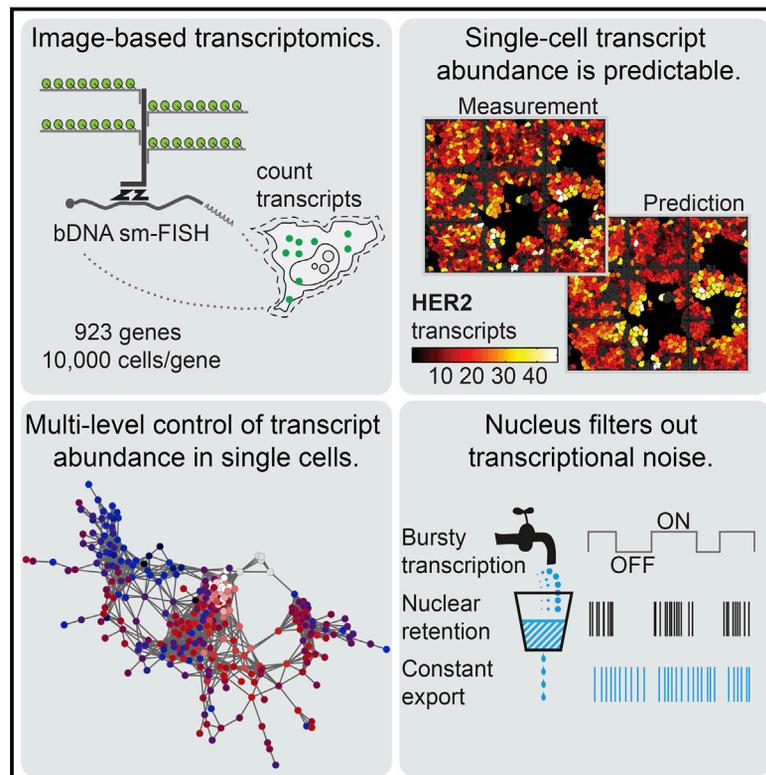


# Control of Transcript Variability in Single Mammalian Cells

## Graphical Abstract



## Authors

Nico Battich, Thomas Stoeger,  
Lucas Pelkmans

## Correspondence

lucas.pelkmans@imls.uzh.ch

## In Brief

Variability between transcript levels in genetically identical cells is not random, but stems from tight adaptation to multiple properties at the single-cell level, including the microenvironment. Nuclear transport operates as a buffer to prevent random fluctuations in transcript levels from being transmitted to the cytoplasm.

## Highlights

- Cytoplasmic transcript abundance can be predicted in single mammalian cells
- Variability in cytoplasmic transcript abundance is only minimally stochastic
- Variability is determined by multilevel transcript homeostasis in single cells
- Nuclear compartmentalization acts as a global buffer of noise in transcription



# Control of Transcript Variability in Single Mammalian Cells

Nico Battich,<sup>1,2,3</sup> Thomas Stoeger,<sup>1,2,3</sup> and Lucas Pelkmans<sup>1,\*</sup>

<sup>1</sup>Faculty of Sciences, Institute of Molecular Life Sciences, University of Zurich, 8006 Zurich, Switzerland

<sup>2</sup>Systems Biology PhD Program, Life Science Zurich Graduate School, ETH Zurich and University of Zurich, 8057 Zurich, Switzerland

<sup>3</sup>Co-first author

\*Correspondence: [lucas.pelkmans@imls.uzh.ch](mailto:lucas.pelkmans@imls.uzh.ch)

<http://dx.doi.org/10.1016/j.cell.2015.11.018>

## SUMMARY

A central question in biology is whether variability between genetically identical cells exposed to the same culture conditions is largely stochastic or deterministic. Using image-based transcriptomics in millions of single human cells, we find that while variability of cytoplasmic transcript abundance is large, it is for most genes minimally stochastic and can be predicted with multivariate models of the phenotypic state and population context of single cells. Computational multiplexing of these predictive signatures across hundreds of genes revealed a complex regulatory system that controls the observed variability of transcript abundance between individual cells. Mathematical modeling and experimental validation show that nuclear retention and transport of transcripts between the nucleus and the cytoplasm is central to buffering stochastic transcriptional fluctuations in mammalian gene expression. Our work indicates that cellular compartmentalization confines transcriptional noise to the nucleus, thereby preventing it from interfering with the control of single-cell transcript abundance in the cytoplasm.

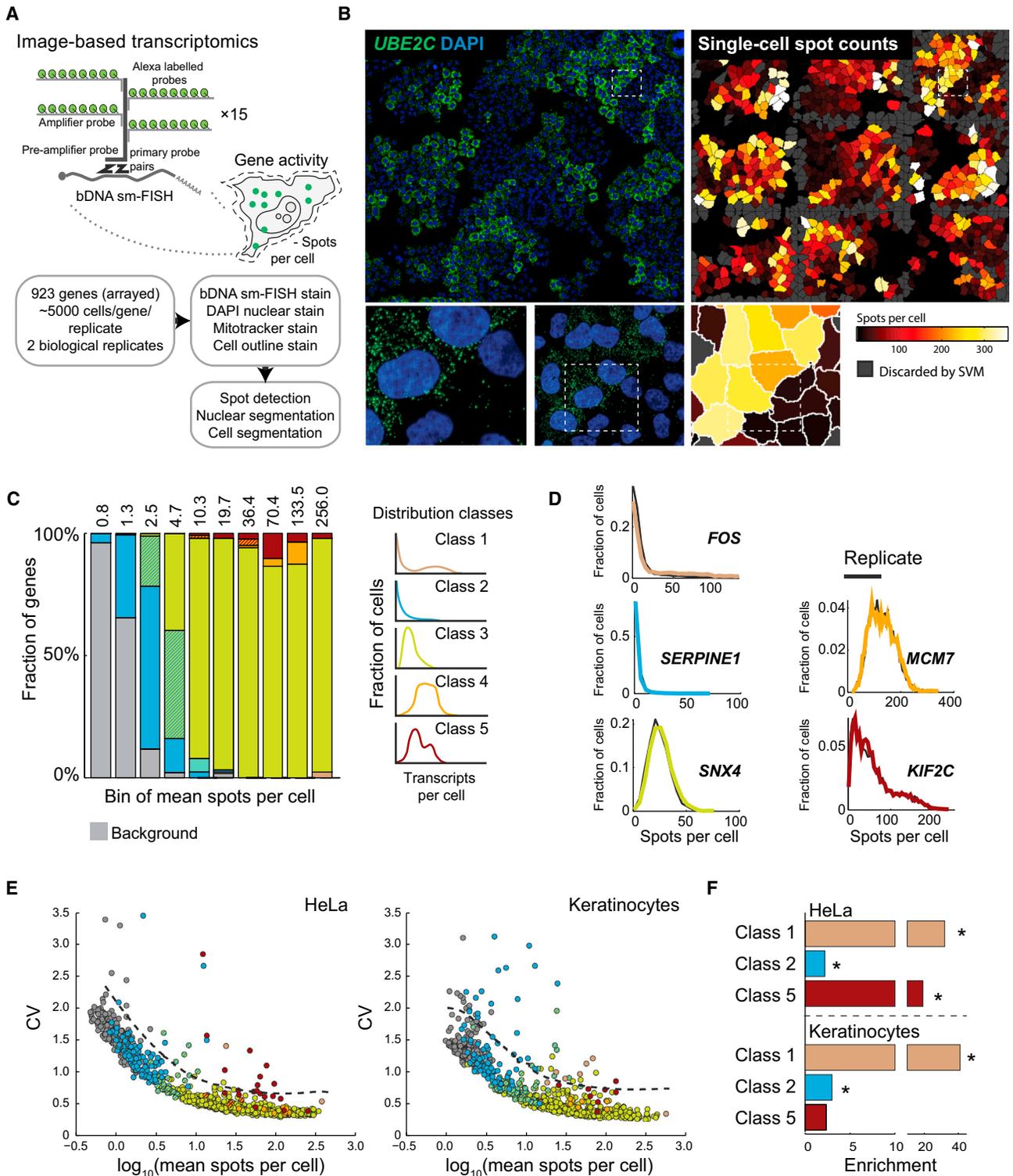
## INTRODUCTION

Gene expression in isogenic cells exposed to the same conditions is heterogeneous, a phenomenon referred to as gene expression noise (Eldar and Elowitz, 2010; Raj and van Oudenaarden, 2008). The origin of this noise can be divided between intrinsic and extrinsic sources (Elowitz et al., 2002). Intrinsic noise is seen as the inherent consequence of stochastic fluctuations in biochemical reactions and interactions between the components that transcribe and translate genes into mRNA and proteins, respectively (Eldar and Elowitz, 2010; Raj and van Oudenaarden, 2008). For instance, stochastic switching of promoters between a closed, transcription-prohibiting state and an open, permissive state can lead to bursts in transcription and consequently large variations in transcript abundance between individual cells (Golding et al., 2005; Raj et al., 2006; Suter et al., 2011; Zenklusen et al., 2008). Extrinsic noise is defined as noise that originates from upstream variations in the cellular state

that results in higher or lower rates of gene expression or degradation and is usually the major source of cell-to-cell variability (Raser and O'Shea, 2005; Snijder and Pelkmans, 2011). Extrinsic noise is not necessarily of a stochastic nature, but is often considered and modeled stochastically given the complexity of the involved processes, an apparent stochasticity in distributions of single-cell measurements, and an assumed inability to predict these variations at the single-cell level.

