching and rescaling of images into preview JPG files, CellProfiler image analysis, population context measurements and correction, SVM classifications, additional statistical analyses, as well as preliminary data visualization and basic lab information management.

Statistical analysis

Quantile multidimensional bin (QMB) models

To model virus infection in a robust and non-parametric way, we performed quantile multidimensional binning of cells (for the small-scale screens on average ~5.8 million cells per assay, after discarding poor quality cells) for the following population context and cell state parameters: Local cell density (LCD), population size (POP.SIZE), cell islet edges (EDGE), cell size (SIZE), mitotic phase (MIT) and apoptosis (APOP), where MIT and APOP are mutually exclusive, and their combined alternative state (i.e. non-mitotic and non-apoptotic) denote interphase cells (See Figure 1F, and Supplementary Figure 3). Each parameter was quantile binned, and cells with equal bin-values for all six dimensions were grouped into multidimensional bins (Supplementary Figure 3), and average readout values (infection levels, or log₁₀ transformed intensities for non-virus readouts) were calculated over all cells from each multidimensional bin. See Supplementary Methods for more details. The QMB method presented here is fast, general, accommodates mixes of discrete and continuous parameters, and has been successfully applied to many other responder and predictor parameter combinations. A Matlab implementation for bootstrapped QMB modelling and subsequent correction using these models is implemented in a CellProfiler v1 module (PopulationContextCorrect), and can be downloaded (with an example data set) from <http://www.infectome.org>.

Clustering QMB models (heterogeneity patterns)

The clustering results presented in Figure 1B were obtained from QMB models including all cells from a single infection assay. For clustering and visualization purposes, we transformed each six-dimensional QMB model per assay into a row vector by taking the median value of all bins at a given value along all six dimensions (for instance, median value of all bins with an LCD = 1, 2, ..., 11, SIZE = 1, 2, ..., 11, etc. for all six dimensions). We find that the clustering of these non-parametric representations of the data yields better (more robust) results compared with a clustering of parameter values estimated with various regression approaches, likely due to the non-linear nature of the data and inherent differences in LCD and SIZE distributions of the different cell lines used in this study. The data were clustered using a novel bootstrapped hierarchical clustering implementation in Matlab made specifically for this study (available for download from

<http://www.infectome.org> and see Supplementary Methods), using a correlation distance measure and average cluster method. Prior to clustering, infection indices were transformed in \log_2 relative infection indices, relative against the median bin infection index per QMB.

Correction of population context effects using QMB models

To correct for population context and cell state-specific effects, we first created QMB models (as described above) for all cells from a given plate (as opposed to all cells from a given assay). The QMB models describe the average, expected, infection index for cells from a given population context and cellular state, per plate. Next, expected infection indices per well were calculated as the sum of QMB expected infection indices for all cells from a given well, given those cells' population context and cellular state parameters and the plate-wise QMB models. The expected infected index equals the mean QMB-derived infection probabilities for all cells in a given well, equal to the summed probability of infection divided by the total number of cells. Finally, the corrected score is the difference between observed readout and expected readout.

Bootstrapped Bayesian network inference

Bayesian network learning was performed essentially as described before (Snijder et al, 2009), and see Supplementary Methods. Causal networks were inferred from single-cell measurements from cells treated with the same siRNA as well as from mixed cells to estimate the siRNA effects. Single-cell measurements in each bootstrap sample were discretized maximizing the Akaike Information Criterion, as implemented in the BNT Structure Learning Package. Bayesian network learning was performed using a greedy search algorithm in the Markov equivalent space (Chickering, 2003). Retrieval rates for each edge were calculated as the fraction of bootstrap runs in which that particular edge was found. See Supplementary Methods for further details.

Rank-based siRNA consistency scores

To measure the consistency between phenotypes of siRNAs we calculated a score that tests the significance of siRNA consistency beyond that expected by random data (i.e. hypergeometric distribution function). We set the top- x strongest siRNAs of each of the three siRNA sets to 1 (hit) and all other siRNAs of each set to 0 (non-hit). We next calculated for increasing x how many genes had 0/3, 1/3, 2/3, and 3/3 siRNA hits. We bootstrapped this procedure for both direct and total effects 130,000 times, by adding an amount of normally distributed noise (with a sigma of $0.3 \times$ s.d. of each siRNA set) to the data set. Average values over 100,000 bootstraps are shown in the main text. If the lists would be completely random, the probabilities for each gene-case can be calculated using the hypergeometric distribution function:

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

where M is the number of genes (~ 7000), K is the number of hit-siRNAs in each siRNA set (0 to ~ 7000), N is the number of siRNA sets (3), and x the number of positives we sample (0 to 3).

