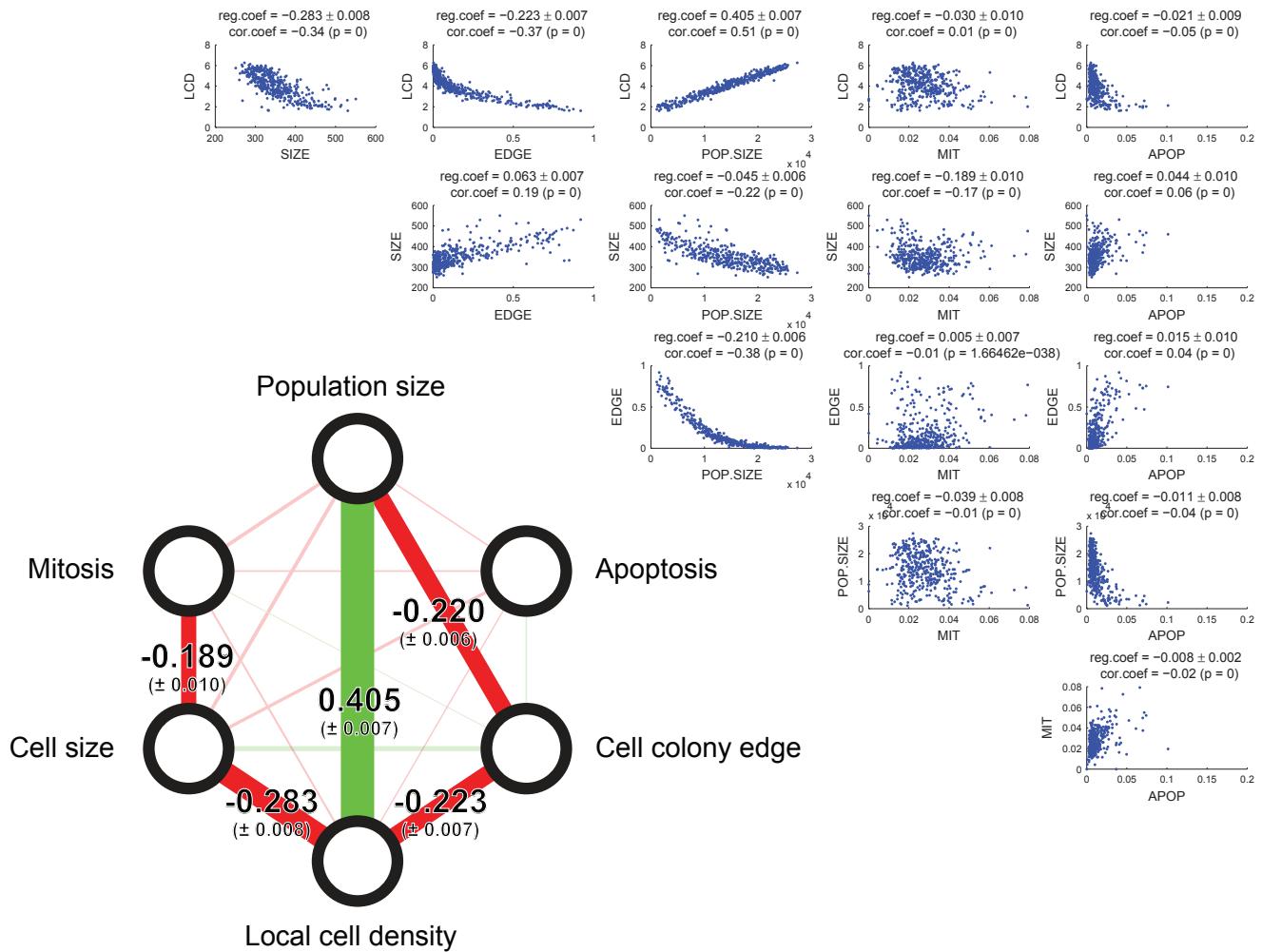


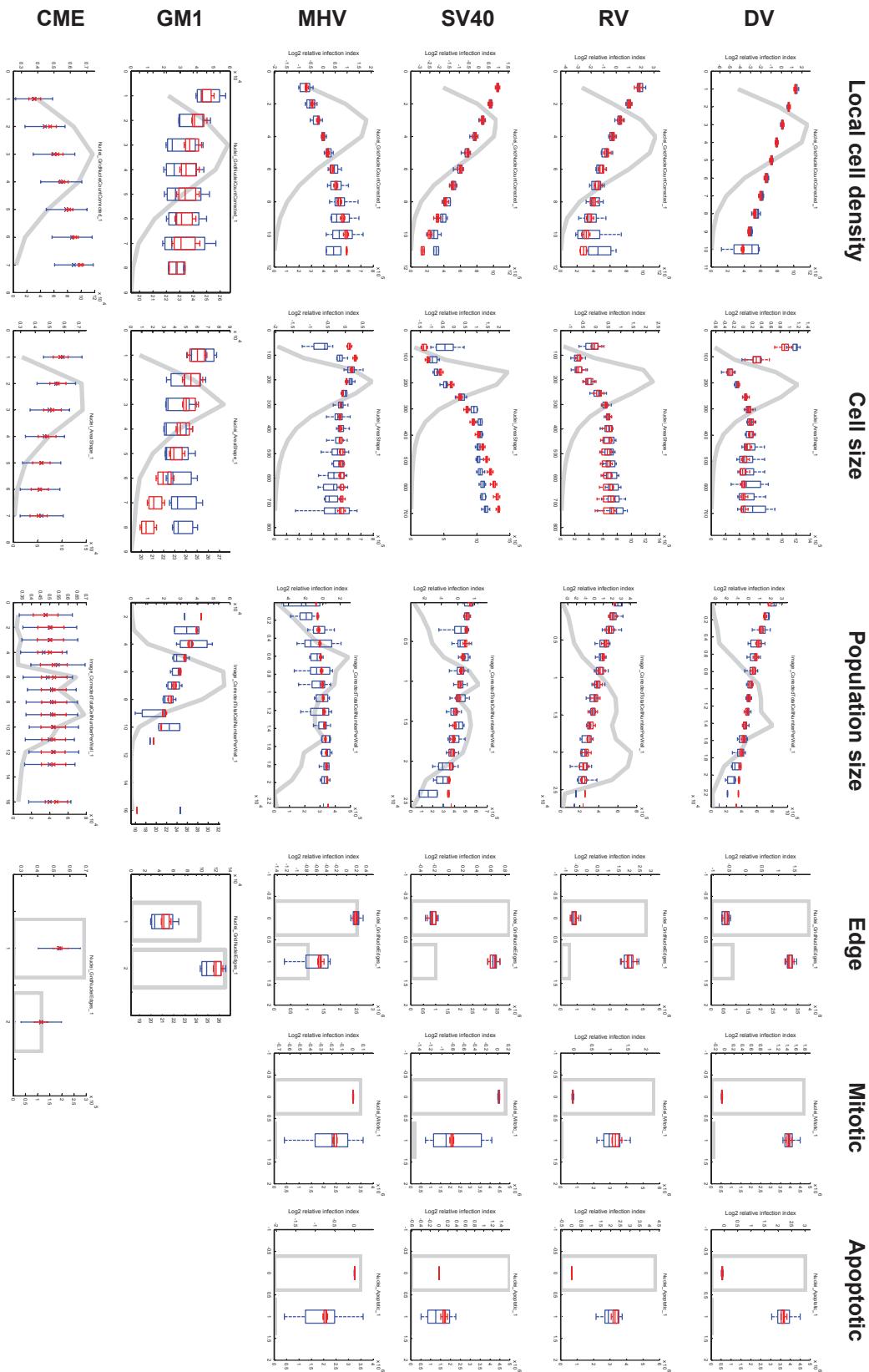
Supplementary results:

