bioRxiv preprint doi: https://doi.org/10.1101/411447; this version posted September 7, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Gene-specific transcriptional memory in mammalian cell lineages

Nicholas E. Phillips[1,2], Aleksandra Mandic[1,2], Saeed Omidi[1], Felix Naef[1,3*], David M. Suter[1,3*]

[1]Institute of Bioengineering, School of Life Sciences, Ecole Polytechnique Fédérale de Lausanne, Lausanne CH-1015, Switzerland.

[2]These authors contributed equally to this work

[3]These authors contributed equally to this work

*Corresponding authors: david.suter@epfl.ch and felix.naef@epfl.ch

Keywords: transcriptional memory, transcriptional fluctuations, single-cell analysis, live imaging, mathematical modelling, cell lineages

Abstract

Phenotypically identical mammalian cells often display considerable variability in transcript levels of individual genes. How transcriptional activity propagates in cell lineages, and how this varies across genes is poorly understood. Here we combined live-cell imaging of short-lived transcriptional reporters in mouse embryonic stem cells with mathematical modelling to quantify the propagation of transcriptional activity over time and across cell generations. In sister cells we found mean transcriptional activity to be strongly correlated and transcriptional dynamics tended to be synchronous; both features control how quickly sister cells diverge in a gene-specific manner. Mean transcriptional activity was also highly correlated between mother and daughter cells, leading to multi-generational transcriptional memory whose duration scaled with the spread of transcriptional activities in the population. The resulting family-specific transcriptional levels suggest a potential role of transcriptional memory in patterning tissue gene expression.

bioRxiv preprint doi: https://doi.org/10.1101/411447; this version posted September 7, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Introduction

Major changes in transcriptional states that then propagate through cell generations is characteristic of embryonic development. Such dynamics often result in irreversible changes in phenotypic states that are then transmitted through cell division[1]. While genome-wide alterations of gene expression profiles are characteristic of cell differentiation, even phenotypically identical cells display significant variability in the levels at which individual genes are expressed[2–4]. The dynamic properties of such fluctuations are determined by both intrinsic noise resulting from the randomness in biochemical reactions controlling gene expression, as well as extrinsic variability caused by differences in cellular parameters[5], such as size[6,7], mitochondrial content[8,9], cell cycle stage[6,10–12], or differences in cellular microenvironment[9,13,14]. In particular, transcriptional bursting causes intrinsic fluctuations with a time scale on the order of one to several hours[15–17], while extrinsic fluctuations in cellular parameters can be significantly longer-lived[18]. How such transcriptional fluctuations structure gene expression dynamics in families of phenotypically homogenous cells, and to which extent genes exhibit transcriptional memory, is largely unknown. Trans-generational transcriptional memory might have direct consequences in generating spatial gene expression patterns, for instance in solid tissues where cells sharing a common ancestor typically remain in close proximity.

Only few studies have investigated transcriptional memory in lineages of phenotypically identical cells. For example, transcriptional parameters in *Dictyostelium* were found to be correlated both between sister and mother-daughter cells[19]. In the developing *Drosophila* embryo, higher transcriptional activity in mother nuclei increases the probability of rapid re-activation in daughter nuclei[20]. In mammalian cells, a study showed that endogenous protein expression levels could be passed on during cell division, with memory time scales typically ranging between 1 to 3 cell cycles[21]. Such protein memory may largely reflect mRNA and

bioRxiv preprint doi: https://doi.org/10.1101/411447; this version posted September 7, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

protein half-lives[22], which can easily exceed the duration of the cell cycle[23]. Altogether, little is known about the times scales of transcriptional memory in mammalian cells.

Here we used short-lived transcriptional reporters to determine how transcriptional fluctuations are propagated over time and across cell division in mouse embryonic stem cells. We found that genes differ broadly in the dynamics of their transcriptional fluctuations at both short (in the hour range) and long (cell generations) time-scales, which results in large differences in the propagation of transcriptional activity. We also found a remarkably large correlation in transcriptional activity of sister cells, suggesting that inherited factors from the mother cell and/or similarity in cellular microenvironment contribute to transcriptional dynamics in dividing cells. Our results suggest that inheritance of transcriptional activity structures temporal fluctuations and overall mRNA production across cell lineages.

Results

Spread and relatedness of transcriptional fluctuations are gene specific

To monitor how transcriptional levels fluctuate and propagate over cell generations, we inserted a short-lived transcriptional luminescent reporter by gene trapping into endogenous genes (Supplementary Fig. 1). This method allows sensitive monitoring of transcriptional activity by luminescence imaging at high time resolution without observable toxicity over long periods of time[15]. In total we produced eight different gene trap cell lines, and an additional cell line where a construct driving the expression of the short-lived luciferase from the pGK promoter was integrated as a single copy in the genome[15]. The insertion sites of the constructs were mapped using splinkerette PCR (Supplementary Fig. 2)[24]. To analyse how temporal transcriptional activity profiles compare both in pairs of mother-daughter as well as

bioRxiv preprint doi: https://doi.org/10.1101/411447; this version posted September 7, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

sister cells (Fig.1a, b), we monitored total transcriptional reporter levels with a time resolution of 5 minutes, and manually tracked approximately 50 pairs of sister cells per cell line from division to division to obtain single-cell traces. In addition, for three clones we quantified transcriptional activity profiles of mother and daughter cells over two cell generations.