In human cells, transcript abundance scales with cellular volume (Kempe et al., 2015; Padovan-Merhar et al., 2015), which varies between single cells of the same population. Cellular volume is thus an important source of extrinsic noise in gene expression, as has been observed previously in yeast (Newman et al., 2006; Raser and O'Shea, 2004). Similarly, mitochondrial content, which is known to vary between individual mammalian cells, is a source of extrinsic noise (das Neves et al., 2010). In proliferating mammalian cells that adapt to their multicellular context, cell-to-cell variability in these and other properties is strongly influenced by the available space to expand cell surface and volume, the relative location of a cell within a population, its local crowdedness, the amount and type of physical force it experiences, the extent to which it faces empty space, and its position in the cell cycle (Dupont et al., 2011; Engler et al., 2006; Frechin et al., 2015; Gut et al., 2015). Since numerous signaling pathways that sense the cellular state and phenotypic properties of single cells and their microenvironment exist, this can result in large-scale adaptation of the transcriptome in single isogenic cells experiencing the same culture conditions. This raises the question of to which extent variability in transcript abundance in mammalian cells is of a deterministic nature and can be predicted once the relevant variables of single cells that drive such adaptation are known. Particularly in the context of development and tissue homeostasis, where tight control of gene expression at the single-cell level is required, such variables could influence cell-fate decisions that may have previously been considered fully stochastic (Arias and Hayward, 2006; Graf and Stadtfeld, 2008; Macarthur et al., 2009). Furthermore, if most variability in transcript abundance in mammalian cells can be predicted, it raises the question of how stochastic fluctuations that arise during transcription are effectively filtered out while deterministic variability is maintained.

Addressing these questions requires highly accurate measurements of single-cell transcript abundance. A suboptimal efficiency in detecting an individual transcript molecule in a single cell yields for most transcripts single-cell distributions that are



**Figure 1. Image-Based Transcriptomics of Cell-to-Cell Variability in Cytoplasmic Transcript Abundance**

(A) Scheme of in situ detection of single transcript molecules (spots per cell).

(B) Left: a HeLa cell population stained for cytoplasmic UBE2C transcripts (bDNA sm-FISH in green). Right: visualization of the quantified cytoplasmic transcript abundance (spots per cell) by pseudo-coloring single-cell segmentations. Cells discarded by machine learning (SVM) are gray.

(C) Classification of single-cell distributions of cytoplasmic transcript abundance in HeLa cells. Genes are binned by their mean spot number per cell. Hatched pattern indicates occurrence of class 2 and class 3 in different subsamples of the observed distributions.

(legend continued on next page)

largely determined by random detection error (Shapiro et al., 2013). Since single-cell RNA sequencing has detection efficiencies between 5%–20% (Deng et al., 2014; Grün et al., 2014), it cannot be used for sensitive analysis of sources of cell-to-cell variability in transcript abundance. Equally important for obtaining highly accurate measurements for large numbers of single cells is to avoid sampling bias of the cellular states and microenvironments experienced by single cells in a population (Battich et al., 2013). Furthermore, it is essential to quantify features of the cellular state and microenvironment of the same single cell in which transcript abundance is being measured. Finally, such measurements are ideally obtained for a large number of genes to compare distributions and identify common and gene-specific variables that determine cell-to-cell variability in transcript abundance.

Here, we applied image-based transcriptomics, a high-throughput automated single-molecule fluorescence in situ hybridization (sm-FISH) method that we recently developed (Battich et al., 2013), which meets these requirements. Using large-scale single-cell datasets acquired with this approach, we show that cell-to-cell variability in cytoplasmic transcript abundance in human adherent cells can be accurately predicted at the single-cell level with a multivariate set of features that quantify properties of the cellular state and microenvironment, and we experimentally verify some of the underlying causality. We find that for most genes, the unexplained variability in cytoplasmic transcript abundance approaches a limit of minimal stochasticity imposed by a Poisson process. The few genes that deviate from this limit also show a high amount of explained variability, suggesting high-level regulation rather than high stochasticity. Through computational multiplexing, we uncover the existence of multilevel transcript homeostasis in single cells to achieve specific adaptation of transcript abundance to the cellular state and microenvironment, according to function of the proteins they encode. Finally, we show that the mammalian nucleus acts as a potent and global buffer to stochastic fluctuations arising from bursts in gene transcription by temporally retaining transcripts in the nucleus. This explains how cytoplasmic transcript abundance in mammalian cells can be minimally stochastic, while deterministic variation is maintained.

## RESULTS

### Single-Cell Distributions of Cytoplasmic Transcript Abundance in a Human Cancer-Derived Cell Line and Primary Keratinocytes

To study cell-to-cell variability of transcript abundance in human cells, we applied image-based transcriptomics to HeLa cells and freshly isolated primary keratinocytes. This provides high-quality images of large numbers of single cells in which each transcript is visible as a bright spot that can be robustly detected, resolved

from other spots, and assigned to the corresponding cell using fully automated computer vision algorithms (Battich et al., 2013; Stoeger et al., 2015) (Figure 1A). As a result, accurate and reproducible transcript counts in the cytoplasm of millions of single cells and thousands of genes are obtained (Battich et al., 2013). When visualized across cell populations, this reveals gene-specific patterns in single cells (Figures 1B and S1A).

In both cell types, we identified, in a semi-automated manner, five classes of single-cell distributions of cytoplasmic transcript abundance across 932 genes (see <http://image-based-transcriptomics.org>). The vast majority of genes show a unimodal distribution (Figures 1C, 1D, and S1B), which shifts from a one-tailed distribution (class 2) to a skewed two-tailed distribution (class 3) as the mean cytoplasmic transcript abundance increased. Genes displaying a skewed two-tailed distribution with a flattened peak (class 4) enrich for genes acting during the replication of DNA (8 of 9 genes in HeLa and 5 of 7 genes in keratinocytes). Rarely (1.6% in HeLa and 2.8% in keratinocytes), bimodal distributions were observed, with either one mode representing no expression and the other mode expression (class 1), or with both modes representing two different levels of expression (class 5). This low fraction of bimodality compared to stimulated dendritic cells (Shalek et al., 2013) likely results from the fact that cells were unperturbed and did not experience a sudden change in culture conditions (e.g., addition of growth factor after serum starvation). Concordantly, the majority of the genes that show bimodal distributions under these culture conditions act during the M phase of the cell cycle (60.0% of HeLa class 5, 71.4% of keratinocytes class 1, and 50.0% of keratinocytes class 5). We also noticed that the coefficient of variation (CV) in single-cell transcript abundance decreased in both cell types monotonically from ~2 to ~0.3 as the mean transcript abundance increased, with only a few outlier genes (3%–6%) that show a higher CV than the bulk (Figure 1E). Expectedly, these outliers are enriched in the one-tailed and bimodal distributions of cytoplasmic transcript abundance (classes 1, 2, and 5) (Figure 1F).

### Cytoplasmic Transcript Abundance in Single Human Cells Can Be Predicted and Is Minimally Stochastic

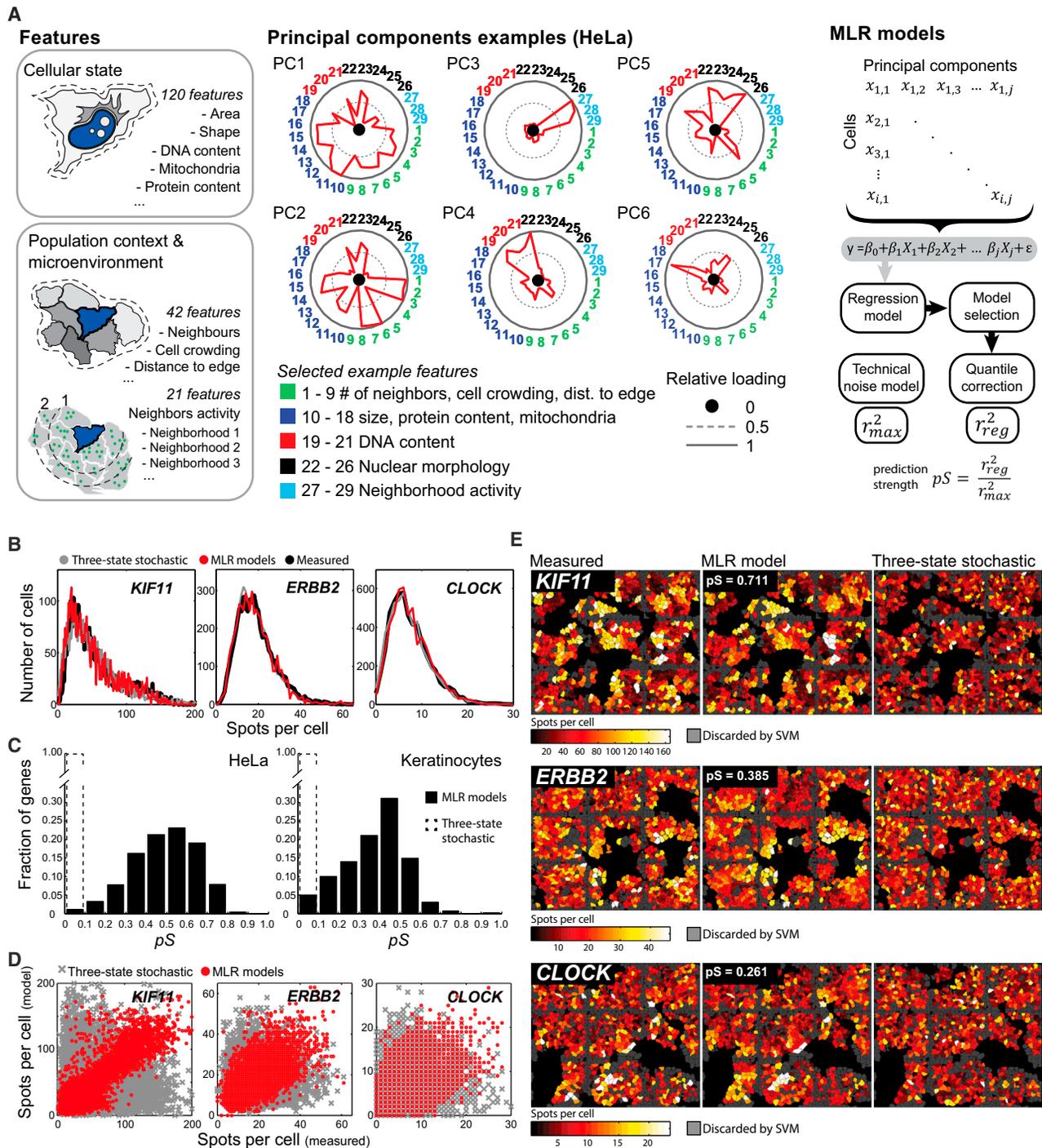
In addition, we collected from each single cell a multivariate set of 183 features that quantify properties of cell and nucleus shape and area, of protein, DNA, and mitochondrial content and texture, and of the extent of local cell crowding, number of neighbors, and relative location to other cells and to empty space in the cell population (Figure 2A). For genes that are expressed (Figure S2A), we observed that many of these features show a correlation with transcript abundance (Figure S2B), prompting us to investigate the extent to which these features can collectively predict cytoplasmic transcript abundance in single cells. To address this, we learnt data-driven models for each gene

(D) Gene examples with single-cell cytoplasmic transcript abundance distributions. For more examples, see <http://image-based-transcriptomics.org>.