Functional annotation enrichment scores on tree

We tested the functional annotation enrichment by a non-parametric two-sample Kolmogorov-Smirnov test, comparing the ranks of genes of a functional annotation (positive genes) to that of all other genes. To test functional annotation enrichment of branch points in the tree (i.e. multiple assays), we calculated the two-sample Kolmogorov-Smirnov test for the gene ranks of all ranks averaged over the underlying assays.

Additional statistical analyses

See Supplementary Methods for more information on the calculation of virus-induced cellular state controls, calculation of the

heterogeneity pattern strength and reproducibility, and an explanation and motivation of the normalized z-score.

Data availability

Screening data is available at <http://www.infectome.org> and in the Supplementary information.

Supplementary information

Supplementary information is available at the *Molecular Systems Biology* website (www.nature.com/msb).

Acknowledgements

We acknowledge T Steiger and O Byrde from the ETH IT services for large-scale data storage and high-performance cluster computing, as well as G Csucs and A Vonderheit of the RNAi and image-based screening center of the ETH Zurich, and Frank Wippich for experimental help. PR is supported by the Human Frontiers Science Program, PL by the Federation of European Biochemical Societies. JG is supported by the Swiss National Science Foundation, the SystemsX.ch RTD LipidX and the NCCR in Chemical Biology. CCS has been supported by LipidX, the CIHR and the HFSP. LP is supported by the University of Zurich, the Swiss National Science Foundation and the European Union. This work was supported by the SystemsX.ch RTD projects PhosphoNetX, LipidX, and InfectX.

Author contributions: LP supervised and conceived the project. RS, KM, BSn, NW, LB, MHV, JM, SM, The, KT, AJ, PL, GB, and MS performed experiments, BSn and PR developed computational image analysis methods, BSn performed all computational image analysis, BSn and PR conceived the statistical analysis methods. CCS and JG performed and analysed the genome-wide cholesterol RNAi screens. BSn and PR performed all statistical analysis, CAMH, VM, THy, PJ MR, BSo, MM, AA, UG, and AH discussed results and provided reagents, LP and BSn wrote the manuscript.

Conflict of interest

The authors declare that they have no conflict of interest.

References

- Bushman FD, Malani N, Fernandes J, D'Orso I, Cagney G, Diamond TL, Zhou H, Hazuda DJ, Espeseth AS, König R (2009) Host cell factors in HIV replication: meta-analysis of genome-wide studies. *PLoS Pathogens* **5**: e1000437
- Carpenter A, Jones T, Lamprecht M, Clarke C, Kang I, Friman O, Guertin D, Chang J, Lindquist R, Moffat J (2006) CellProfiler: image analysis software for identifying and quantifying cell phenotypes. *Genome Biol* **7**: R100
- Cherry S (2009) What have RNAi screens taught us about viral-host interactions?. *Curr Opin Microbiol* **12**: 446–452
- Chickering D (2003) Optimal structure identification with greedy search. *J Mach Learn Res* **3**: 554
- Collinet C, Stöter M, Stöter M, Bradshaw CR, Samusik N, Rink JC, Kenski D, Habermann B, Buchholz F, Henschel R, Mueller MS (2010) Systems survey of endocytosis by multiparametric image analysis. *Nature* **464**: 243–249
- Colman-Lerner A, Gordon A, Serra E, Chin T, Resnekov O, Endy D, Pesce CG, Brent R (2005) Regulated cell-to-cell variation in a cell-fate decision system. *Nature* **437**: 699–706
- Damm EM, Pelkmans L (2006) Systems biology of virus entry in mammalian cells. *Cell Microbiol* **8**: 1219–1227
- Dennis Jr G, Sherman BT, Hosack DA, Yang J, Gao W, Lane HC, Lempicki RA (2003) DAVID: database for annotation, visualization, and integrated discovery. *Genome Biol* **4**: P3
- Doherty G, McMahon H (2009) Mechanisms of endocytosis. *Annu Rev Biochem* **78**: 857–902