We first aimed to determine whether differences in transcriptional levels across cells decayed quickly or if they were maintained over longer timescales and transmitted to daughter cells (Fig. 1c). The live-cell imaging of sister cells generated pairs of time traces, and exploratory data analysis revealed several key features of transcriptional dynamics. First, the mean and spread of transcriptional reporter levels across the population of cells in function of time were gene-specific (Fig. 2a, Supplementary Fig. 3). The average transcriptional reporter levels across the population increased during G1 phase (see Methods for cell cycle phase definition), consistent with RNA-seq analysis of pre-mRNA around the cell cycle [25], and then stayed approximately constant during S and G2 phases for most genes. Sorting cells by initial transcriptional reporter levels showed that for the *pGK* clone, cells tended to retain their relative expression levels for longer time than for the *Dstn* gene (Fig. 2a). For *pGK*, transcriptional activity fluctuated around largely different mean levels in individual cells (Fig. 2b), suggesting that cells retained their average transcriptional levels over longer times than for *Dstn*. Unexpectedly, transcriptional profiles of sister cells often showed striking similarity over the cell cycle (Fig. 2c). Moreover, sister cells showed high correlation in reporter levels immediately following cell division, as explained from the partitioning of reporter protein and mRNA molecules. This sister-cell correlation then decreased over the cell cycle in a gene-specific manner at a slower pace than non-sister control pairs matched for similar initial levels (Fig. 2d, all genes shown in Supplementary Fig. 4), suggesting that transcriptional activity is transmitted along cell lineages.

A stochastic model to analyse transcriptional dynamics in dividing cells

bioRxiv preprint doi: https://doi.org/10.1101/411447; this version posted September 7, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Next, we developed a mathematical model to quantitatively assess how transcriptional activity fluctuates in pairs of sister cells, taking into account the features described in Fig. 2. We built a minimal kinetic model (Fig. 3) to fit the transcriptional reporter data for each gene. This model describes fluctuations both at the single-cell and population level (*i.e.* the set of all time traces for paired sister cells). In each cell, our model describes the production and degradation of the transcriptional reporter *R* and consists of two time-dependent and stochastic variables: the transcriptional activity *S* that acts as a source for the transcriptional reporter *R* (Fig. 3a). To account for the spread in mean levels (Fig. 2b), *S* is allowed to fluctuate around a cell-specific mean, and the variances of *S* and *R* are also cell-specific. Reporter levels *R* are produced at rate *S* and their effective half-lives were measured independently by blocking transcription with actinomycin D (Supplementary Fig. 5; assumed to be constant across cells for the analysis). This estimated half-life is therefore dependent both on reporter protein and mRNA half-lives. We further introduced the parameter ρ describing the correlation of transcriptional fluctuations between sister cells, which tunes the extent to which sister cells acquire similar reporter profiles over the cell cycle (Fig. 3b). To set the initial conditions, we modelled the mean, variance and co-variance of *R* and *S* in the beginning of each cell cycle from the predicted steady-state, assuming *R* at the beginning of the cell cycle to be at half its steady state value to reflect cell division.

The population model is then built hierarchically, whereby the cell-specific parameters are related to each other through a population level distribution (Fig. 3c), and these population parameters are estimated within our inference scheme. These global parameters therefore control the distribution of cell-specific parameters over all pairs of sister cells, such as the cell-to-cell variability in mean transcription rates. In this model, high intercellular variability in mean transcription rates leads to a broad range of cell-specific means across the population, whereas lower intercellular transcriptional variability describes cells in the population sharing similar mean levels (Fig. 3c). The correlation in mean transcriptional activity between sister cells is quantified with the parameter λ (Fig. 3d), and along with the similarity in dynamics (ρ)

bioRxiv preprint doi: https://doi.org/10.1101/411447; this version posted September 7, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

these two parameters connect sister cells. Time traces are analysed and model parameters estimated within a Bayesian hierarchical framework, which combines Gaussian processes with Hamiltonian Markov Chain Monte Carlo (MCMC) sampling for efficient inference (Methods).

Mean transcriptional activity is highly correlated between sister cells

We applied our inference scheme to estimate the parameters of our model for each gene individually (parameters estimates for all genes shown in Supplementary Fig. 6, example trace plots shown in Supplementary Fig. 7). To validate our method we simulated bioluminescent time series for a range of parameters using 50 pairs of cells and cell-cycle lengths that were similar to our data (Supplementary Note 1), and we found that for all parameters the true values used for simulation were in the 90% credible intervals (Supplementary Fig. 8), showing our method can reliably recover parameters for data that resembled our experiments. We next used our model to analyse the gene-specificity of the variability of cell-specific means and correlations between sister cells (Fig. 2). The spread of cell-specific means varies significantly across genes, with *pGK* being the most (coefficient of variation, CV=0.7) and *Dstn* the least (CV=0.2) variable (Fig. 4a). To test whether cell-specific means are correlated between the two daughters, we analysed the parameter λ (Fig. 4b). Interestingly, λ was less variable and consistently high across genes, ranging from 0.7 to 0.95 (Fig. 4b). The genes with the highest variability of cell-specific means also exhibited the most highly correlated sister cells (Fig. 4c). Of note this was not due to a structural property of the model, namely, the two parameters were not correlated during inference (Supplementary Fig. 9). These analyses thus suggest that sister cells inherit highly correlated mean transcriptional activities. Below, we investigate the impact on the maintenance of sister-cell correlation over the cell cycle.