Supplementary Figure 1



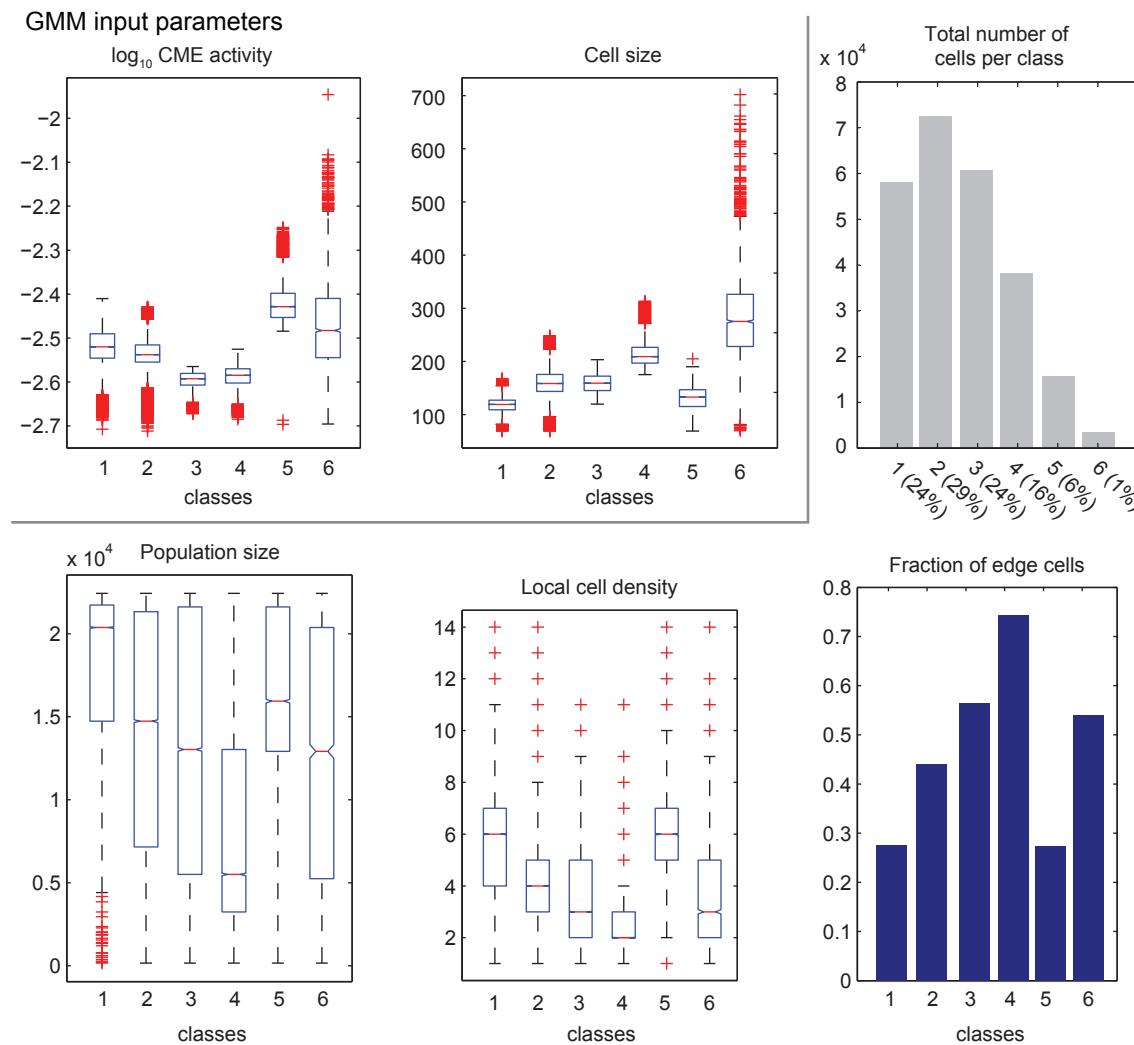
Supplementary figure 1 | Graphical Gaussian modelling reveals links between the population context features. Pairwise scatter plots of average population property values per population for 414 individual populations, with a total of 5.0×10^6 cells. Strong correlations can be observed between population averaged population properties. Correlation coefficients and corresponding p-values are calculated for all population property pairs on individual cell measurements (indicated by *cor.coef*). We further performed linear regression analysis on z-score-normalised population properties. Each population property was taken as dependent variable, and explained by a linear model consisting of all other population properties, a procedure also known as graphical Gaussian modelling¹. Average regression coefficients (*reg.coef*, \pm standard deviations) derived from 100 regression analyses on 2×10^4 randomly selected cells are shown, and regression coefficients stronger than > 0.1 or < -0.1 are visualized as edges connecting nodes in a graph. The network quantifies the intuitive notion that as a population of cells grows, the local cell density increases and both effects lead to a decrease in the number of cells located on edges of cell islets. The increase in local cell density leads to a decrease in the cell size. The link between mitosis and cell size is trivial since mitotic cells are smaller than non-mitotic cells. LCD, local cell density; EDGE, cells residing on cell colony edges, or inside cell colonies; POP.SIZE, total number of cells in a cell population; SIZE, cell size; MIT, mitotic cells; APOP, apoptotic cells.