(E) Coefficient of variation per gene versus cytoplasmic transcript abundance (mean spots per cell), colored according to their distribution class as in (B). The dashed line defines outliers exceeding one SD of a LOESS fit.

(F) Enrichment for cytoplasmic transcript abundance distribution classes among outlier genes over non-outlier genes. Asterisks indicate Fisher's exact test below 0.05.

See also Figure S1.



**Figure 2. Predicting Cytoplasmic Transcript Abundance in Single Cells within a Population**

(A) Features describing cellular state, population context, and microenvironment of single cells (right), the loadings of the first six principal components (middle), the construction of multilinear regression (MLR) models, and the calculation of prediction strength ( $pS$ ).

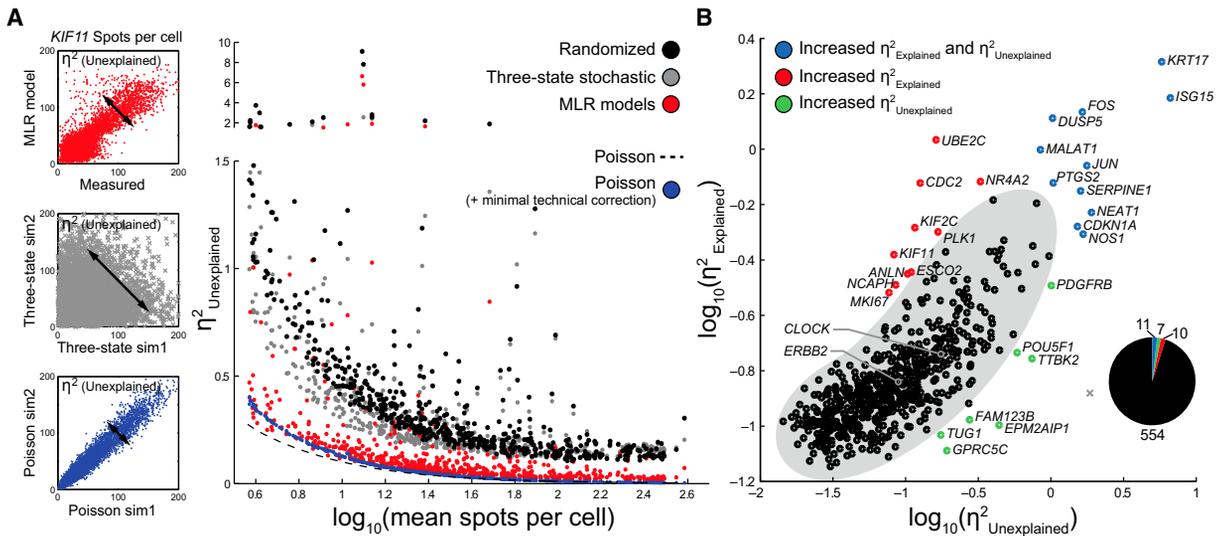
(B) Prediction of single-cell transcript distributions of *KIF11*, *ERBB2*, and *CLOCK* in HeLa cells by MLR models and three-state stochastic models.

(C) Distribution of prediction strengths ( $pS$ ) for 583–598 genes using MLR models (black filled bars) and three-state stochastic models (dashed open bars). The size of each bin is 0.1.

(D) Prediction of *KIF11*, *ERBB2*, and *CLOCK* cytoplasmic transcript abundance in single HeLa cells by MLR models and three-state stochastic models.

(E) Visualization of measured and predicted single-cell cytoplasmic transcript abundance within a population of HeLa cells.

See also Figure S2.



**Figure 3. Cell-to-Cell Variability in Cytoplasmic Transcript Abundance Contains Minimal Stochastic Variability**

(A) Comparison of unexplained variability ( $\eta^2_{\text{Unexplained}}$ ) of MLR models (red), three-state stochastic models (light gray), randomized data (black), and Poisson limit without (dark blue) and with correction for hybridization efficiency (light blue) in HeLa cells.

(B) Correlation between explained variability ( $\eta^2_{\text{Explained}}$ ) and unexplained variability ( $\eta^2_{\text{Unexplained}}$ ) for single genes in HeLa cells. Gray area shows 90% confidence interval of a Gaussian mixture model of  $\eta^2_{\text{Explained}}$  and  $\eta^2_{\text{Unexplained}}$  of all genes. Blue genes show both high explained and unexplained variability and are enriched in immediate early response genes. Red genes have higher explained than unexplained variability and are enriched in cell-cycle genes. Green genes are the lowest abundant genes, with a spot count per cell barely above background (3–4 spots per cell).

See also Figure S3.

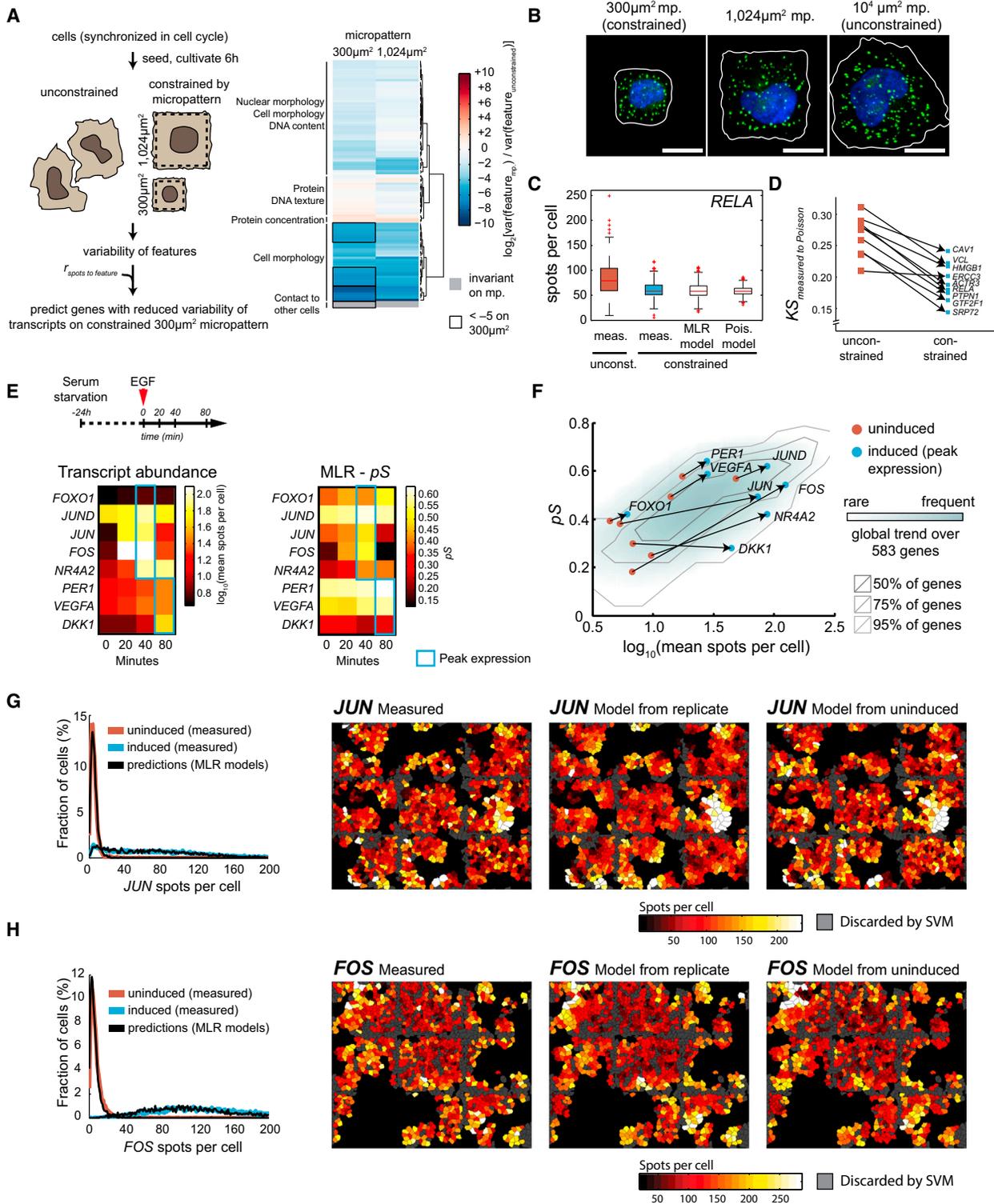
on one dataset using multilinear regression (MLR) in a principal component (PC)-reduced multidimensional space of the multivariate feature set (Figure 2A). Generally, MLR models consisted of  $\sim 20$  PCs, which quantify a variety of different aspects of individual cells. For example, the first six PCs of HeLa cells consist of features of local cell crowding, distance of cells to each other, their number of neighbors, distance to a cell islet edge, cell and nuclear area, cell volume (as measured by protein content, see Figure S2C), mitochondrial content, DNA content (indicating position in the cell cycle), nuclear morphology, cell shape, and the activity (transcript abundance) of neighboring cells (Figure 2A). In keratinocytes, the first six PCs are highly comparable (Figure S2D). Higher PCs used in the MLR models often contain highly specific properties related to cell shape, texture, or micro-environment (data not shown).

Next, we tested the performance of each MLR model by directly predicting cytoplasmic transcript abundance in each single cell of an independently obtained ( $\sim 3$  weeks later) biological replicate dataset for the same gene and comparing single-cell predictions with single-cell measurements. The models accurately reproduced single-cell distributions of cytoplasmic transcript abundance (Figures 2B, S2E, and S2F). More importantly, they also achieved high prediction strength ( $pS$ ; coefficient of determination corrected for technical variability, see the Supplemental Experimental Procedures) at the single-cell level (Figures 2C–2E and S2G). The median  $pS$  was slightly higher in monoclonal HeLa cells (0.503) than in freshly isolated primary keratinocytes (0.400), possibly due to uncontrolled clonality of the latter cells. A partial least-squares regression, as well as a non-linear approach using random forests (Liaw and

Wiener, 2002) on the non-transformed multivariate feature set, achieved virtually identical results (Figure S3A), indicating the robustness of these statistical models.