Transcriptional fluctuations show synchronicity in sister cells

bioRxiv preprint doi: https://doi.org/10.1101/411447; this version posted September 7, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

We next determined whether the similarity in the dynamics of sister cells we observed (Fig. 2c) could be substantiated by our mathematical model. In the model, similarity of dynamics is quantified with the correlation parameter ρ, ranging from -1 to 1. $\rho = 0$ indicates independent fluctuations in *S*, while $\rho = 1$ indicates identical shapes of transcriptional activity over the cell cycle (for identical initial conditions). Intriguingly, the inferred values of ρ were positive for all genes, confirming that sister cells tend to show correlated dynamics (Fig. 4d). The degree of similarity in dynamics was gene-specific but overall lower than λ, ranging from $\rho = 0.3$ for *Spry4* to $\rho = 0.7$ for *Jam2*. Having found that correlated transcriptional fluctuations are detectable for all genes, we wanted to further explore the origins of this similarity in dynamics by analysing pairings of randomised non-sister cells (examples shown in Supplementary Fig. 10). If there was cell-cycle dependent transcriptional control affecting all cells, this would lead to a non-zero ρ value even amongst random pairings of cells. In fact, we found that most ρ values were only slightly above zero for random cell pairings, which suggests a modest contribution of cell cycle progression to ρ (Fig. 4e).

The origin of correlated dynamics between sister cells remains unsolved, but one possible explanation is that sister cells share a common microenvironment, and that transcriptional activity could be regulated by local signalling. To address this question, we compared to a control situation in which non-sister cells separated by same average distance as true sister cells were paired. This showed that while ρ was higher for sisters than non-sisters for both *Rbpj* and *Jam2* (Fig. 4f), the value of ρ for non-sisters pairs was still higher than for fully randomised pairings of cells, the latter being on average more spatially distant (compare Figs. 4e and f). Therefore, the microenvironment can, at least in some cases, increase the synchrony in transcriptional dynamics of cells that are close in space.

Correlated mean transcription levels and similar dynamics control sister cell correlations

bioRxiv preprint doi: https://doi.org/10.1101/411447; this version posted September 7, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

We next aimed to investigate how the correlated levels of mean transcriptional activity (λ) and similarity in dynamics (ρ) between sister cells impact the observed loss of correlation between sister cells for each gene over the cell cycle (Fig. 2d). We therefore used the model to predict how the correlation between sister cells evolves over time, using the inferred parameter values (the posterior means) of each gene and the empirical correlation between sisters at the beginning of the cell cycle. Remarkably, comparing the correlation over the cell cycle from the model (red, Fig. 4g-i) with the empirical correlation from the data (green, Fig. 4g-l, all genes shown in Supplementary Fig. 11) showed very good agreement, even if the model was fitted to the time series and hence not directly fitted to this correlation decay. Next, to quantify the relative contributions of different processes to maintaining similar transcriptional levels between sisters, we dropped certain features from the model. First, we set the parameter *s* to zero (such that all cells share the same mean transcriptional activity), which made the predicted correlation between sister cells decay much faster for most genes (violet, Fig. 4g-i). Similarly, setting ρ to zero led to a faster decorrelation between sister cells (yellow, Fig. 4g-i) (when ρ and *s* are removed from the model before fitting to data the correlation remains underestimated, showing that both features are required to account for the sister-cell correlation in the data (Supplementary Fig. 12)). Therefore, both *s* and ρ positively contributed to the correlation between sister cells, but the relative contributions of these two parameters was gene-specific. For *pGK*, the predicted sister-sister correlation was much lower when variability in cell-specific means was removed (*s*=0, Fig. 4g), which suggests that variable cell-specific means are important to maintain similar transcriptional activity between sisters for this gene. In contrast, the predicted correlation was not changed significantly for *Dstn* when variable cell-specific means were abolished (*s*=0) from the model, and the similarity in dynamics (ρ) were more important for maintaining correlation between sisters (Fig. 4i). Our model therefore shows that not only is the maintenance of sister-sister correlation gene-specific, but also that different processes tune sister-sister correlation for different genes.

bioRxiv preprint doi: https://doi.org/10.1101/411447; this version posted September 7, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Mean transcriptional activity is transmitted to daughter cells

Having observed that transcriptional activity is highly correlated between sister cells, we next explored whether transcriptional states were also propagated through cell division. Given that sister cells inherit highly correlated mean transcriptional activities and display similarity in their transcriptional dynamics (Fig. 4b,d), we asked whether this was also the case for mother-daughter pairs. We thus measured the reporter levels of mother and daughter cells for three genes (*pGK*, *Jam2* and *Rbpj*) and re-fitted our model (examples of two pairs shown in Fig. 5a, *pGK* gene). Similarly to the sister cells, cell-specific means between mother and daughter cell were again highly correlated, showing that mean transcriptional activity can be robustly transmitted across generations (Fig. 5b). In contrast, the similarity in dynamics between mother and daughter cells was low (Fig. 5c). Taken together, this data suggests that while the mother cell may to some extent set temporal patterns of transcriptional fluctuations in daughter cells, the shape of fluctuations is largely independent between cell generations. Therefore, the transmission of cell-specific mean transcriptional activity through cell division is the main contributor to the propagation of transcriptional levels from mother to daughter cells. Given that the mean correlation (λ) between mother and daughter cells for *pGK*, *Jam2* and *Rbpj* was 0.92, 0.87 and 0.86, respectively, a simple extrapolation predicts that it would take 17, 9 and 8 cell generations for this correlation to be reduced by a factor of 1/e~0.37. Thus, this indicates that the inheritance of correlated mean levels of transcriptional activity results in multi-generational transcriptional memory.