Supplementary Figure 2



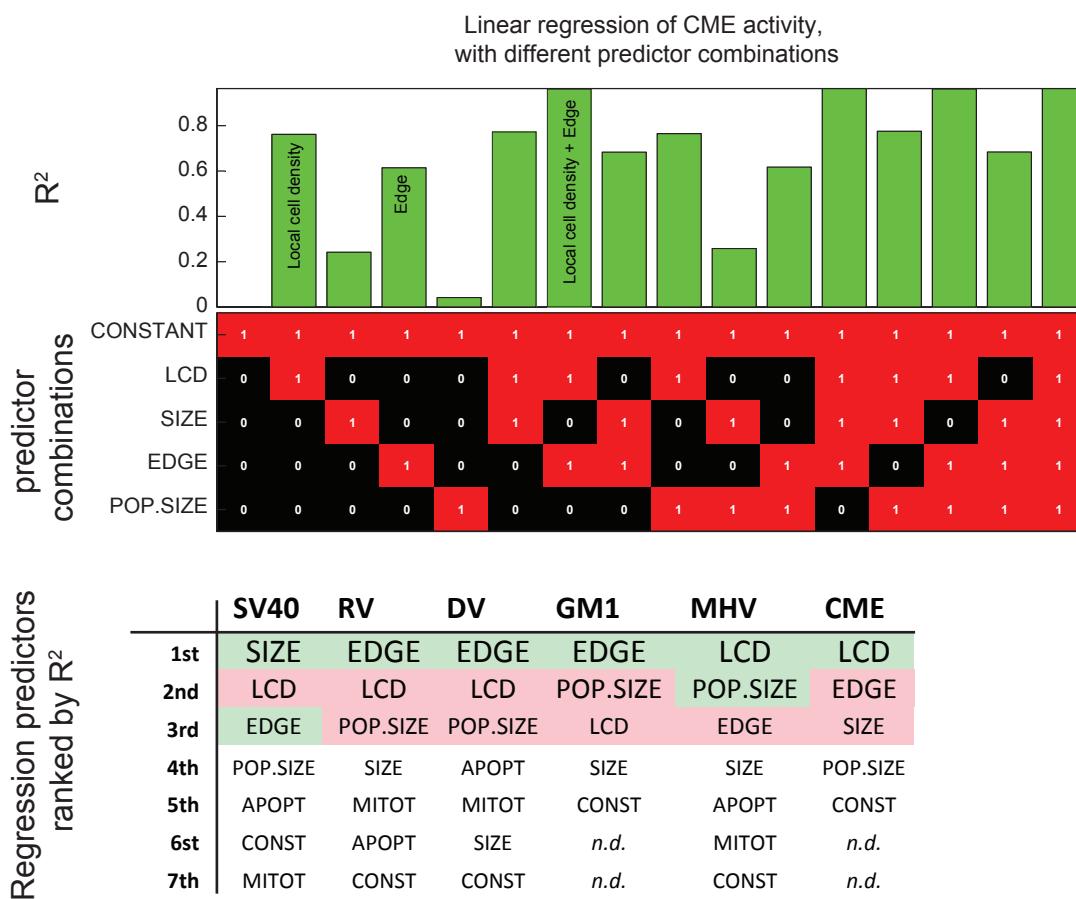
Supplementary figure 2 | Probit regression models of virus infection and weighted linear regression models of GM1 content and CME activity. Measured and modelled \log_2 transformed relative infection-indices (infected cells / total cells), and trends are shown for all population-context properties. Measurements are represented by blue boxplots and their corresponding model fits by red boxplots. Light grey lines and bar plots indicate total number of cells corresponding to the data point (see right axis in each graph). Note that the GM1 model has boxplots of three replicate datasets, whereas the CME model shows the single cell average values (\pm standard deviations) of a single dataset.

Supplementary Figure 3



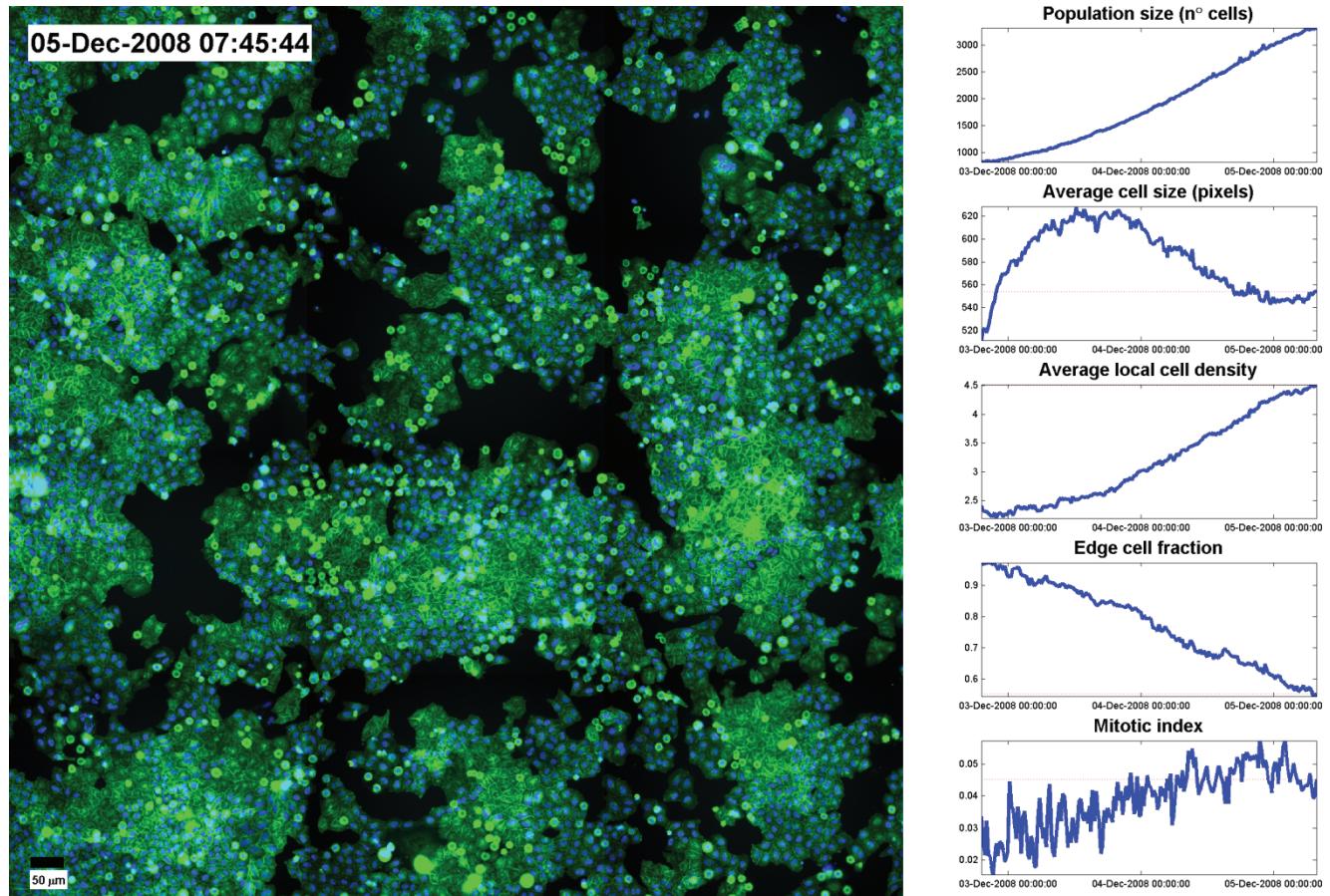
Supplementary figure 3 | Multivariate clustering of CME activity phenotypes using Gaussian mixture modelling identifies diverse subpopulations present in unperturbed cells. Gaussian mixture modelling with a 6 component model on CME activity and cell size readouts reveals distinct subpopulations present in CME, in 2.5×10^6 HeLa cells. However, in particular the first 4 classes comprising 94% of all cells (upper right panel) differ significantly in their micro-environmental variables (lower three panels), indicating that the different phenotypes correlate with changes in the population context, which confirms the correlations found by the regression modelling approach.