The  $pS$  increased as mean cytoplasmic transcript abundance increased, with a median  $pS$  of 0.29 for low-abundant transcripts (3.7–7.4 mean transcripts per cell) and a median  $pS$  of 0.71 for high-abundant transcripts ( $>149$  mean transcripts per cell; see Figure S2H). Furthermore, as the examples of *KIF11* (a kinesin involved in spindle formation and chromosome positioning during mitosis), *ERBB2* (a receptor tyrosine kinase that dimerizes with epidermal growth factor receptors), and *CLOCK* (a transcription factor that regulates circadian rhythms) show, the MLR models predict the observed patterns of single-cell expression in cell populations remarkably well, even for low-abundant transcripts (Figures 2D, 2E, and S2G). Naturally, three-state stochastic models of transcription can only reproduce distributions (Figures 2B, S2E, and S2F) and do not have any single-cell prediction strength (Figures 2C–2E and S2G), nor can they reproduce single-cell expression patterns in cell populations (Figures 2E and S2F).

Strikingly, when we quantified the amount of unexplained variability in cytoplasmic transcript abundance (see the Supplemental Experimental Procedures), we observed that it approaches a limit of minimal stochasticity as described by a simple one-step Poisson process, also for low-abundance transcripts (Figures 3A and S3A). Although some genes did not fall on this limit, we observed that genes whose unexplained variability was furthest away from the Poisson limit, also displayed the highest amount of explained variability (outliers of both increased  $\eta^2_{\text{Explained}}$  and  $\eta^2_{\text{Unexplained}}$ ) (Figures 3B and S3B). This



**Figure 4. Causality between Predictors and Single-Cell Transcript Abundance**

(A) Growing single cells on micropatterns (left) constrains the variance of phenotypic features (right). Boxed features have the strongest reduction of variance. (B) Images show *RELA* transcripts (green) and DAPI (blue) of single constrained cells grown on differently sized micropatterns and an unconstrained cell grown on a 10,000-µm<sup>2</sup> micropattern. Segmented cell outlines are white lines. Scale bar, 16 µm.

(legend continued on next page)

shows that also for these genes, cell-to-cell variability in cytoplasmic transcript abundance originates largely from regulatory processes rather than from intrinsic stochastic sources (Figure 3B).

Thus, cytoplasmic transcript abundance of genes can be accurately predicted at the single-cell level in mammalian adherent cells, both in a cancer-derived laboratory-adapted cell line and in primary cells freshly isolated from a human donor. Single-cell prediction is achieved with features that quantify a variety of different aspects of the phenotypic state of individual cells, their population context, and their microenvironment. The amount of cell-to-cell variability that these features cannot predict approaches for most genes a single-step Poisson limit. This suggests that somewhere along the life of an RNA molecule in mammalian cells, noise buffering occurs to ensure that cytoplasmic transcript abundance becomes minimally stochastic.

### Causality between Predictors and Single-Cell Transcript Abundance

High prediction strength indicates a high correlation between predictors and single-cell transcript abundance, but does not reveal the presence or direction of causality. To reveal the dominant direction of causality in this situation, we used four orthogonal approaches.

First, we applied Bayesian network inference on the initial datasets and on bulk nascent transcript synthesis measurements in single cells (Figures S4A–S4C). For 83% of the genes in which Bayesian networks could reproducibly be inferred, cytoplasmic transcript abundance was placed downstream of one or multiple single-cell features (Figure S4A). The remaining genes (17%) were placed in between, being downstream of cell area or protein content and upstream of DNA content or cell crowding, which correlated with gene function (Figures S4B and S4C). Furthermore, this revealed that cell area and protein content (cell volume) are major determinants of bulk nascent transcript synthesis and cytoplasmic transcript abundance, which are in turn determined by population context effects that arise from the number of cells seeded and DNA content, which reflects position in the cell cycle (Figure S4D).

Second, we grew cells on micropatterns, which constrain the available area for a single cell to spread on, resulting in a strong reduction in the cell-to-cell variability of many single-cell features, particularly in cell size and morphology (Figures 4A and 4B). Based on this, we used the MLR models to predict which

genes would display the strongest reduction in variability in cytoplasmic transcript abundance in cells grown on the smallest micropatterns, and selected from these nine genes covering different biological processes (Figure 4A). For all genes, as exemplified by *RELA*, a subunit of the major transcription factor NF- $\kappa$ B, we observed that constraining the phenotypic state of single cells results in a strong reduction of cell-to-cell variability in cytoplasmic transcript abundance, approaching a Poisson limit (Figures 4C and 4D). Strikingly, the small amount of remaining cell-to-cell variability in transcript abundance was accurately predicted based on the small amount of phenotypic variability remaining between single cells grown on micropatterns (Figure 4C). This shows that constraining single-cell features directly constrains cell-to-cell variability in transcript abundance.

Third, we performed systematic RNA interference against 367 genes. This did not reveal any relationship between the extent to which two dominant features, nuclear area and cell crowding, correlate with cytoplasmic transcript abundance of a gene and the effect that silencing of this gene had on these two features (Figure S4E). The few genes whose silencing resulted in strong effects were all essential for cell viability, leading to reduced population sizes (Figure 4D), which indirectly changes nuclear area and cell crowding.

Fourth, we performed gene induction experiments. Cells grown for 72 hr were serum starved for 24 hr and subsequently treated with epidermal growth factor (EGF). At 20, 40, and 80 min after induction, we fixed cells and performed image-based transcriptomics on eight genes induced by EGF. We then learnt MLR models on each time point after induction, as well as on the serum-starved non-induced state, and used these to predict single-cell cytoplasmic transcript abundance in a replicate experiment (Figure 4E). While  $pS$  was lower in the non-induced state or in the presence of serum, it was higher at peak expression level following induction, matching the global trend that  $pS$  scales with transcript abundance (Figures 4F and S4F). As shown for *JUN* and *FOS*, two immediate early response genes, the MLR models accurately reproduced the change in distributions of cytoplasmic transcript abundance during induction, including the emergence of bimodality, as well as single-cell patterns of EGF-induced gene expression in cell populations (Figures 4G, 4H, and S4G). Strikingly, even MLR models learnt on serum-starved non-induced cells were able to predict the single-cell expression patterns in induced cell populations, when correcting for the difference in mean expression levels

(C) Boxplots show measured single-cell spot count distributions of *RELA* transcripts in unconstrained cells and constrained on 300- $\mu\text{m}^2$  micropatterns ( $n = 1,874$  and 694, respectively). Spot count distribution in constrained cells predicted by the MLR model from unconstrained cells and a distribution arising only from Poisson noise is also shown.

(D) Kolmogorov-Smirnov distance ( $KS$ ) between measured single-cell transcript distributions and Poisson distributions ( $n = 10,000$ ) of nine different genes in unconstrained and constrained cells.

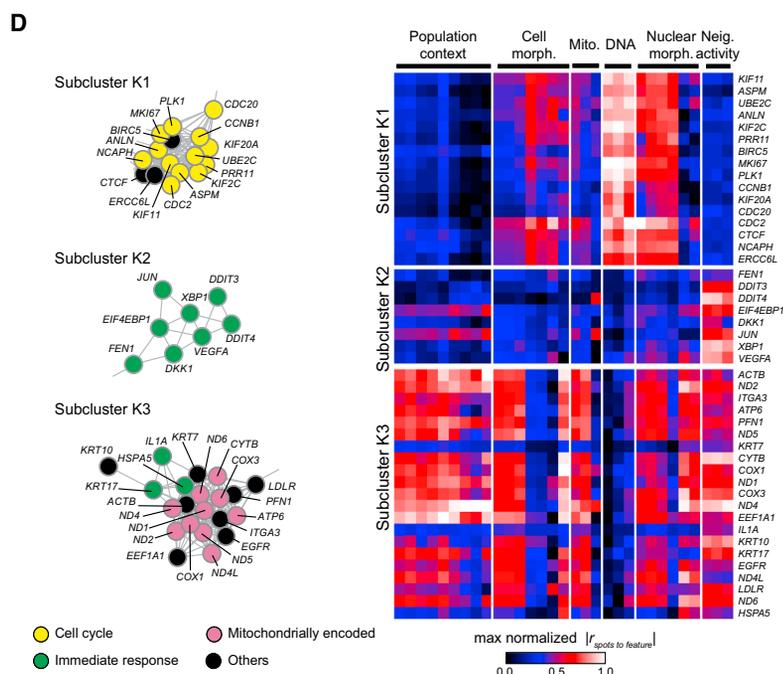
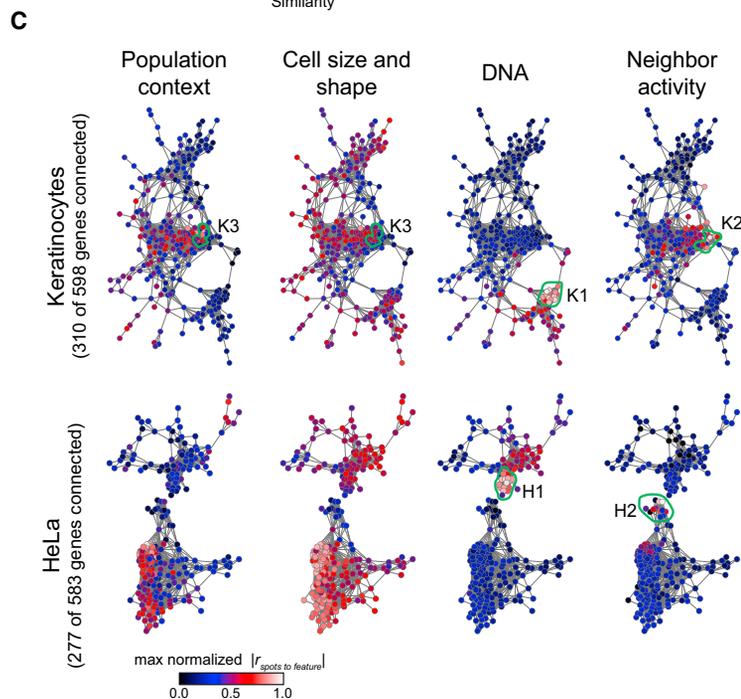
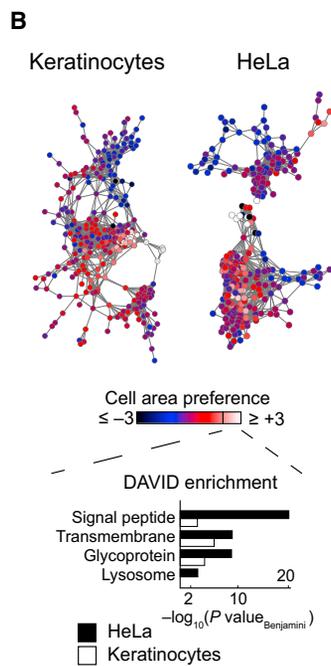
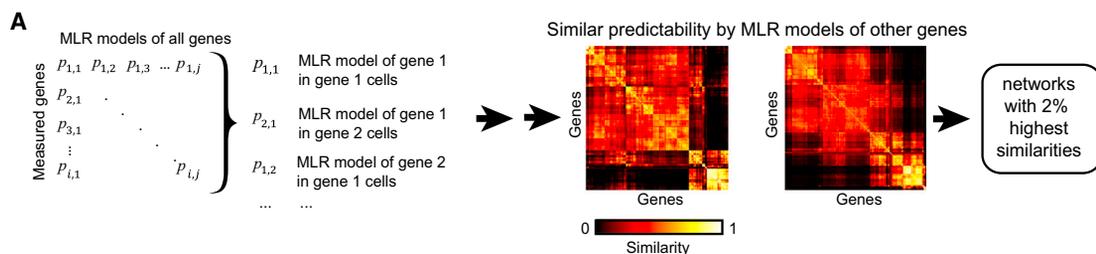
(E) Heatmaps of mean cytoplasmic transcript abundance at various time points after serum starvation and addition of EGF and of the  $pS$  of MLR models. Blue boxes highlight the highest observed mean cytoplasmic transcript abundance.