Gene-specific transcriptional memory generates gene expression patterns across cell families

bioRxiv preprint doi: https://doi.org/10.1101/411447; this version posted September 7, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

As mother-daughter analysis showed that transcriptional activity could be transmitted across generations, we reasoned that this should lead cells sharing a common ancestor to display more similar transcriptional activities than unrelated cells. To test this hypothesis, we first measured transcriptional activities within and across families of at least four cells (*i.e.* all cells sharing the same grandmother) for the *pGK*, *Jam2* and *Dstn* genes, as examples of genes with slow, intermediate and fast loss of correlation between sister cells (Supplementary Fig. 4). To minimise biases linked to cell-cycle related changes in expression levels, we averaged luminescence levels from three image frames preceding nuclear envelope breakdown in families of at least four cells. We found that the family-averaged levels of transcriptional activity were more variable for *pGK* and *Jam2* than for *Dstn* (Fig. 6a-d). In contrast, transcriptional activity within cell families was less variable for *pGK* and *Jam2* than for *Dstn* (Fig. 6e), in line with the higher strength of transcriptional noise (*i.e.* the steady state distribution of *S*) of *Dstn* (Fig. 6f). These results suggest that the maintenance of cell-specific mean transcriptional activity across cell generations allows cells to propagate intercellular variability in transcriptional activity, thus generating differences between average transcriptional levels of cell families.

While it is technically difficult to quantify transcriptional memory over a much larger number of cell generations, we reasoned that we could obtain qualitative insights into longer-term memory by comparing the average transcriptional activity of ES cell colonies resulting from about five cell divisions. We thus seeded the *pGK, Jam2* and *Dstn* cell lines at low density, and imaged the resulting cell colonies 60 hours later. In contrast to the *Dstn* and *Jam2* colonies, the *pGK* colonies displayed markedly different average luminescence between them (Fig. 6d), suggesting that transcriptional activity can be inherited over many cell generations.

Discussion

One of the major challenges in quantitative biology is to understand how gene expression

bioRxiv preprint doi: https://doi.org/10.1101/411447; this version posted September 7, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

dynamics of single cells are related in the context of multicellularity. The combination of lineage-tracing and mRNA measurements has previously been used to quantify the dynamics of cell-fate transitions[26,27]. However, thus far, it remained unclear how transcriptional fluctuations are propagated in lineages of phenotypically homogenous cells, and to which extent this transmission is gene-specific. Previous studies in fixed mammalian cell lines have reported higher similarity of mRNA levels of neighbouring cells[9] and that population context can predict cellular features such as membrane lipid composition and endocytosis[13], but the impact of lineage relationships on such microenvironment-related correlations was not addressed in these studies.

Lineage information was found to be an important contributor to patterning gene expression in bacterial microcolony formation[28,29], where it can act as the dominant cause of spatial correlations[30]. While properties such as cell cycle duration have been shown to propagate in mammalian cell lineages[31,32], the importance of genealogy for transcriptional activity in mammalian cells is still poorly studied. Here, we used live-cell imaging to measure and compare transcriptional activity of lineage-related mammalian cells over time. We developed a simple yet powerful stochastic model of gene expression fluctuations, which combined with Bayesian inference allowed us to identify the key processes and parameters underlying the observed correlation patterns of transcriptional reporter levels within lineage-related cells. This quantitative analysis allowed us to separate short-term transcriptional fluctuations from long-term trends, which both contribute to population heterogeneity in the dynamics of the observed reporter levels.

In particular, we found that transcriptional activities in each cell within the population fluctuate around cell-specific mean levels, which propagate through cell division in a gene-specific manner and result in multigenerational transcriptional memory. Remarkably, we also showed that the rate at which transcriptional activity of sister cells diverge from each other to be correlated with the spread of transcriptional activity in the population (Fig. 4c). This implies that the time required for a cell to explore the full range of expression levels scales

bioRxiv preprint doi: https://doi.org/10.1101/411447; this version posted September 7, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

with variability – in other words, a gene displaying a large spread of transcriptional activities will transmit cell-specific activities for a longer period of time.

Surprisingly, sister cells displayed not only similar cell-specific mean levels, but also correlation of their transcriptional dynamics over the cell cycle. Our model uses a dedicated parameter (ρ) to capture similarity in dynamics, which contrasts with previous mathematical modelling assuming that transcriptional dynamics of sister cells are independent [33,34]. At the mechanistic level, this correlation in dynamics could be caused by correlated inheritance of factors from the mother cell that control transcriptional dynamics. The much lower correlated dynamics between mother and daughter cell pairs suggest that the factors controlling transcriptional activity may fluctuate significantly over the course of one cell cycle, and thus set a different transcriptional dynamics program in the next cell generation. For some genes non-sister cells in the same spatial proximity as sister cells also exhibit correlated transcriptional fluctuations (Fig. 4f), which could be due to the exposure to shared extracellular signals, compared to cells that are more distant in space. However, we cannot exclude that such non-sister cells could be more distantly related. Notably, both the inherited and microenvironmental factors may have indistinguishable consequences on transcriptional dynamics similarity of proliferating adherent cells, as related cells will typically remain in close spatial proximity.