Supplementary Figure 4



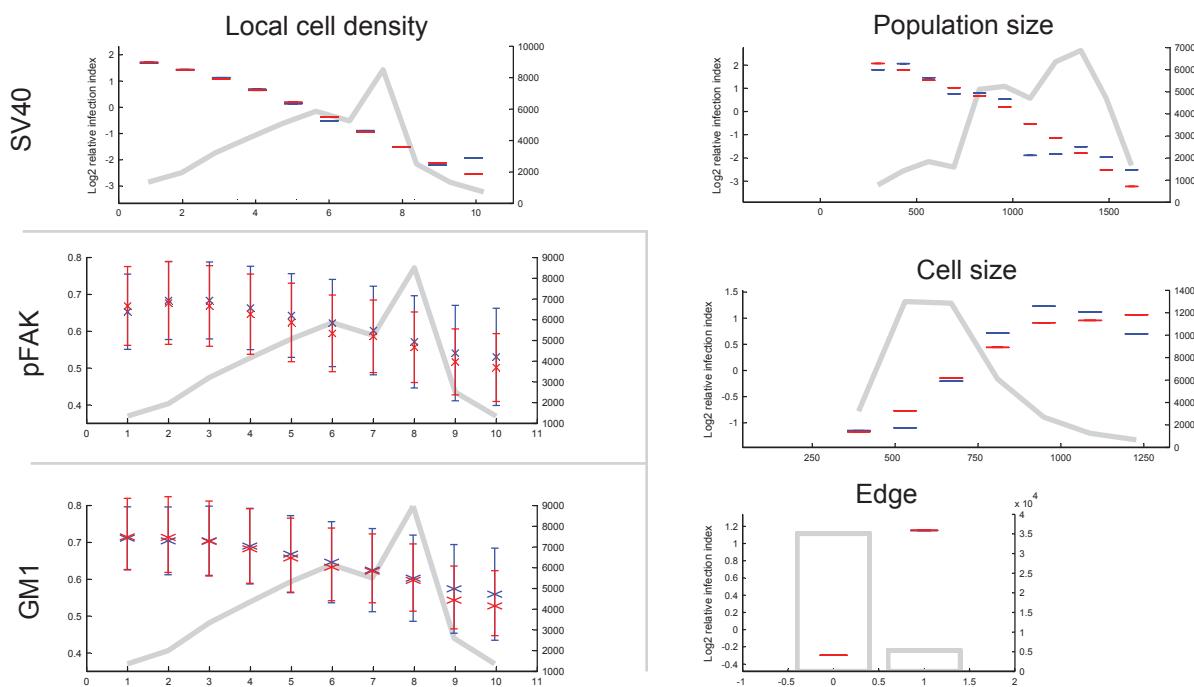
Supplementary figure 4 | Ranked explanatory power determined per individual model parameter reveals the strongest determining parameters for each activity. **a.** The explanatory power, or R^2 values, of linear models of CME activity against all possible combinations of population context predictors, identify the local cell density and cell islet edges as the first and second strongest regulators of CME activity. **b.** A similar analysis was done for all models, in which probit-regression was used for the infection assays, and weighted least squares regression for the CME and GM1 models. Explanatory power of the individual components was estimated using either the weighted residuals for the weighted least squares models, or using the squared correlation of the predicted number of infected cells with the measured number of infected cells per particular population context bin. For each cellular activity the population context parameters are ranked according to the strongest explanatory power for each individual parameter. The top 3 ranking population-context parameter are indicated in Fig. 1d of the main text by the strength of colour. (green background indicates positive regression coefficient, red background a negative one).

Supplementary Movie 1



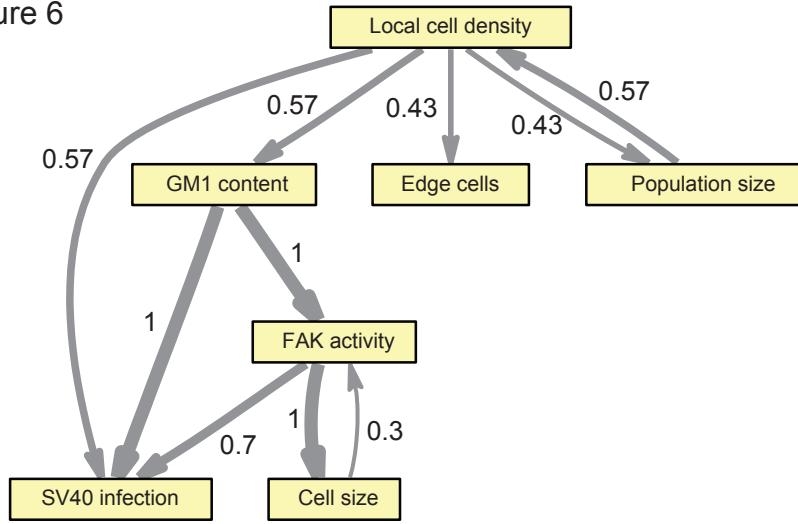
Supplementary movie 1 | Population properties are determined during growth of adherent human cells. Live cell imaging of growing HeLa Kyoto cells over-expressing GPI-GFP (green channel) and H2B-mCherry (blue channel) during a period of ~3 days. Note how during growth the local cell density and position of cells on cell islet edges are largely determined. Scale bar, 50μm. You can find the full movie in the online supplementary information.