(F) The  $pS$  in EGF-induced cells at peak expression (blue dots) is higher than in uninduced cells (red dots), as expected from the increase in transcript abundance (gray-shaded contoured area).

(G) Left: single-cell cytoplasmic transcript abundance distributions for *JUN* in uninduced cells (red) and in cells 40 min after EGF induction (blue). Distributions are predicted with MLR models (black). Right: prediction of EGF-induced cytoplasmic transcript abundance of *JUN* at the single-cell level with MLR models learnt from a replicate experiment or from uninduced cells.

(H) As in (G), except for *FOS* transcripts.

See also Figure S4.



(legend on next page)

(Figures 4G and 4H). This shows that it is largely the predetermined phenotypic state and microenvironment of a single cell that determines its response to EGF.

Together, this shows that in human adherent cells grown in culture, the emergence of heterogeneity in phenotypic state, population context, and microenvironment of single cells is the dominant source of cell-to-cell variability in cytoplasmic transcript abundance, making it for most genes largely predictable. This does not only apply to cells at quasi steady-state continuously grown in serum but also, and more profoundly, during an acute induction of gene expression by EGF, also when this leads to bimodal gene expression.

### Computational Multiplexing of Cytoplasmic Transcript Abundance Reveals Multilevel Transcript Homeostasis in Single Cells

Next, we studied the biological information that the MLR models contain, taking advantage of their generally high prediction strength at the single-cell level. This allowed us to perform computational multiplexing (Figure 5A), in which we predicted the transcript abundance of one gene in each cell of a population in which we had measured the transcript abundance of another gene. In this manner, we could calculate pairwise correlations between the predicted and measured single-cell transcript abundances for  $\sim 2.5 \times 10^5$  gene-gene combinations across  $\sim 5,000$  single cells (Figures 5A and S5A). We then calculated the similarity between two genes in their pairwise single-cell correlations with all other genes and created a similarity matrix for each cell type. The matrices contained a high degree of modality with various sub-clusters (Figure 5A; Table S1), presenting a systems-level map of single-cell transcript homeostasis in human adherent cell populations (Figures S5B and S5C).

To visualize this map, we created a gene interaction network for each cell type, in which two genes are connected when they were within the top 2% highest similarity scores (Figures 5A, 5B, and S5A). To reveal patterns in these networks, we first looked at the two most dominant predictors: cell area and cell volume. Plotting the ratio of the correlation of cytoplasmic transcript abundance with these two predictors on the networks revealed areas of genes with a higher correlation to cell volume or higher correlation to cell area (Figures 5B and S5D; Data S1). The latter are strongly enriched in genes encoding for proteins that contain a signal peptide, a transmembrane domain, or that are N-glycosylated (Figure 5B), as well as for cytosolic proteins with important membrane-related functions (data not shown). This indicates the existence of mechanisms that allow distinct adaptation of transcript abundance to the volume or surface

area of a single cell, depending on whether it encodes for a protein with cytosolic or membrane-related functions.

Next, we plotted on the networks the mean absolute correlations of transcript abundance to selected sets of features related to the population context, to cell size and shape, DNA content, and neighbor activity (Figure 5C). This revealed multiple sub-regions in the networks that consist of groups of genes whose cytoplasmic transcript abundance is adapted in specific ways to different combinations of features. For example, a particularly outstanding sub-cluster present in both networks (K1 in keratinocytes and H1 in HeLa; see Figures 5D and S5E) shows high correlations with features of DNA content and texture and nuclear morphology and is enriched in genes that function in the cell cycle.

We also noticed that within dense regions of the networks, highly differentiated and specific adaptation is visible. For instance, sub-clusters K2 and K3 lie next to each other in the keratinocyte network (Figure 5C). Sub-cluster K2 consists of genes whose transcript abundance shows a specific and strong correlation with neighbor activity and contains immediate early genes (e.g., *JUN*), including secreted molecules (e.g., *VEGFA* and *DKK1*) (Figure 5D). A similar sub-cluster was also found in HeLa cells (H2; see Figure S5E). H2 contains genes that display high levels of both explained and unexplained variability (compare to Figure 3B), including the early response genes in the EGF induction experiment (such as *JUN*, *FOS*, and *NR4A2*). This indicates that both cell types show a highly variable expression of a group of genes that respond quickly to signals in a correlated manner determined by the activity of cell neighbors, suggesting the involvement of paracrine signaling (Avraham and Yarden, 2011). Sub-cluster K3 consists of genes whose single-cell transcript abundance shows strong correlation with multiple sets of selected features, including those of the population context, of cell size and shape, of mitochondrial abundance, nuclear morphology and also neighbor activity (Figure 5D). It contains 10 of the 13 mitochondrially encoded protein-coding genes, indicating that multilevel control of single-cell transcript abundance also occurs for genes not transcribed in the nucleus.

The high degree of modularity and the presence of multiple subgroups of genes whose transcript abundance is adapted in highly differentiated and specific ways in single cells exposed to the same conditions, demonstrates the existence of a complex multilevel transcript homeostasis system that drives cell-to-cell variability in gene expression. This ensures that levels of transcripts are precisely adapted to the physiological state of a single cell and its microenvironment according to the function of the RNA or the protein they encode for.

### Figure 5. Multilevel Transcript Homeostasis in Single Mammalian Cells

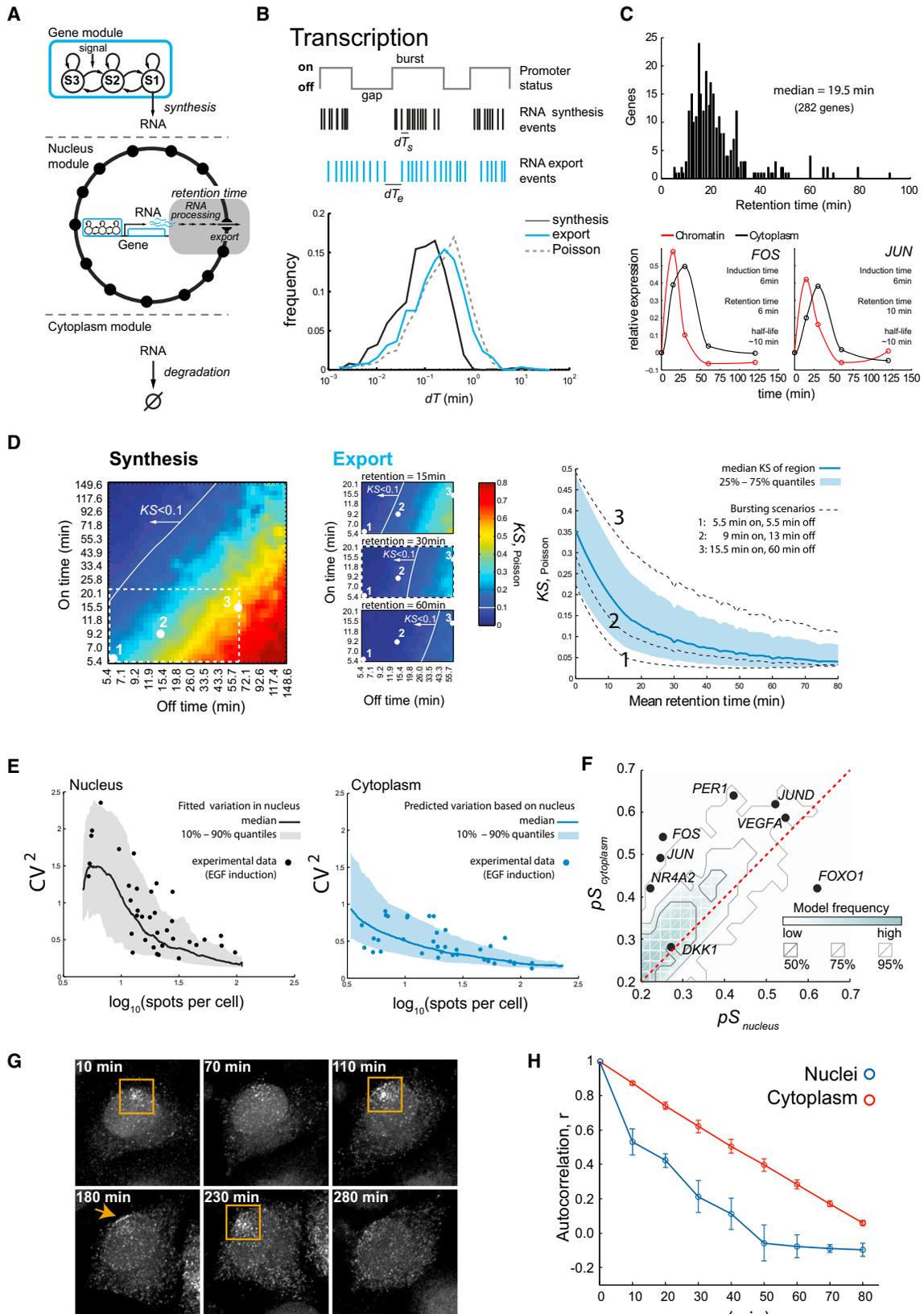
(A) Gene-specific MLR models are applied to other genes in a pairwise manner, and genes are clustered by their similarity in being predicted by the MLR models of all other genes.