Several potential regulators could determine the timescale of transcriptional memory. While physiological parameters such as cell size variability could explain differences in mRNA counts across cells[9], these global factors are unlikely to fully explain our data as they are common for all genes examined, and for example the CV of the cell-specific transcriptional activities ranges from 0.2-0.7 across the genes we measured (Fig. 4a). Potential gene-specific factors include both *cis*-regulatory elements such as epigenetic marks of promoters and enhancers, or *trans*-regulatory elements – *i.e.* molecules such as transcription factors. While memory conferred by *trans*-regulatory elements essentially depends on their half-lives and inheritance of their expression levels, *cis*-regulatory mechanisms have the potential to

bioRxiv preprint doi: https://doi.org/10.1101/411447; this version posted September 7, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

fine-tune inheritance of transcriptional activity over a broad range, since different types of chromatin modifications fluctuate on drastically different time scales[35].

The findings we describe here suggest a potential role for propagation of transcriptional activity in tissue patterning during developmental processes in the sense that transcriptional memory may act as a pattern generator. As lineage-related cells often remain spatially close in a growing tissue, transcriptional memory may contribute to the formation of cell clusters retaining similar gene activity profiles. This passive mechanism could thereby initiate changes in expression patterns between groups and families of cells, which may be further reinforced and stabilised by diverging cell fate decisions. Future studies shall investigate whether genes encoding master regulators of cell fate may exhibit long-term transcriptional memory, allowing cells to propagate their expression levels across several generations to initiate tissue patterning during developmental processes.

Methods

Experimental methods

Generation of lentiviral constructs

To generate the pSTAR-GTX gene trap lentiviral vector, ten repeats of the 9-nucleotide IRES element derived from the 5'UTR sequence of the *gtx* mRNA (Chappell, Edelman, Maura, PNAS 2000), interspersed with 9-nt spacers based on a segment of the β-globin 5' UTR (nt 9-17), were inserted upstream of bsdF2ANLSLuc by restriction cloning into the pSTAR lentiviral vector[36]. To generate the pGK-Luc lentiviral construct, the pGK promoter was PCR-amplified from the pLV-pGK-rtTA3G-IRES-Bsd[36] and inserted upstream of bsdF2ANLSLuc by restriction cloning into the pSTAR lentiviral vector.

bioRxiv preprint doi: https://doi.org/10.1101/411447; this version posted September 7, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Stable cell line generation

The stable gene trap (GT) cell lines were generated by transducing E14 mouse embryonic stem (ES) cells (kindly provided by Didier Trono, EPFL) with the concentrated virus carrying the pSTAR-GTX or pGK-Luc construct. Virus production was performed by co-transfection of HEK 293T cells with the construct of interest, the envelope (PAX2) and packaging (MD2G) constructs using calcium phosphate, and concentrated 120-fold by ultracentrifugation as described previously[15]. ES cells were then seeded at a density of 125,000 cells per 10 cm dish and transduced with 125 µl of virus. Antibiotic selection was started by addition of 10 µg/ml of blasticidin 3 days after transduction, while the outgrown colonies were picked 14-21 days after. The small number of outgrown colonies per 10 cm dish (two on average) ensured we obtained a single active insertion per clone. Colonies were then expanded in the selection medium and subsequently frozen. The FUCCI ES cell line was generated by transducing ES cells with 50µl of 120-fold concentrated lentiviral vectors encoding mKO2-hCdt1 and mAG-hGem[37], followed by FACS to sort cells positive for both mKO2 and mAG fluorescence.

Cell culture

ES cell lines were cultured at 37°C and 5% CO_2, on dishes coated with 0.1% gelatin type B (Sigma), in GMEM (Sigma) medium supplemented with 10% ES cell-qualified FBS, 1x nonessential amino acids (NEAA), 2 mM L-glutamine, sodium pyruvate, 100µM 2-mercaptoethanol, 1% penicillin and streptomycin, home-made leukemia inhibitory factor (LIF), CHIR99021 at 3µM and PD184352 at 0.8µM. Cells were split every 2-3 days. The pGK-Luc cell line was constantly maintained in the presence of 10µg/ml of blasticidin to prevent silencing of the reporter.

HEK 293T cells were cultured at 37°C and 5% CO_2, in DMEM medium (Sigma) supplemented with 10% FBS and 1% penicillin and streptomycin (BioConcept, 4-01F00H).

bioRxiv preprint doi: https://doi.org/10.1101/411447; this version posted September 7, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Mapping of insertion sites in gene trap cell lines

To identify the endogenous gene into which the pSTAR-GTX was inserted in each GT cell line, we used splinkerette PCR (spPCR) [24] with modified primer sequences adapted to our lentiviral gene trap construct (Supplementary Table 1). This method allows the amplification of a portion of DNA between the GT cassette and a known DNA sequence (adaptor). Genomic DNA (gDNA) was extracted from cells of each clone using the Qiagen gDNA Extraction Kit (Qiagen). gDNA was cut with 4-cutter restriction enzyme MluCI, followed by ligation to the annealed small and long adaptor. The ligation was followed by HindIII digestion, allowing removal of the adaptors and most of the GT cassette. Then, the portion of DNA between the adaptor and the GT cassette was amplified through two rounds of PCR. The bands from the nested PCR were purified using the QIAquick gel extraction kit (Qiagen) and directly sequenced using nested primers (Supplementary Table 2; F2 and R2). Sequences derived from spPCR were used to identify the insertion site through the BLAT genome alignment tool (http://genome.ucsc.edu) (Supplementary Fig. 2)[38]. At the same time, since MluCI and EcoR V cut both LTRs, an additional 200 bp DNA segment was amplified in all samples, which was used as a control of successful nested PCR amplification.