Supplementary Figure 5



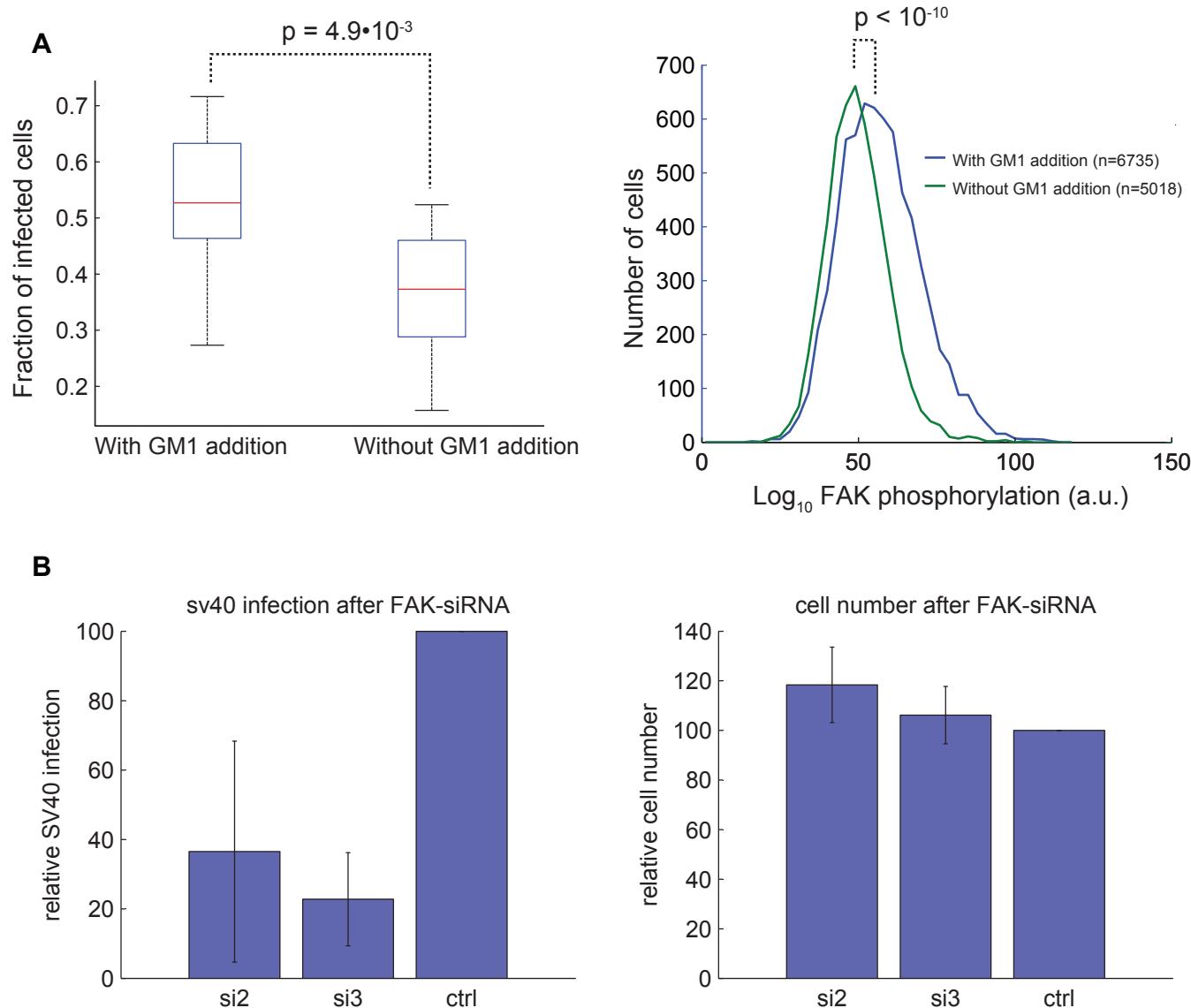
Supplementary Figure 5 | Heterogeneity patterns in diploid human MCF10A cells. **a.** The probit-model fit of SV40 infection in 4.0×10^5 human diploid MCF10A cells is shown (single replicate, blue lines measured infection, red lines model-fit), on a \log_2 transformed relative infection index scale. Note that the pattern of infection is the same as in HeLa cells, with large and sparse cells being most highly infected by SV40. **b.** FAK phosphorylation levels and GM1 content on the plasma membrane show similar heterogeneity patterns in MCF10A cells as SV40 infection does, with sparse cells having the highest levels for both activities.

Supplementary Figure 6



Supplementary figure 6 | Edge retrieval rates of bootstrapped Bayesian network learning on population context parameters, GM1 content, FAK phosphorylation and SV40 infection in MCF10A cells. Single cell MCF10A data was discretized while maximizing Akaike's information criterion, and 100 bootstrap runs of Bayesian structure learning was performed using a greedy search algorithm that searches the Markov equivalent space on randomly selected 75% of the data. For each edge the fraction of 100 bootstrap runs in which the edge was retrieved is indicated.

Supplementary Figure 7



Supplementary figure 7 | Exogenous GM1 increases SV40 infection and FAK phosphorylation in MCF10A cells. **a.** Exogenous addition of GM1 to the cell culture media of MCF10A cells leads to an increase in SV40 infection and (left panel, $p = 4.9 * 10^{-3}$), and FAK phosphorylation (right panel, $p < 10^{-10}$) as compared to MCF10A cells to which no GM1 was added. **b.** RNAi knockdown of two independent siRNAs against FAK strongly reduces SV40 infection. Measurements show average and standard deviation of 3 replicate measurements, normalized to control (100%).

Supplementary methods:

Supplemental Materials and Methods

Cell lines & 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 A431 cells from ATCC (Molsheim Cedex). Primary human endothelial (HDMEC) cells were a kind gift from Michael Detmar, Zurich, and originally obtained from PromoCell. HeLa Kyoto cells over-expressing GPI-GFP and H2B-RFP fusion proteins were a kind gift from Daniel Gerlich (ETH Zurich)². Epithelial MCF10A cells were obtained from Ruth Kroschewski (ETH Zurich).

Infection and endocytosis assays

MHV infection was performed in HeLa Kyoto cells stably transfected with CEACAM1a, as described by Verheij et al.³; SV40 infection was performed in HeLa MZ cells as described earlier⁴; The RV infection assay was performed in HeLa Kyoto cells as described⁵; DV infection was performed in HeLa Kyoto cells stably expressing DC-SIGN, essentially as described by Lozach et al.,⁶. The GM1 binding assay was performed with Cholera toxin B (ChTxB) conjugated to AlexaFluor568⁷, which was incubated on A431 cells for 5' prior to fixing. Tf_n was incubated on cells for 15 minutes, according to standard protocol. Prior to fixation, an acid wash removed non-internalised Tf_n⁷. All assays were performed in 96-well plates. Cells were fixed and the nuclei were stained with DAPI before imaging.

Imaging

Per well, 3×3 adjacent images were taken per well for the virus infection and ChTxB assays, and 5×5 images for the Tf_n assay. Images were acquired on automated widefield cellWoRx™ microscopes (Applied Precision) or ImageXpress Micro microscopes (Molecular Devices) with a 10× objective for the infection assays, and with a 20× objective for the ChTxB binding assay and the Tf_n uptake assay. Either image based auto-focussing was performed on the DAPI signal, or laser based autofocussing on the multi-well plate bottom. The images were recorded with 12-bit CCD cameras, and stored as 16-bit uncompressed TIFF files.

RNAi knockdown of FAK

siRNA transfection was done in a reverse manner in 384-well plates, seeding 1000 HeLa CNX cells into wells already containing 10ul of transfection mix. The transfection mix for one well consisted of 1 μM siRNA diluted in DEPC-water, mixed 1:1 with Optimem. Lipofectamine2000 (Qiagen) was incubated for 15 min in Optimem to a concentration of 1% and was then added 1:1 to the siRNA. 10 μl of this final mix were added to each well. To each well 1000 cells were added, suspended in 40 μl of growth medium (DMEM, Gibco#41965039; 10% FCS, Sigma #F7524 lot.# 085K3396; 1% Glutamax, Gibco#35050038 and 1% of penicillin-streptomycin, Invitrogen #15140). The plates were incubated for 62 h at 37°C before infection. For infection, the liquid within each well was replaced with 40 μl virus suspension diluted in R-medium. R-medium consisted of 0.2% BSA in RPMI buffered with 10 mM pH 6.8 Hepes. Plates were imaged with 10× objective and infection was quantified for triplicate measurements for two independent siRNAs as described below.

MCF10A infection, immunofluorescence and GM1 addback

Cells were grown to 3 different subconfluent densities in a 96 well plate. 16 hours before fixing SV40 infection was performed as described above. 5 minutes prior to fixation fluorescently labelled ChTxB

was added to the cells to visualize the GM1 distribution of cells. Cells were fixed in 4% formaldehyde, permeabilized with 0.1 % Triton X-100, blocked with 0.5% BSA and incubated with the rabbit-anti phosphorylated-FAK [Y397] antibody (Bioscience, Camarillo, CA, USA.) for 1.5 h at RT, followed by incubation with an Alexa Fluor labeled secondary antibody for 45 min at RT. Images were acquired with 20 \times magnification, taking 16 pictures per well. To test the effect of exogenous GM1 addition on SV40 infection and pFAK levels, MCF10A cells were grown to different confluencies in 96-a well plate either in SFM, or in SFM supplemented with 14 μ g/ml GM1 complexed to fatty acid free BSA o.n. at 37°C. Next, SV40 infection, GM1 levels and pFAK levels were assessed as described above.