(B) Networks formed by connecting 2 genes (nodes) that show the 2% highest similarities. Node color indicates genes with higher correlation of cytoplasmic transcript abundance with cell area or with cell volume. Bar graphs show functional annotation enrichments for genes with a 2-fold higher correlation with cell area than cell volume.

(C) Genes in the networks colored according to max-normalized correlation between cytoplasmic transcript abundance and sets of features. Green borders indicate sub-clusters K1-3 and H1-2.

(D) Enlargement of sub-clusters K1 and K2 present in keratinocytes. Heatmap shows max-normalized correlation between cytoplasmic transcript abundance and selected individual features. Grouping of features as in (C).

See also Figure S5, Table S1, and Data S1.



(legend on next page)

### Transcript Retention in the Nucleus and Export into the Cytoplasm Efficiently Buffers Stochastic Bursts in Gene Transcription

A high degree of predictability in cytoplasmic transcript abundance at the single-cell level contradicts the view that it arises from stochasticity in gene transcription, caused by, among others, the stochastic switching of promoters between a closed transcription-prohibiting state and an open permissive state. However, we realized that bDNA sm-FISH, unlike most methods that quantify single-cell transcript abundance, specifically detects transcripts in the cytoplasm. This suggests that random fluctuations in transcript abundance arising from bursts in transcription are filtered out during nuclear processing and/or export from the nucleus to the cytoplasm (Singh and Bokes, 2012; Xiong et al., 2010).

To test if nuclear compartmentalization can theoretically act as a buffer of transcriptional noise in mammalian cells, we developed an agent-based single-cell mathematical model and performed computer simulations (Figure 6A). In the model, gene activation and transcription is governed by a three-state stochastic model (Neuert et al., 2013; Raj and van Oudenaarden, 2008), in which the gene switches randomly between an “off” state (S3) and a transcription-competent state (S2), which switches randomly to a transcription-initiated (S1) or “on” state and back. Once transcription is initiated, RNA synthesis occurs at randomly fluctuating transcription rates. The time spent in the “on” and “off” states can be varied. Each transcript is then retained for a certain amount of time in the nucleus, after which it is transported into the cytoplasm. Nuclear retention time is used as a general term to comprise the various events between birth of a single transcript molecule and its emergence into the cytoplasm, including chromatin dissociation, nuclear diffusion, processing, and binding to and transport across the nuclear pore. It is modeled as a combination of a 3D diffusion process and a probabilistic interaction with the nuclear pore, and can be varied. Nuclear degradation of transcripts is not considered. Finally, transcript degradation in the cytoplasm is modeled as a single probabilistic function that can also be varied (Figure 6A). To quantify the effect that nuclear retention

has on the amount of stochasticity in transcript abundance in this model, time distributions between simulated transcript production events ( $dT_s$ ) and between nuclear export events ( $dT_e$ ) are obtained, and the distance of these distributions to a Poisson distribution determined (Figure 6B). Physiological boundaries for nuclear retention times of transcripts were obtained from a recently collected high-quality RNA-seq dataset on lipopolysaccharide (LPS)-induced transcription in mouse bone marrow-derived macrophages (Bhatt et al., 2012). From 282 genes, we derived a nuclear retention time of newly synthesized transcripts between ~5–90 min, with a median of ~20 min (Figure 6C). Because these genes are enriched in fast-responding genes during stress signaling in macrophages, these nuclear retention times are an underestimation for most other genes.

We performed model simulations with different burst-like gene transcription scenarios, ranging from transcription “on” times (corresponding to state S1 in the model of the gene module; Figure 6A) between 5–20 min and “off” times (corresponding to state S2 or S3 in Figure 6A) between 5–60 min, which lie in the range of observed bursting dynamics of endogenous genes in mammalian cells (Ochiai et al., 2014). Longer “on” times do not reflect burst-like gene expression and are already close to a Poisson limit during synthesis in the nucleus (Figure 6D), while longer “off” times reflect, in our opinion, non-stochastic regulation, such as refractory periods, feedbacks, or oscillating cellular states (Sanchez and Golding, 2013). Within these boundaries, variability in transcript synthesis is far away from the Poisson limit (Figure 6D). However, export of the produced transcripts into the cytoplasm was efficiently converted into a Poisson process as mean retention time increased (Figures 6D and S6A). Over all bursting scenarios, a mean nuclear retention time of 15 min was able to buffer ~57% of the stochastic fluctuations introduced by bursts, which increased to ~90% at 40 min of mean nuclear retention time. Importantly, when we modeled bursting scenarios with “on” and “off” times of ~5.5 min (scenario 1 in Figure 6D), we observed ~50% buffering already at a mean nuclear retention time of 6 min. This corresponds to the measured induction and nuclear retention times of *FOS* and

#### Figure 6. Nuclear Compartmentalization Efficiently Buffers Stochastic Bursts in Gene Transcription

(A) Mathematical model separated in gene, nucleus, and cytoplasm modules of gene transcription, transcript processing, diffusion, retention, nuclear export, and transcript degradation.

(B) The effect of nuclear retention on how variation between transcript synthesis events ( $dT_s$ ) converts to variation between transcript export events ( $dT_e$ ).

(C) Distribution of nuclear retention times of transcripts of 282 genes induced by LPS in mouse bone-marrow-derived macrophages (Bhatt et al., 2012). The lower panels display kinetics for *FOS* and *JUN* and derived  $t_{1/2}$  of gene induction, nuclear retention, and degradation.

(D) Left: Kolmogorov-Smirnov distances (KS) of  $dT_s$  (synthesis) distributions to a Poisson distribution for various “on” and “off” times. Middle: KS distances of  $dT_e$  (export) distributions to a Poisson distribution (using “on” and “off” times boxed on the left) and nuclear retention times of 15, 30, and 60 min. The white line indicates regions in which the KS distance is 0.1. Right: median KS distance of  $dT_e$  distributions to a Poisson distribution over all “on” and “off” times as a function of retention time, as well as of specific combinations indicated with white dots on the left.

(E) Model-predicted  $CV^2$  of single-cell nuclear (left, black) and cytoplasmic (right, blue) transcript abundance against mean transcript abundance. Shaded areas, interquartile range. Measured  $CV^2$  in both compartments are solid dots.

(F) MLR model prediction strengths ( $pS$ ) of measured single-cell transcript abundance in the nucleus and cytoplasm (black dots) and as predicted with the agent-based model (shaded area).

(G) Time-lapse imaging of doxycycline-induced transcription of HeLa 11ht MS2 cells (Halstead et al., 2015). Squares, transcriptional bursts; arrow, accumulation of transcripts at the nuclear envelope.

(H) Autocorrelation function of mRNA spot counts in the nucleus and cytoplasm of single cells ~1 hr after induction.  $n = 4$ , data are mean  $\pm$  SEM. at given  $\tau$ ,  $p < 0.05$  for all  $\tau > 0$  min.

See also Figure S6, Movie S1, and Data S2.

*JUN* (Figures 6C and S6B), indicating that the short retention times observed for fast-responding genes also have a noise buffering effect. Thus, for most genes, nuclear retention seems long enough to reduce stochastic variation arising from bursts in transcription, also for immediate early genes. Moreover, it may be that timescales of transcript retention in the nucleus are adapted to the rate of their induction.

To test experimentally whether nuclear retention increases the predictability of cytoplasmic transcript abundance in single mammalian cells, we adapted bDNA sm-FISH to detect transcripts in the nucleus, and performed an EGF induction experiment where we measured both nuclear and cytoplasmic transcript abundance. As expected, the increase in cytoplasmic transcript abundance of genes reacting to EGF followed with a delay the increase in nuclear transcript abundance (Figures S6B and S6C). Moreover, bursts of transcription were clearly visible in the nucleus (Figure S6C). Importantly, we found that the coefficient of variation ( $CV^2$ ) was higher in the nucleus than in the cytoplasm, in particular in cases when transcripts were less abundant. This drop in transcript variability in the cytoplasm compared to the nucleus was predicted by the model (Figure 6E). Moreover, we found that MLR models have higher prediction strength on cytoplasmic transcript abundance than on nuclear transcript abundance, (e.g., 2.5-fold higher for *FOS* and *JUN*), which is consistent with the agent-based model (Figure 6F). In addition, when we overexpressed NUP153, a nuclear pore component that upon overexpression reduces mRNA nuclear export (Bastos et al., 1996), we observed a reduction in cytoplasmic transcript variability for *JUN*, 20 min after induction with EGF (Figure S6D).