Luminescence Microscopy

Luminescence imaging was performed on an Olympus LuminoView LV200 microscope equipped with an EM-CCD camera (Hamamatsu photonics, EM-CCD C9100-13), a 60-fold oil-immersion magnification objective (Olympus UPlanSApo 60x, NA 1.35, oil immersion) in controlled environment conditions (37°C, 5% CO_2). 16 to 24 hours before imaging, 50,000-75,000 cells were seeded on FluoroDishes (WPI, FD35-100) coated with E-cadherin, allowing to obtain a monolayer of individual cells suitable for single cell tracking[39]. The medium was supplemented with 0.5 mM luciferin (NanoLight Technology, Cat#306A) two to four hours before imaging. Fields of view with about 10 to 30 cells were imaged every 5 minutes with an exposure time of 299 seconds for 24 to 48 hours. To examine propagation of gene expression levels within ES cell colonies (Fig. 6d), 500-1000 cells were seeded on

bioRxiv preprint doi: https://doi.org/10.1101/411447; this version posted September 7, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Fluorodishes coated with gelatin, and grown as colonies for 60 hours. For each clone, two consecutive images with an exposure time of 5 (Dstn and Jam2) or 3 (pGK) minutes in at least 10 fields of view were acquired.

Reporter half-life measurements

Single cell reporter half-lives were determined by treating cells with 5 µg/ml of Actinomycin D, which inhibits RNA elongation and thus results in transcriptional arrest[40]. Luminescence imaging was performed as described above for 3 to 5 hours, starting immediately after addition of Actinomycin D. Although both protein and mRNA half-lives contribute to overall reporter half- life (τ_R), the decay curve was well fitted by a first order exponential function (Supplementary Fig. 5).

Cell cycle phase durations

In order to determine the durations of the different cell cycle phases, we combined different approaches. We first used time-lapse imaging of ES cells expressing both components of the FUCCI system[41] to measure the duration of the whole cell cycle and of G1 phase. The FUCCI systems relies on biphasic cell cycle-dependent activity and proteolysis of the ubiquitination oscillators Cdt1 and Geminin, whose fragments are fused to mKO2 and mAG, respectively. Cells were seeded on E-cadherin at a density of 50,000 cells per well of a black 96-well plate (Sigma) 16 to 24 hours before imaging. Time-lapse fluorescence imaging was performed using an inverted Olympus Cell xCellence microscope equipped with a 20x objective (Olympus UPlanSApo 20x, NA 0.75) in controlled environment conditions (37°C, 5% CO_2). Green and red fluorescence were measured using the GFP and Cy3 channel, respectively, every 10 minutes with an exposure time of 300 ms for 24 hours. The fluorescence time-lapse acquisitions were analysed manually using the Fiji software. mKO2 expression allowed us to define the duration of G1, while mAG was expressed in the S, G2 and M phases. To directly measure the length of M phase in mES cells, we used single cell traces from the luminescence time-lapse acquisitions in which nuclear envelope breakdown

bioRxiv preprint doi: https://doi.org/10.1101/411447; this version posted September 7, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

is clearly visible as a sudden increase in the area occupied by the luminescence signal of an individual cell. We thus manually determined the number of frames from the moment of nuclear breakdown in the prophase of the cell cycle, until the moment when we see formation of two new nuclei manually. Using this information we were able to calculate the average length of M phase in single cells (Supplementary Fig. 13).

Cell tracking and image analysis

Prior to quantification of single cell gene expression from luminescence microscopy movies, we removed imaging artifacts known as cosmic rays using the Min operation of the Fiji software Image Calculator function. To track cells, we used Fiji to manually draw the outlines around cells, using a fixed area with shape adjustment when required. Background measurements were performed close to every tracked cell, in regions devoid of luminescent signal separately for each time point of the movie, and these values were subtracted from cell measurements. Cells were tracked from the time they were born (just after division of their mother cell) until the last frame before cytokinesis, either as pairs of sisters or pairs of mother and daughter cells. For the experiments investigating the impact of microenvironment on similarity of gene expression between cells, the distance between sister cells and non-sister cells was measured by hand-drawing a line between the approximate centres of two nuclei. The distance between sister cells was measured every ten frames for 500 minutes, starting from the tenth frame after their birth. In the case of non-sister cells, the distances between cells present over the same time period in the field of view were measured every ten frames for 500 minutes. For the *Jam2* gene, the average distance of sisters was 0.07 $\pm$ 0.01 μm (standard error), and for non-sisters the average distance was 0.08 $\pm$ 0.01 μm. For the *Rbpj* gene, the average distance of sisters was 0.11 $\pm$ 0.01 μm, and for non-sisters the average distance was 0.12 $\pm$ 0.01 μm. We defined the first measurement as the time frame when the later cell in a pair was born. Additionally, for the

bioRxiv preprint doi: https://doi.org/10.1101/411447; this version posted September 7, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

cell family experiments (Fig. 6a-c), we tracked families of 4 cells that were from the middle towards the end of the cell cycle for 3 frames.

Data analysis

The objective of the mathematical model was to capture the key processes that underlie the observed correlation patterns of transcriptional reporter levels within lineage-related cells. We first describe the stochastic model of single-cell dynamics that captures noisy fluctuations amongst pairs of cells, and then describe how cell-specific parameters are connected via a population model. Parameter inference of the model is performed for each gene using Markov Chain Monto Carlo within a Bayesian framework.