Live cell imaging

Cells stably transfected with H2B-mCherry and GPI-GFP were cultured in selective medium (G418 & purimycin). For imaging, 800 cells were seeded in a 96 well plate without selective pressure and imaged with an MD microscope equipped with environmental temperature and CO₂ control. Acquisition started 60 minutes after initial seeding, and images were acquired with a 20 minute interval, with a 50 millisecond exposure time per channel. Images were analyzed as described below, measuring population properties for each time point from the H2B signal.

Computational image analysis

The images were analysed with open-source cell image analysis software⁸ (www.cellprofiler.org) combined with image analysis methods specifically designed for this study and implemented in MATLAB. All computational image analysis was run on a high performance computer cluster. The general image analysis pipeline was as follows: first, nuclei were identified based on the DAPI stain. Next, cell boundaries were estimated using nuclear expansion for the virus infection and Tfn uptake assays, or detected using the ChTxB signal in the GM1 binding assay. Standard CellProfiler texture, intensity, size and shape features were extracted from nucleus and cell regions. We additionally developed and implemented several image analysis steps for the purpose of the automated detection of infected nuclei, detection of out of focus images, for the measurement of the local cell density and cell colony edges, and for the SVM-based classification of interphase, mitotic, and apoptotic nuclear classes.

Detection of out-of-focus images

Images were marked as out-of-focus if the minimal value of the DAPI image granularity spectrum measured by CellProfiler was below a sliding threshold. This threshold was set lower for higher nucleus-counts per image, correcting for the observation that images with many DAPI stained nuclei are less granular than image with fewer cells. Data obtained from out of focus images were discarded.

Local cell density and cell colony edge measurements

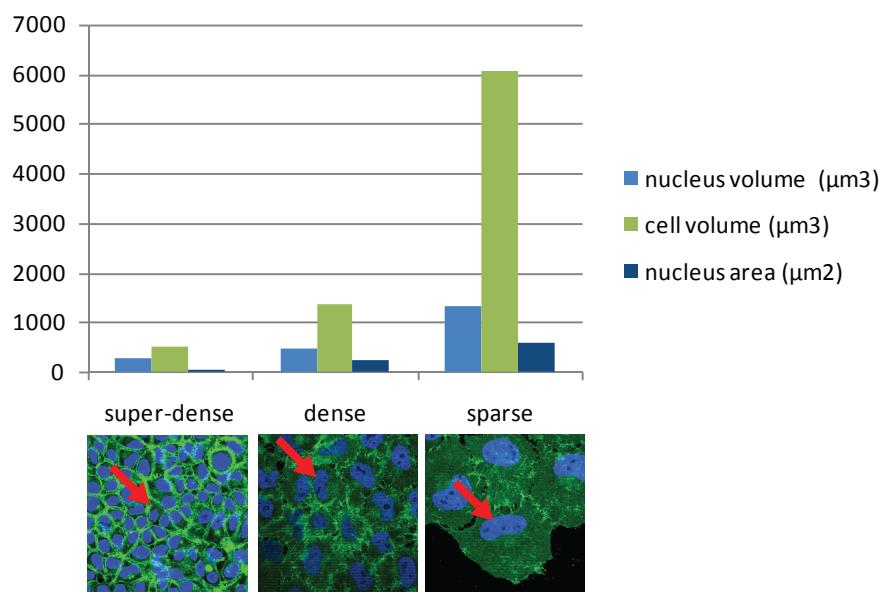
Local cell density and cell colony edge detection was analysed by overlaying an exact matching grid with a grid size of 24 \times 26 pixels for 10 \times or 48 \times 52 pixels grid size for 20 \times images. The local cell density for a given nucleus was calculated as the total number of nuclei with weighted centres falling in the grid-unit in which that nucleus resides. Grid sizes were optimised to give a linear range of local cell densities for the observed total cell numbers, as can be seen in the population size vs. average local cell density (LCD) scatter-plot in Supplementary Figure 1. This resulted in grid sizes such that LCD values range from 1 to approximately 14. Additionally, nuclei residing at grids bordering empty grid units, including those touching the corners or the empty grid, were marked as cell colony edge nuclei, whereas nuclei in grids that did not border empty grids were marked as cells inside cell colony edges. Example LCD measurement and EDGE detection is shown in Figure 1b (main text). Notably, this local cell density measurement differs from a number-of-neighbours type of local cell density measurement in that those

cells with maximum values are not larger cells surrounded by many small objects, as is the case with number-of-neighbours type of local cell density measurements, but maximum values are obtained for rather smaller cells in crowded areas.

Cell size measurements

Because the nuclei of adherent human cells have more regular shapes than the cells themselves, and because they are more faithfully detectable with object segmentation in automated image analysis, we used the nucleus 2D surface-area as a representative for the cell size. Although it is well known that the regulation of nucleus volume is tightly coupled to the regulation of cell volume^{9 10 11}, we additionally tested if the nucleus areas is a good representative for the cell size. We measured the nucleus and cell volumes from confocal z-stack images of HeLa cells grown at high, intermediate and low local cell densities (Supplementary Figure 8). Manual segmentation of these nuclei and cell borders confirms the correlation between the observed nucleus area, nucleus volume and cell volume, additionally confirming that cells reduce their cell and nucleus volume in dense areas. We therefore used the surface area of the nucleus as our readout for cell size.

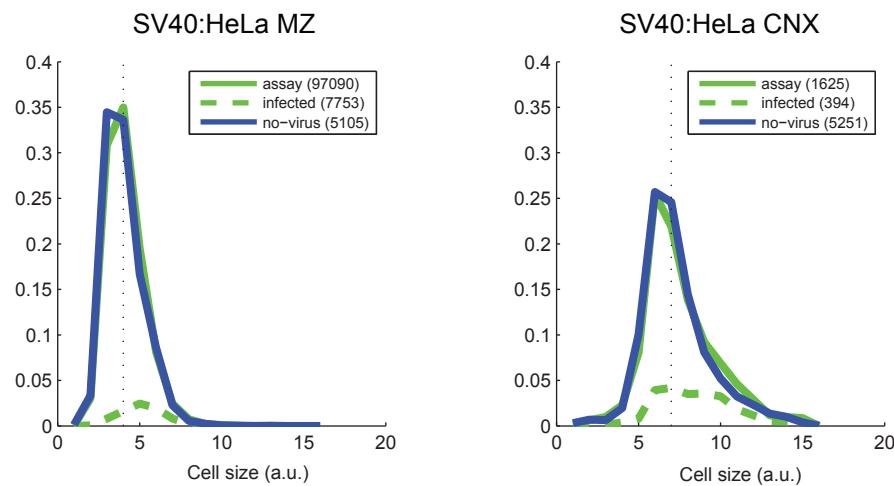
Supplementary Figure 8



Supplementary figure 8 | Cell and nucleus volumes negatively correlate with local cell density. Z-stack immunofluorescence images were acquired on a confocal microscope covering the entire height of HeLa cells transfected with GPI-GFP to mark the plasma membrane and stained with DAPI to mark the nucleus. Manual segmentation was applied on selected cells (red arrows) to reconstruct the cell volumes, nuclear volumes and nucleus areas at different local cell densities. The cell and nucleus volumes of dense cells are significantly smaller than those of sparse cells. Note that the nucleus area measure underestimates the increase in cell volume.