- Echeverri CJ, Perrimon N (2006) High-throughput RNAi screening in cultured cells: a user's guide. *Nat Rev Genet* **7**: 373–384
- Eldar A, Elowitz M (2010) Functional roles for noise in genetic circuits. *Nature* **467**: 167
- Elowitz MB, Levine AJ, Siggia ED, Swain PS (2002) Stochastic gene expression in a single cell. *Science* **297**: 1183–1186
- Forsthoefel D, Liebl E, Kolodziej P, Seeger M (2005) The Abelson tyrosine kinase, the Trio GEF and Enabled interact with the Netrin receptor Frazzled in *Drosophila*. *Development* **132**: 1983
- Fukuhara S, Sako K, Minami T, Noda K, Kim H, Kodama T, Shibuya M, Takakura N, Koh G, Mochizuki N (2008) Differential function of Tie2 at cell–cell contacts and cell–substratum contacts regulated by angiopoietin-1. *Nat Cell Biol* **10**: 513–526
- Gumbiner B (1996) Cell adhesion: the molecular basis of tissue architecture and morphogenesis. *Cell* **84**: 345
- Helenius A, Kartenbeck J, Simons K, Fries E (1980) On the entry of Semliki Forest virus into BHK-21 cells. *J Cell Biol* **84**: 404
- Hopkins AL, Groom CR (2002) The druggable genome. *Nat Rev Drug Discov* **1**: 727–730
- Huh D, Paulsson J (2010) Non-genetic heterogeneity from stochastic partitioning at cell division. *Nat Genet* **43**: 95–100
- Hurwitz M, Vanderzalm P, Bloom L, Goldman J, Garriga G, Horvitz H (2009) Abl kinase inhibits the engulfment of apoptotic cells in *Caenorhabditis elegans*. *PLoS Biol* **7**: e1000099
- Jackson AL, Bartz SR, Schelter J, Kobayashi SV, Burchard J, Mao M, Li B, Cavet G, Linsley PS (2003) Expression profiling reveals off-target gene regulation by RNAi. *Nat Biotechnol* **21**: 635–637
- Kim J, Lee J, Heynen-Genel S, Suyama E, Ono K, Lee K, Ideker T, Aza-Blanc P, Gleeson J (2010) Functional genomic screen for modulators of ciliogenesis and cilium length. *Nature* **464**: 1048–1051
- Kittler R, Putz G, Pelletier L, Poser I, Heninger AK, Drechsel D, Fischer S, Konstantinova I, Habermann B, Grabner H (2004) An endoribonuclease-prepared siRNA screen in human cells identifies genes essential for cell division. *Nature* **432**: 1036–1040
- Lamaze C, Chuang T, Terlecky L, Bokoch G, Schmid S (1996) Regulation of receptor-mediated endocytosis by Rho and Rac. *Nature* **382**: 177–179
- Liebl E, Forsthoefel D, Franco L, Sample S, Hess J, Cowger J, Chandler M, Shupert A, Seeger M (2000) Dosage-sensitive, reciprocal genetic interactions between the Abl tyrosine kinase and the putative GEF trio reveal trio's role in axon pathfinding. *Neuron* **26**: 107–118
- Li Q, Brass AL, Ng A, Hu Z, Xavier RJ, Liang TJ, Elledge SJ (2009) A genome-wide genetic screen for host factors required for hepatitis C virus propagation. *Proc Natl Acad Sci* **106**: 16410–16415
- Luo KQ, Chang DC (2004) The gene-silencing efficiency of siRNA is strongly dependent on the local structure of mRNA at the targeted region. *Biochem Biophys Res Commun* **318**: 303–310
- Martin D, Souillard A, Hall M (2004) TOR regulates ribosomal protein gene expression via PKA and the Forkhead transcription factor FHL1. *Cell* **119**: 969–979
- Moffat J, Sabatini DM (2006) Building mammalian signalling pathways with RNAi screens. *Nat Rev Mol Cell Biol* **7**: 177–187
- Mohr S, Bakal C, Perrimon N (2010) Genomic screening with RNAi: results and challenges. *Biochemistry* **79**: 37
- Neumann B, Walter T, Jean-Karim H (2010) Phenotypic profiling of the human genome by time-lapse microscopy reveals cell division genes. *Nature* **464**: 721–727
- Newman J, Ghaemmaghami S, Ihmels J, Breslow D, Noble M, DeRisi J, Weissman J (2006) Single-cell proteomic analysis of *S. cerevisiae* reveals the architecture of biological noise. *Nature* **441**: 840–846