Single-cell reporter level dynamics

For two sister cells labelled $i \in \{1,2\}$, we model the total production rate of the bioluminescent reporter with the variable *S*, which we interpret as a total transcriptional activity. The dynamics of the transcriptional activity for cell *i* follows the stochastic differential equation

$$\frac{dS_i(t)}{dt} = -\frac{1}{\tau_S}(S_i(t) - \mu_i) + \epsilon_i(t), \quad (1)$$

where the first term describes the relaxation to a cell-specific mean level (μ_i). The time scale τ_S controls the rate at which *S* fluctuates (*i.e.* slow or rapid fluctuations for large or small τ_S, respectively). The distribution of the cell-specific means μ_i is further modelled at the population level (described below). The term $\epsilon_i(t)$ models biological noise, for example arising from the stochastic biochemical processes occurring in single cells, and acts to continuously deliver random perturbations to the transcriptional activity. $\epsilon_i(t)$ is modelled as Gaussian white noise with zero mean and variance

bioRxiv preprint doi: https://doi.org/10.1101/411447; this version posted September 7, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

$$\mathrm{Cov}[\epsilon_i(t)\epsilon_i(t')] = \frac{2{\sigma_{S,i}}^2}{\tau_S}\delta(t-t'),$$

where ${\sigma_{S,i}}^2$ controls the size of the perturbations on S_i and is cell specific. In the stationary state, the covariance of the transcriptional activity is $\mathrm{Cov}[S_i(t)S_i(t)] = {\sigma_{S,i}}^2$. To account for the similarity in dynamics observed in sister cells we introduced a correlation parameter ρ linking the noise terms of two sisters :

$$\mathrm{Cov}[\epsilon_1(t)\epsilon_2(t')] = \rho\frac{\sigma_{S,1}\sigma_{S,2}}{\tau_S}\delta(t-t').$$

ρ can vary between -1 and 1. When ρ=0 the cells are fluctuating independently and have uncorrelated trajectories, but when ρ>0 (or ρ<0) the perturbations are correlated (or anti-correlated) between the cells.

The measured total transcriptional reporter level is modelled with the variable *R*. The reporter *R* is produced at rate *S* and is degraded with half-life τ_R

$$\frac{dR_i(t)}{dt} = \frac{\ln(2)}{\tau_R}\big(S_i(t) - R_i(t)\big) + \eta_i(t), \quad (2)$$

where $\eta_i(t)$ corresponds to noise at the reporter level. Note that to save parameters, mRNA is not explicitly modelled; we estimated the net reporter half-life (which thus depends on both the mRNA and protein half-life) by blocking transcription with actinomycin D and by fitting a first order exponential decay to the decrease in reporter levels (values shown in in Supp Fig. 4). $\eta_i(t)$ is taken as Gaussian white noise, and represents effective noise combining both molecular fluctuations in reporter levels as well as experimental noise. $\eta_i(t)$ is assumed to be independent between two cells. The variance of the reporter Gaussian white noise terms is given by

$$\mathrm{Cov}[\eta_i(t)\eta_i(t')] = \frac{2{\sigma_{R,i}}^2\ln(2)}{\tau_R}\delta(t-t'),$$

where ${\sigma_{R,i}}^2$ controls the cell-specific variance in reporter levels. Our model consists of a system of two linear stochastic differential equations (Equations 1 and 2), and if the initial

bioRxiv preprint doi: https://doi.org/10.1101/411447; this version posted September 7, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

conditions of the two variables are normally distributed then the model can thus be analysed within the framework of Gaussian processes (Supplementary Note 1).

Initial conditions

As for any system of SDEs the distribution for the initial conditions at time *t*=0 (*i.e.* following cell division) need to be specified. Here, the distributions over *R* and *S* were taken from the steady state solution of the model, with the modification that the *R* variable was divided by two, reflecting the fact that we measure the total levels of transcriptional reporter, which are approximately halved at cell division (Supplementary Note 1).

Population level

The above model (Equations 1 and 2) introduced cell-specific mean levels μ_i (Fig. 2d), as well as cell-specific transcriptional noise (ϵ_i) and noise in reporter dynamics (η_i). Across the population, we assumed that these quantities are log-normally distributed. For example, this captures the heavy tails of expression levels (e.g. data in Fig. 2a show few high-expressing cells). Moreover, we introduce a parameter λ representing the correlation in mean transcriptional activities between sister cells (*i.e.* the population correlation between μ_1 and μ_2 for pairs of cells across the population). Together the population distributions of $\mu_i, \sigma_{S,i}$ and $\sigma_{R,i}$ are parameterised as follows

$$\log\binom{\mu_1}{\mu_2} = N\left(\begin{bmatrix} m \\ m \end{bmatrix}, \begin{bmatrix} s^2 & \lambda s^2 \\ \lambda s^2 & s^2 \end{bmatrix}\right),$$

$$\log\binom{\sigma_{S,1}}{\sigma_{S,2}} = N\left(\begin{bmatrix} \Lambda_S \\ \Lambda_S \end{bmatrix}, \begin{bmatrix} \Sigma_S & 0 \\ 0 & \Sigma_S \end{bmatrix}\right),$$

$$\log\binom{\sigma_{R,1}}{\sigma_{R,2}} = N\left(\begin{bmatrix} \Lambda_R \\ \Lambda_R \end{bmatrix}, \begin{bmatrix} \Sigma_R & 0 \\ 0 & \Sigma_R \end{bmatrix}\right).$$

bioRxiv preprint doi: https://doi.org/10.1101/411447; this version posted September 7, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

where *N* stands for a 2-variable normal distribution. Thus, the population mean of log $\mu_i, \sigma_{S,i}$ and $\sigma_{R,i}$ are parameterised with m, Λ_S and Λ_R, respectively. The intercellular population variances of $\mu_i, \sigma_{S,i}$ and $\sigma_{R,i}$ are parameterised with s^2, Σ_S and Σ_R, respectively.