We next tested if virus infection induces changes in the cell state, as opposed to the cell state determining virus infection. Since SV40 infection was observed to occur mostly in large spread out cells, we tested whether SV40 infection might lead to an increase in nucleus size. However, results obtained from two different HeLa cell lines (CNX and MZ) confirms that SV40-infected cells do not have a bigger nucleus as compared to cells in the same population context that were not exposed to SV40 virus (Supplementary Figure 9).

Supplementary Figure 9



Supplementary figure 9 | SV40 infection does not induce an increase in cell size. A comparison of cell size distributions of cells that were infected with SV40 and cells that were not infected with SV40, but that are otherwise from equal population contexts, shows no difference between the size distribution of virus-exposed cells compared to non virus-exposed cells. This indicates that SV40 infection does not induce an increase in cell size, but that SV40 infection preferentially occurs in cells that are large. Results are shown for SV40 infection in two different HeLa cell lines. Dashed green lines correspond to the fraction of virus-exposed cells (solid green line) that is infected.

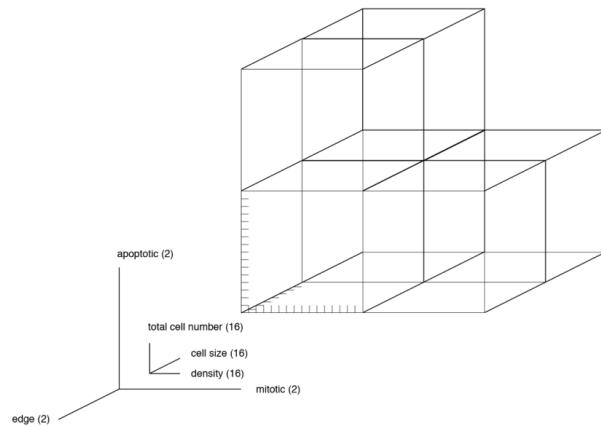
Nuclear classification

We performed nuclear classification to detect the mitotic, apoptotic and interphase states of individual cells, and to discard cells resulting from erroneous object detection. We manually classified nuclei for each class, in a binary fashion, i.e. “interphase” or “non-interphase”. We next applied a support vector machine (SVM) learning implementation (STPR-toolbox)¹² for the classification of each nucleus in each class separately, correcting training errors in an iterative fashion. Classification was performed using all CellProfiler extracted features from the DAPI channel and nuclei objects, and training was corrected in an iterative fashion, correcting wrongly classified nuclei, until satisfactory classification of phenotypes was obtained. Although this training on wrongly classified nuclei biases the training set towards difficult to score nuclei, we obtained good results with the SVM. The interphase class was focussed on well-segmented nuclei, whereas the non-interphase class was trained to include typical object detection problems such as incorrect fusion or separation of nuclei, or nuclei detected from out of focus areas of images. The mitotic class was focussed on M-phase nuclei, and the apoptotic class on blebbing and fragmentation of nuclei. Finally, a scoring scheme was applied to decide the per nucleus class. Nuclei that were classified as interphase and additionally not as mitotic were regarded as interphase cells. Nuclei that were classified as mitotic and additionally not as interphase were regarded as mitotic cells. Cells that were exclusively classified as apoptotic were regarded as apoptotic. All nuclei that were differently classified were excluded from further analysis. Note that for both the Tfⁿ uptake assay and for the ChTx^B binding assays mitotic and apoptotic classes were not classified due to the relatively low abundance of those phenotypes, and cells were classified either as interphase or as non-interphase. Non-interphase cells were excluded from further analysis.

Statistical analysis

Modelling cellular activity on population-context properties

To model the effect of the population properties we first annotated each individual nucleus that passed the classification by the SVM for the following 6 properties. (1) The population size (PS), which we calculated as the total number of cells counted for all images coming from the same well, compensating for the loss of nuclei due to out of focus images of that well, but including nuclei that would be discarded based on the SVM classification. (2) The local cell density (LCD), (3) the cell size (SIZE), (4) the position inside or on the edge of a cell colony (EDGE), and (5) the mitotic (MIT) or (6) apoptotic (APO) state. The latter 5 classes were calculated as described above. Note that all 6 population-context properties can be computed from images of just a nuclear stain. We next subdivided this densely occupied feature space in discrete groups or bins of cells with similar population properties. This binning was optimised per assay such that the data for PS, and SIZE were distributed over 16 linearly increasing bins, with all bins containing cells. Binning of LCD, which has integer values, was optimised such that the number of bins matches the difference between the lowest and highest value present, often being slightly less than 16. The boolean values of the population properties EDGE, MIT, and APO did not require further binning. Each cell was annotated with the value for all 6 population properties, where values determined by binning of PS, SIZE or LCD were shifted back 1 value such that 0 was the minimal value for all population properties. Therefore, cells that are neither apoptotic nor mitotic are by default interphase, and they form the bulk of the data. These values lead to the formation of a multi-dimensional feature space where each dimension corresponds to an individual population property, as is depicted in Supplementary Figure 10.



Supplementary figure 10 | The population properties define a multidimensional feature space. The population properties define a multi-dimension feature space describing a cell's state and its population context. The typical dimensions of this multi-dimensional space can be visualised as depicted above, where interphase cells are cells that are neither mitotic nor apoptotic, and thus occupy the lower left two of the larger cubes, and cells on cell colony edges occupy the front set of three cubes, whereas cells inside cell colony edges occupy the back set of three cubes. Note that 'total cell number' is equal to the 'population size', and that the dimension defined by the local cell density can be smaller than 16 depending on the data.

We next calculated the corresponding cellular activity for each assay, over all cells for a given bin. The cellular activity of infection was expressed as the total number of infected cells over the total number of cells per bin. The cellular activity of the ChTxB or Tfn assays was defined as the \log_{10} transformed average cellular intensity per cell. To reduce the memory footprint of the intensity measurements we additionally binned these \log_{10} transformed intensity readouts over all cells over 128 bins, allowing us to reduce the data type to unsigned 8 bit integers, but still ensuring a reasonable resolution of cellular

activity. The cellular activity per population-context bin for both Tfn and ChTx B was finally calculated as the average intensity-bin value over all cells in that bin.