Next, we used long-term time-lapse imaging of single HeLa cells expressing an inducible transcript containing 24 bacteriophage MS2 stem loops, as well as Halo-tagged MS2 coat protein, which binds to the stem loops (Halstead et al., 2015). Time-lapse imaging carried out for 5–13 hr after gene induction showed repeated bursts of transcription in the nucleus (Figure 6G). We also observed a transient accumulation of transcripts at the inner nuclear envelope, and a delay between the increase in nuclear transcript abundance and cytoplasmic transcript abundance, both indicative of nuclear retention (Figure 6G). From the movies, we estimated that the length of bursts (“on” times) were 10–60 min, the intervals between bursts (“off” times) were 20–100 min, and nuclear retention time was ~40 min. These values are within the range of the modeled parameter space, and thus predict that cytoplasmic transcript abundance should display less stochastic variability than nuclear transcript abundance. To measure this within the same single cells, we calculated the autocorrelation in transcript abundance over time in both the nucleus and the cytoplasm. A low autocorrelation is indicative of stochastic fluctuations. Consistent with the model’s predictions, we observed that autocorrelation measurements of transcript abundance over up to 1-hr time periods are higher in the cytoplasm than in the nucleus (Figures 6H and S6E). This directly shows that during gene induction, transcript abundance shows more stochastic fluctuations over time in the nucleus than in the cytoplasm, indicative of buffering through nuclear compartmentalization and retention.

Taken together, this showed that cellular compartmentalization separating the nucleus from the cytoplasm is an efficient mechanism to dampen stochastic fluctuations arising from bursts in gene transcription for most genes. This explains how cytoplasmic transcript abundance in single cells can approach a Poisson limit of minimal stochasticity despite the occurrence of burst-like gene transcription.

## DISCUSSION

In this study, we perform highly accurate measurements of transcript abundance in large numbers of single adherent human cells with single-molecule resolution for a thousand genes using image-based transcriptomics. We combine these measurements with a multivariate set of features from the same single cells that quantify multiple properties of the cellular state, their population context, and their microenvironment. We show that multilinear regression models based on these features can predict single-cell distributions, have high prediction strength on single-cell transcript abundance, and can accurately predict single-cell expression patterns. The amount of variability not explained by multilinear regression approaches a system of minimal stochasticity given by a Poisson process. The causality underlying this high predictability stems from mechanisms by which the cellular state, the population context, and the microenvironment determine cytoplasmic transcript abundance in single cells, for which we provide a systems-level map across several hundred genes. Finally, we show that mammalian cells can achieve minimal stochasticity in cytoplasmic transcript abundance by means of nuclear compartmentalization, which, through temporally retaining transcripts in the nucleus, provides a general and potent mechanism to buffer stochastic fluctuations caused by bursts in gene transcription. An independent simultaneous study confirms that also within tissues, mammalian cells display nuclear retention of transcripts to buffer noise (Halpern et al., 2015).

Our findings pertain to virtually all of the genes analyzed in adherent human cells, both when cells are at quasi steady state in the continuous presence of serum, as well as during acute gene induction experiments after a period of serum starvation. This illustrates that even at time-scales of less than 1 hr, a differential response in the upregulation of cytoplasmic transcript abundance in single adherent mammalian cells is largely of non-stochastic origin. Only a few genes display simultaneously a high degree of explainable variability as well as a high degree of unexplainable variability. These are immediate early response genes, the transcripts of which accumulate rapidly in the cytoplasm after induction of expression, are only shortly retained in the nucleus and are subject to high cytoplasmic turnover. While this limits the nucleus’ ability to completely filter out stochastic variability caused by bursts in gene transcription for these genes, their relatively brief nuclear retention still has a sufficient noise dampening effect. Thus, while cell-to-cell variability in cytoplasmic transcript abundance in mammalian cells is often large, our findings show that the cause of this variability is not stochastic, but is determined by a multilevel system regulating transcript homeostasis in single cells.

The use of nuclear retention for noise filtering underscores the notion that mammalian cells do not rely on the induction of gene

transcription for very fast responses. For the fastest responding genes in mammalian cells, such as *FOS* and *JUN*, nuclear retention times appear adjusted to the rate of induction, short enough to minimize the delay in response, but long enough to enable efficient noise buffering.

In prokaryotes, where a nucleus is absent and RNA pre-processing is minimal, transcriptional responses can make use of co-transcriptional translation and can thus be very fast (Bird, 1995; Martin and Koonin, 2006). Also in single-cell eukaryotes, such as yeast, which show less extensive nuclear processing of transcripts and have considerably smaller nuclei, transcriptional responses may overall be somewhat faster than in mammalian cells (Kresnowati et al., 2006). This suggests that as cells acquired a nucleus during evolution and formed multicellular organisms, the increased complexity in nuclear RNA processing came with the additional benefit of filtering out stochasticity in gene transcription, at a slight expense of response time.

Several mechanisms of buffering noise in mammalian gene expression have been proposed, mostly involving gene-specific solutions, such as feedback or feedforward motifs in their transcriptional regulation, or the co-expression of its own microRNA (Arias and Hayward, 2006; Li et al., 2009; Milo et al., 2002; Schmiedel et al., 2015). Cellular compartmentalization into the nucleus and the cytoplasm however acts more globally. Thus, regulation of nuclear retention may be a primary mechanism for noise buffering of gene transcription in mammalian cells, with additional mechanisms allowing further gene-specific adaptation. While a relatively slow rate of transcript degradation in mammalian cells may also contribute to buffering stochastic fluctuations, this also affects the mean abundance of a transcript and its ability to respond. Buffering through nuclear retention does not or to a much lesser extent have these drawbacks.

The broad range of nuclear retention times for individual genes in mammalian cells suggests the existence of mechanisms to couple nuclear retention time to transcription dynamics, which is more sophisticated than our simplified model. Indeed, transcript release from the nucleus, or transcript storage within sub-compartments of the nucleus, is additionally regulated (Bhatt et al., 2012; Cuijkovic-Kraljacic et al., 2012; Prasanth et al., 2005; Taddei et al., 2006), and such a coupling may involve association of active transcription sites to nuclear pores (Taddei et al., 2006), and the direct participation of nuclear pore components in the regulation of transcription machinery (Schneider et al., 2015).

Finally, besides the generally accepted view that nuclear compartmentalization of the genome during the course of evolution allowed more complex gene regulation and the rise of multicellular organisms, we speculate that it provides another important advantage: it allows a buffering of transcriptional noise, resulting in a tighter control of gene expression variability that is essential for successful multicellular development.

## EXPERIMENTAL PROCEDURES

All details of all experimental and computational procedures are described in the [Supplemental Experimental Procedures](#). CellProfiler modules are available at <http://github.com/pelkmanslab>. Single-cell distributions can be browsed at <http://image-based-transcriptomics.org>.

## Cell Cultivation

HeLa cells were cultivated and seeded for experiments as described previously (Battich et al., 2013). Keratinocytes were cultivated in CnT-57 medium (CELLnTEC) supplemented at 1:100 (v:v) with Pen Strep (GIBCO).

## Image-Based Transcriptomics

Image-based transcriptomics, including sample processing and computational object detection, were performed as described previously (Battich et al., 2013; Stoeger et al., 2015). Briefly, cells were seeded in 384-well plates, and transcripts of distinct genes were stained in separate wells by branched DNA single-molecule fluorescence in situ hybridization using ViewRNA reagents (Affymetrix) on an automated experimental platform and imaged using a CellVoyager 7000 (Yokogawa) with an enhanced CSU-X1 spinning disk (MicroLens-enhanced dual Nipkow disk confocal scanner, wide view type) and a 40× Olympus objective of 0.95 NA and Neo sCMOS cameras (Andor; 2,560 × 2,160 pixels).

## Predictions of Spots per Cell

Multilinear regression (MLR) models of spots per cell were trained using the robustfit function of MATLAB and applied to an independent biological replicate. Stochastic simulations were carried out using the Gillespie algorithm.

## In Vivo Imaging of Transcripts

HeLa 11ht cells (Weidenfeld et al., 2009) stably expressing a doxycycline-inducible Renilla luciferase transcript that contains a chimeric  $\beta$ -globin/immunoglobulin G intron in the 5' UTR, and 24 copies of the MS2 stem loops in the 3' UTR (HeLa 11ht MS2) were kindly provided by Jeffrey Chao (Friedrich Miescher Institute). To visualize transcripts tagged with the MS2 stem loops, HeLa 11ht MS2 cells were imaged in an inverted Nikon Eclipse Ti-E microscope equipped with the Yokogawa Spinning Disc System W1 and a Nikon CFI PlanApo 100× oil-immersion objective.

## SUPPLEMENTAL INFORMATION

Supplemental Information includes Supplemental Experimental Procedures, six figures, one table, one movie, and two data files and can be found with this article online at <http://dx.doi.org/10.1016/j.cell.2015.11.018>.

## AUTHOR CONTRIBUTIONS

L.P. initiated the study. N.B., T.S., and L.P. designed and analyzed the experiments and wrote the manuscript. N.B. and T.S. performed the experiments. The order of appearance of the first authors of this and related studies (Battich et al., 2013; Stoeger et al., 2015) reflects the outcome of a single random event (coin toss).

## ACKNOWLEDGMENTS

We thank Y. Yakimovich for help with computational infrastructure, R. Holtackers for help with experiments, J. Patterson for assistance, J. Chao (Friedrich Miescher Institute) and J. Ellenberg (European Molecular Biology Laboratory) for reagents, and all members of the lab for useful comments on the manuscript. L.P. acknowledges financial support for this project from SystemsX.ch, the Swiss National Science Foundation, the University of Zurich, and the University of Zurich Research Priority Program in Systems Biology and Functional Genomics.

Received: June 18, 2015

Revised: October 8, 2015

Accepted: November 4, 2015

Published: December 17, 2015

## REFERENCES

Arias, A.M., and Hayward, P. (2006). Filtering transcriptional noise during development: concepts and mechanisms. *Nat. Rev. Genet.* 7, 34–44.