- Ohno S (2001) Intercellular junctions and cellular polarity: the PAR-aPKC complex, a conserved core cassette playing fundamental roles in cell polarity. *Curr Opin Cell Biol* **13**: 641–648
- Ozbudak E, Thattai M, Kurtser I, Grossman A, van Oudenaarden A (2002) Regulation of noise in the expression of a single gene. *Nat Genet* **31**: 69–73
- Parsons J, Horwitz A, Schwartz M (2010) Cell adhesion: integrating cytoskeletal dynamics and cellular tension. *Nat Rev Mol Cell Biol* **11**: 633–643
- Pelkmans L (2009) Systems biology of virus infection in mammalian cells. *Curr Opin Microbiol* **12**: 429–431
- Pelkmans L, Fava E, Grabner H, Hannus M, Habermann B, Krausz E, Zerial M (2005) Genome-wide analysis of human kinases in clathrin- and caveolae/raft-mediated endocytosis. *Nature* **436**: 78–86
- Pelkmans L, Kartenbeck J, Helenius A (2001) Caveolar endocytosis of simian virus 40 reveals a new two-step vesicular-transport pathway to the ER. *Nat Cell Biol* **3**: 473–483
- Raj A, van Oudenaarden A (2009) Single-molecule approaches to stochastic gene expression. *Annu Rev Biophys* **38**: 255–270
- Ramo P, Sacher R, Snijder B, Begemann B, Pelkmans L (2009) CellClassifier: supervised learning of cellular phenotypes. *Bioinformatics* **25**: 3028–3030
- Raser J, O'Shea E (2005) Noise in gene expression: origins, consequences, and control. *Science* **309**: 2010
- Sarbasov D, Ali S, Sabatini D (2005) Growing roles for the mTOR pathway. *Curr Opin Cell Biol* **17**: 596–603
- Schellhaas M, Malmström J, Pelkmans L, Haugstetter J, Ellgaard L, Grünewald K, Helenius A (2007) Simian Virus 40 depends on ER protein folding and quality control factors for entry into host cells. *Cell* **131**: 516–529
- Schultz N, Marenstein DR, De Angelis DA, Wang WQ, Nelander S, Jacobsen A, Marks DS, Massagué J, Sander C (2011) Off-target effects dominate a large-scale RNAi screen for modulators of the TGF- pathway and reveal microRNA regulation of TGFBR2. *Silence* **2**: 3
- Scita G, Di Fiore P (2010) The endocytic matrix. *Nature* **463**: 464–473
- Scita G, Nordstrom J, Carbone R, Tenca P, Giardina G, Gutkind S, Bjarnegård M, Betsholtz C, Di Fiore P (1999) EPS8 and E3B1 transduce signals from Ras to Rac. *Nature* **401**: 290–293
- Shapiro J (1998) Thinking about bacterial populations as multicellular organisms. *Annu Rev Microbiol* **52**: 81–104
- Sigal A, Milo R, Cohen A, Geva-Zatorsky N, Klein Y, Liron Y, Rosenfeld N, Danon T, Perzov N, Alon U (2006) Variability and memory of protein levels in human cells. *Nature* **444**: 643–646
- Singh DK, Ku CJ, Wichaidit C, Steiner III RJ, Wu LF, Altschuler SJ (2010) Patterns of basal signaling heterogeneity can distinguish cellular populations with different drug sensitivities. *Mol Syst Biol* **6**: 369
- Sini P, Cannas A, Koleske A, Di Fiore P, Scita G (2004) Abl-dependent tyrosine phosphorylation of Sos-1 mediates growth-factor-induced Rac activation. *Nat Cell Biol* **6**: 268–274
- Snijder B, Pelkmans L (2011) Origins of regulated cell-to-cell variability. *Nat Rev Mol Cell Biol* **12**: 119–125
- Snijder B, Sacher R, Ramo P, Damm EM, Liberali P, Pelkmans L (2009) Population context determines cell-to-cell variability in endocytosis and virus infection. *Nature* **461**: 520–523
- Takeichi M (1991) Cadherin cell adhesion receptors as a morphogenetic regulator. *Science* **251**: 1451
- Zandy NL, Pendergast AM (2008) Abl tyrosine kinases modulate cadherin-dependent adhesion upstream and downstream of Rho family GTPases. *Cell Cycle* **7**: 444–448



Molecular Systems Biology is an open-access journal published by *European Molecular Biology Organization* and *Nature Publishing Group*. This work is licensed under a Creative Commons Attribution-Noncommercial-Share Alike 3.0 Unported License.