Model parameter estimation

To identify the individual effects of the population properties on the cellular activity of interest, we performed multi-linear regression taking the cellular activity readouts per bin as response variable (Y) and the population properties per bin as explanatory variables (x). Since infection is a probabilistic event following a Binomial distribution, we used probit-regression to model the likelihood of infection for individual cells, where the probability of being infected (P) given the discretized values of the population-context predictors (x) follows: $P(Y = 1|X = x) = \Phi(x'\beta)$, and Φ is the cumulative distribution function of the standard normal distribution, and β the regression coefficients. R^2 values were estimated by the squared correlation coefficient of the measured number of infected cells for each x , against the model-predicted number of infected cells. For the models of GM1 content and CME activity, we used weighted least squares solution to solve the linear model. We calculated the weight of an observation as $\sqrt[p]{n}$, where n is the number of cells with that particular population context values, and we empirically determined $p = 3$ for all assays. For bins $n < 64$ for the CME and GM1 based assays we set the weight to 0, effectively discarding these rare observations from the parameter estimation. R^2 values were calculated using the weighted squared errors of the model fit. Confidence values were calculated for both the probit and weighted types of model parameters based on the t-statistic of each parameter. Non-significant parameters were discarded from the model, and parameter estimation was repeated. Parameter estimation was found to be very reproducible, as model parameters estimated for the same activity but from different experiments showed very little variation (data not shown).

To address the issue that explanatory variables need to be uncorrelated for optimal parameter estimation using the weighted least squares solution and the probit model, we tested if multicollinearity occurred in our data. We first tested the robustness of model parameters during the omission of explanatory variables, and observed very little variation. We next calculated the variance inflation factors based on the population properties of individual cells themselves. We found values lower than 2, well below values where multicollinearity might cause problems in linear regression. Importantly, it is the high density of our dataset that allows the least squares solution to accurately identify the model parameters corresponding to individual population properties. We therefore conclude that multicollinearity did not pose a problem, and that the linear regression models are suited for parameter estimation.

Variance analysis

To compare the amount of variation explained by the population context with other sources of variation, we calculated the F-statistic, or ratio of variances, as follows: the variance between average cellular activity values per bin, over the average variance for cells of the same bins. Since we scored infection as a binary process, we approximated the within-bin variance as the variance observed over the measured activity level for the same bin between 4 to 9 replicate experiments. Note that the variance was normalised by $n - 1$, where n is the sample size, for unbiased estimation of the variance for small sample sizes. For the ChTx B and Tfn experiments the within bin variance could be calculated as the variance over the single cell intensity readouts within a single bin. The number of bins available for variance calculation of each assay is reported in Supplementary Table 1. The variance-ratios are expressed in the main text figure 2a as the percentage: $100 \frac{\sigma_{\text{between}}^2}{\sigma_{\text{between}}^2 + \hat{\sigma}_{\text{within}}^2}$.

Supplementary Table 1 | Details of the variance ratio calculations

	DV	SV40	RV	MHV	Tfn	ChTxb
Variance ratio	6.996	3.8235	1.8606	0.6311	0.5473	0.4575
Average of random. Ctrl.	0.101365	0.102345	0.14203	0.135095	0.007087	0.001573
Standard deviation of random. Ctrl.	0.008935	0.006206	0.008399	0.007906	0.000483	0.000777
number of bins	501	994	847	775	603	23
minimal number of cells per bin	100	100	100	100	50	50

Prediction images

To be able to visually compare model predicted and measured cellular activity levels, we produced synthetic images displaying the predicted cellular activity in natural cellular contexts. Example images of infection and Tfn uptake were selected, and the model predicted activity levels were calculated based on the corresponding models and on the population properties of the nuclei present in the images. Nuclei-objects detected by CellProfiler on the DAPI stain were colour coded with the corresponding cellular activity levels. All patterns resembled the activity levels of the corresponding assays. We realised however that the example images of infection all showed more striking patterns than measured by the model on the complete datasets. For all infection assays, we therefore rescaled infection profiles of the images to 1 and 0. We next applied a vertical shift of the predicted infection levels such that the overall predicted infection matches the original overall infection level measured on the example images. Next, nuclei were randomly chosen to be infected based on the adjusted infection levels for the corresponding images. Resultant images are shown in figure 1d. For Tfn, activity levels were colour-coded and compared to measured Tfn activity directly.

Bayesian network learning

Bayesian network learning was performed on the MCF10A data using the Bayesian Network Toolbox developed in MATLAB by Kevin Murphy (<http://www.cs.ubc.ca/~murphyk/Software/BNT/bnt.html>) and the BNT Structure Learning Package developed by Philippe Leray and Olivier Francois (http://bnt.insa-rouen.fr/programmes/BNT_StructureLearning_v1.3.pdf). Single cell measurements of SV40 infection, GM1 levels, FAK phosphorylation, and the population context parameters local cell density, population size, cell islet edges and cell size were discretized maximizing the Akaike Information Criterion, as implemented in the Structure Learning Package. Bootstrapped Bayesian network learning was performed using a greedy search algorithm in the Markov equivalent space¹³. The recovery rates, i.e. the number of times an edge between two nodes was inferred by the structure learning over 100 bootstrap runs on randomly selected 75% of the full single cell data set, are shown in supplementary figure 5. Note that the Bayesian network results confirm the results obtained using the graphical Gaussian modelling approach applied in supplementary figure 1 on HeLa cells.

References

1. Opgen-Rhein, R. & Strimmer, K. From correlation to causation networks: a simple approximate learning algorithm and its application to high-dimensional plant gene expression data. *BMC Syst Biol* 1, 37 (2007).
2. Steigemann, P. et al. Aurora B-mediated abscission checkpoint protects against tetraploidization. *Cell* 136, 473-84 (2009).
3. Verheijen, M. H. et al. Mouse hepatitis coronavirus RNA replication depends on GBF1-mediated ARF1 activation. *PLoS Pathog* 4, e1000088 (2008).
4. Pelkmans, L., Burli, T., Zerial, M. & Helenius, A. Caveolin-stabilized membrane domains as multifunctional transport and sorting devices in endocytic membrane traffic. *Cell* 118, 767-80 (2004).
5. Sacher, R. et al. In preparation. (2009).
6. Lozach, P. Y. et al. Dendritic cell-specific intercellular adhesion molecule 3-grabbing non-integrin (DC-SIGN)-mediated enhancement of dengue virus infection is independent of DC-SIGN internalization signals. *J Biol Chem* 280, 23698-708 (2005).
7. Pelkmans, L. et al. Genome-wide analysis of human kinases in clathrin- and caveolae/raft-mediated endocytosis. *Nature* 436, 78-86 (2005).
8. Carpenter, A. E. et al. CellProfiler: image analysis software for identifying and quantifying cell phenotypes. *Genome Biol* 7, R100 (2006).
9. Wilson, E. B. *The karyoplasmic ratio* (The Macmillan Company, New York, 1925).
10. Huber, M. D. & Gerace, L. The size-wise nucleus: nuclear volume control in eukaryotes. *J Cell Biol* 179, 583-4 (2007).
11. Neumann, F. R. & Nurse, P. Nuclear size control in fission yeast. *J Cell Biol* 179, 593-600 (2007).
12. Schlesinger, M. I. & Hlavac, V. *Ten lectures on statistical and structural pattern recognition* (Kluwer Academic Publishers, Dordrecht/Boston/London, 2002).
13. Chickering, D. M. Optimal structure identification with greedy search. *Journal of Machine Learning Research* 3, 507-554 (2002b).