- Avraham, R., and Yarden, Y. (2011). Feedback regulation of EGFR signalling: decision making by early and delayed loops. *Nat. Rev. Mol. Cell Biol.* *12*, 104–117.
- Bastos, R., Lin, A., Enarson, M., and Burke, B. (1996). Targeting and function in mRNA export of nuclear pore complex protein Nup153. *J. Cell Biol.* *134*, 1141–1156.
- Battich, N., Stoeger, T., and Pelkmans, L. (2013). Image-based transcriptomics in thousands of single human cells at single-molecule resolution. *Nat. Methods* *10*, 1127–1133.
- Bhatt, D.M., Pandya-Jones, A., Tong, A.J., Barozzi, I., Lissner, M.M., Natoli, G., Black, D.L., and Smale, S.T. (2012). Transcript dynamics of proinflammatory genes revealed by sequence analysis of subcellular RNA fractions. *Cell* *150*, 279–290.
- Bird, A.P. (1995). Gene number, noise reduction and biological complexity. *Trends Genet.* *11*, 94–100.
- Culjkovic-Kraljacic, B., Bague, A., Volpon, L., Amri, A., and Borden, K.L. (2012). The oncogene eIF4E reprograms the nuclear pore complex to promote mRNA export and oncogenic transformation. *Cell Rep.* *2*, 207–215.
- das Neves, R.P., Jones, N.S., Andreu, L., Gupta, R., Enver, T., and Iborra, F.J. (2010). Connecting variability in global transcription rate to mitochondrial variability. *PLoS Biol.* *8*, e1000560.
- Deng, Q., Ramsköld, D., Reinius, B., and Sandberg, R. (2014). Single-cell RNA-seq reveals dynamic, random monoallelic gene expression in mammalian cells. *Science* *343*, 193–196.
- Dupont, S., Morsut, L., Aragona, M., Enzo, E., Giulitti, S., Cordenonsi, M., Zanconato, F., Le Digabel, J., Forcato, M., Bicciato, S., et al. (2011). Role of YAP/TAZ in mechanotransduction. *Nature* *474*, 179–183.
- Eldar, A., and Elowitz, M.B. (2010). Functional roles for noise in genetic circuits. *Nature* *467*, 167–173.
- Elowitz, M.B., Levine, A.J., Siggia, E.D., and Swain, P.S. (2002). Stochastic gene expression in a single cell. *Science* *297*, 1183–1186.
- Engler, A.J., Sen, S., Sweeney, H.L., and Discher, D.E. (2006). Matrix elasticity directs stem cell lineage specification. *Cell* *126*, 677–689.
- Frechin, M., Stoeger, T., Daetwyler, S., Gehin, C., Battich, N., Damm, E.M., Stergiou, L., Riezman, H., and Pelkmans, L. (2015). Cell-intrinsic adaptation of lipid composition to local crowding drives social behaviour. *Nature* *523*, 88–91.
- Golding, I., Paulsson, J., Zawilski, S.M., and Cox, E.C. (2005). Real-time kinetics of gene activity in individual bacteria. *Cell* *123*, 1025–1036.
- Graf, T., and Stadtfeld, M. (2008). Heterogeneity of embryonic and adult stem cells. *Cell Stem Cell* *3*, 480–483.
- Grün, D., Kester, L., and van Oudenaarden, A. (2014). Validation of noise models for single-cell transcriptomics. *Nat. Methods* *11*, 637–640.
- Gut, G., Tadmor, M.D., Pe'er, D., Pelkmans, L., and Liberali, P. (2015). Trajectories of cell-cycle progression from fixed cell populations. *Nat. Methods* *12*, 951–954.
- Halpern, K.B., Caspi, I., Lemze, D., Levy, M., Landen, S., Elinav, E., Ulitsky, I., and Itzkovitz, S. (2015). Nuclear retention of mRNA in mammalian tissues. *Cell Rep.* *13*, <http://dx.doi.org/10.1016/j.celrep.2015.11.036>.
- Halstead, J.M., Lionnet, T., Wilbertz, J.H., Wippich, F., Ephrussi, A., Singer, R.H., and Chao, J.A. (2015). Translation. An RNA biosensor for imaging the first round of translation from single cells to living animals. *Science* *347*, 1367–1671.
- Kempe, H., Schwabe, A., Crémazy, F., Verschure, P.J., and Bruggeman, F.J. (2015). The volumes and transcript counts of single cells reveal concentration homeostasis and capture biological noise. *Mol. Biol. Cell* *26*, 797–804.
- Kresnowati, M.T., van Winden, W.A., Almering, M.J., ten Pierick, A., Ras, C., Knijnenburg, T.A., Daran-Lapujade, P., Pronk, J.T., Heijnen, J.J., and Daran, J.M. (2006). When transcriptome meets metabolome: fast cellular responses of yeast to sudden relief of glucose limitation. *Mol. Syst. Biol.* *2*, 49.
- Li, X., Cassidy, J.J., Reinke, C.A., Fischboeck, S., and Carthew, R.W. (2009). A microRNA imparts robustness against environmental fluctuation during development. *Cell* *137*, 273–282.
- Liaw, A., and Wiener, M. (2002). Classification and regression by randomForest. *R News* *2*, 18–22.
- Macarthur, B.D., Ma'ayan, A., and Lemischka, I.R. (2009). Systems biology of stem cell fate and cellular reprogramming. *Nat. Rev. Mol. Cell Biol.* *10*, 672–681.
- Martin, W., and Koonin, E.V. (2006). Introns and the origin of nucleus-cytosol compartmentalization. *Nature* *440*, 41–45.
- Milo, R., Shen-Orr, S., Itzkovitz, S., Kashtan, N., Chklovskii, D., and Alon, U. (2002). Network motifs: simple building blocks of complex networks. *Science* *298*, 824–827.
- Neuert, G., Munsky, B., Tan, R.Z., Teytelman, L., Khammash, M., and van Oudenaarden, A. (2013). Systematic identification of signal-activated stochastic gene regulation. *Science* *339*, 584–587.
- Newman, J.R., Ghaemmaghami, S., Ihmels, J., Breslow, D.K., Noble, M., DeRisi, J.L., and Weissman, J.S. (2006). Single-cell proteomic analysis of *S. cerevisiae* reveals the architecture of biological noise. *Nature* *441*, 840–846.
- Ochiai, H., Sugawara, T., Sakuma, T., and Yamamoto, T. (2014). Stochastic promoter activation affects Nanog expression variability in mouse embryonic stem cells. *Sci. Rep.* *4*, 7125.
- Padovan-Merhar, O., Nair, G.P., Bialesch, A.G., Mayer, A., Scarfone, S., Foley, S.W., Wu, A.R., Churchman, L.S., Singh, A., and Raj, A. (2015). Single mammalian cells compensate for differences in cellular volume and DNA copy number through independent global transcriptional mechanisms. *Mol. Cell* *58*, 339–352.
- Prasanth, K.V., Prasanth, S.G., Xuan, Z., Hearn, S., Freier, S.M., Bennett, C.F., Zhang, M.Q., and Spector, D.L. (2005). Regulating gene expression through RNA nuclear retention. *Cell* *123*, 249–263.
- Raj, A., and van Oudenaarden, A. (2008). Nature, nurture, or chance: stochastic gene expression and its consequences. *Cell* *135*, 216–226.
- Raj, A., Peskin, C.S., Tranchina, D., Vargas, D.Y., and Tyagi, S. (2006). Stochastic mRNA synthesis in mammalian cells. *PLoS Biol.* *4*, e309.
- Raser, J.M., and O'Shea, E.K. (2004). Control of stochasticity in eukaryotic gene expression. *Science* *304*, 1811–1814.
- Raser, J.M., and O'Shea, E.K. (2005). Noise in gene expression: origins, consequences, and control. *Science* *309*, 2010–2013.
- Sanchez, A., and Golding, I. (2013). Genetic determinants and cellular constraints in noisy gene expression. *Science* *342*, 1188–1193.
- Schmiedel, J.M., Klemm, S.L., Zheng, Y., Sahay, A., Blüthgen, N., Marks, D.S., and van Oudenaarden, A. (2015). Gene expression. MicroRNA control of protein expression noise. *Science* *348*, 128–132.
- Schneider, M., Hellerschmied, D., Schubert, T., Amlacher, S., Vinayachandran, V., Reja, R., Pugh, B.F., Clausen, T., and Köhler, A. (2015). The nuclear pore-associated TREX-2 complex employs mediator to regulate gene expression. *Cell* *162*, 1016–1028.
- Shalek, A.K., Satija, R., Adiconis, X., Gertner, R.S., Gaubomme, J.T., Raychowdhury, R., Schwartz, S., Yosef, N., Malboeuf, C., Lu, D., et al. (2013). Single-cell transcriptomics reveals bimodality in expression and splicing in immune cells. *Nature* *498*, 236–240.
- Shapiro, E., Biezuner, T., and Linnarsson, S. (2013). Single-cell sequencing-based technologies will revolutionize whole-organism science. *Nat. Rev. Genet.* *14*, 618–630.
- Singh, A., and Bokes, P. (2012). Consequences of mRNA transport on stochastic variability in protein levels. *Biophys. J.* *103*, 1087–1096.
- Snijder, B., and Pelkmans, L. (2011). Origins of regulated cell-to-cell variability. *Nat. Rev. Mol. Cell Biol.* *12*, 119–125.
- Stoeger, T., Battich, N., Herrmann, M.D., Yakimovich, Y., and Pelkmans, L. (2015). Computer vision for image-based transcriptomics. *Methods* *85*, 44–53.

- Suter, D.M., Molina, N., Gatfield, D., Schneider, K., Schibler, U., and Naef, F. (2011). Mammalian genes are transcribed with widely different bursting kinetics. *Science* 332, 472–474.
- Taddei, A., Van Houwe, G., Hediger, F., Kalck, V., Cubizolles, F., Schober, H., and Gasser, S.M. (2006). Nuclear pore association confers optimal expression levels for an inducible yeast gene. *Nature* 441, 774–778.
- Weidenfeld, I., Gossen, M., Löw, R., Kentner, D., Berger, S., Görlich, D., Bartsch, D., Bujard, H., and Schönig, K. (2009). Inducible expression of coding and inhibitory RNAs from retargetable genomic loci. *Nucleic Acids Res.* 37, e50.
- Xiong, L.P., Ma, Y.Q., and Tang, L.H. (2010). Attenuation of transcriptional bursting in mRNA transport. *Phys. Biol.* 7, 016005.
- Zenklusen, D., Larson, D.R., and Singer, R.H. (2008). Single-RNA counting reveals alternative modes of gene expression in yeast. *Nat. Struct. Mol. Biol.* 15, 1263–1271.