Parameter inference

Because of the population parameters the full model is a so-called hierarchical model. Parameter inference for each gene was performed within a Bayesian framework. The joint posterior distribution over all parameters (of all cell pairs of a given gene) was inferred using Hamiltonian Markov Chain Monte Carlo (MCMC) sampling, which uses the gradients of the posterior to improve the efficiency of the sampling. We discarded the first 200 samples of each chain as burn-in and then obtained 2500 samples from 4 parallel chains. The inference procedure (including the priors for all parameters) is fully described in Supplementary Note 1. Data and code to generate all figures will be available on a public repository.

References

1. D'Urso, A. & Brickner, J. H. Mechanisms of epigenetic memory. *Trends Genet.* **30,** 230–236 (2014).

2. Raj, A. & van Oudenaarden, A. Nature, nurture, or chance: stochastic gene expression and its consequences. *Cell* **135,** 216–26 (2008).

3. Coulon, A., Chow, C. C., Singer, R. H. & Larson, D. R. Eukaryotic transcriptional dynamics: from single molecules to cell populations. *Nat. Rev. Genet.* **14,** 572–584 (2013).

4. Kærn, M., Elston, T. C., Blake, W. J. & Collins, J. J. Stochasticity in gene expression: from theories to phenotypes. *Nat. Rev. Genet.* **6,** 451–464 (2005).

5. Bruggeman, F. J. & Teusink, B. Living with noise: on the propagation of noise from molecules to phenotype and fitness. *Curr. Opin. Syst. Biol.* **8,** 144–150 (2018).

bioRxiv preprint doi: https://doi.org/10.1101/411447; this version posted September 7, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

6. Padovan-Merhar, O. *et al.* Single Mammalian Cells Compensate for Differences in Cellular Volume and DNA Copy Number through Independent Global Transcriptional Mechanisms. *Mol. Cell* **58,** 339–352 (2015).

7. Kempe, H., Schwabe, A., Cremazy, F., Verschure, P. J. & Bruggeman, F. J. The volumes and transcript counts of single cells reveal concentration homeostasis and capture biological noise. *Mol. Biol. Cell* **26,** 797–804 (2015).

8. Márquez-Jurado, S. *et al.* Mitochondrial levels determine variability in cell death by modulating apoptotic gene expression. *Nat. Commun.* **9,** (2018).

9. Battich, N., Stoeger, T. & Pelkmans, L. Control of Transcript Variability in Single Mammalian Cells. *Cell* **163,** 1596–1610 (2015).

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

11. Skinner, S. O. *et al.* Single-cell analysis of transcription kinetics across the cell cycle. *Elife* (2016).

12. Zopf, C. J., Quinn, K., Zeidman, J. & Maheshri, N. Cell-Cycle Dependence of Transcription Dominates Noise in Gene Expression. *PLoS Comput. Biol.* **9,** 1–12 (2013).

13. Snijder, B. *et al.* Population context determines cell-to-cell variability in endocytosis and virus infection. *Nature* **461,** 520–523 (2009).

14. Snijder, B. & Pelkmans, L. Origins of regulated cell-to-cell variability. *Nat. Rev. Mol. Cell Biol.* **12,** 119–125 (2011).

15. Suter, D. M. *et al.* Mammalian Genes are Transcripbed with Widely Different Bursting Kinetics. *Science (80-.).* **332,** 472–474 (2011).

16. Zoller, B., Nicolas, D., Molina, N. & Naef, F. Structure of silent transcription intervals and noise characteristics of mammalian genes. *Mol. Syst. Biol.* **11,** 823 (2015).

bioRxiv preprint doi: https://doi.org/10.1101/411447; this version posted September 7, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

17. Nicolas, D., Phillips, N. E. & Naef, F. What shapes eukaryotic transcriptional bursting? *Mol. BioSyst.* **13,** 1280–1290 (2017).

18. Nikolov, E. N., Dabeva, M. D. & Nikolov, T. K. Turnover of ribosomes in regenerating rat liver. *Int. J. Biochem.* **15,** 1255–1260 (1983).

19. Muramoto, T., Müller, I., Thomas, G., Melvin, A. & Chubb, J. R. Methylation of H3K4 Is Required for Inheritance of Active Transcriptional States. *Curr. Biol.* **20,** 397–406 (2010).

20. Ferraro, T. *et al.* Transcriptional Memory in the Drosophila Embryo. *Curr. Biol.* **26,** 212–218 (2016).

21. Sigal, A. *et al.* Variability and memory of protein levels in human cells. *Nature* **444,** 643–646 (2006).

22. Corre, G. *et al.* Stochastic fluctuations and distributed control of gene expression impact cellular memory. *PLoS One* **9,** 1–22 (2014).

23. Cambridge, S. B. *et al.* Systems-wide proteomic analysis in mammalian cells reveals conserved, functional protein turnover. *J. Proteome Res.* **10,** 5275–5284 (2011).

24. Uren, A. G. *et al.* A high-throughput splinkerette-pcr method for the isolation and sequencing of retroviral insertion sites. *Nat. Protoc.* **4,** 789–798 (2009).

25. Hsiung, C. C. S. *et al.* A hyperactive transcriptional state marks genome reactivation at the mitosis-G1 transition. *Genes Dev.* **30,** 1423–1439 (2016).

26. Itzkovitz, S., Blat, I. C., Jacks, T., Clevers, H. & Van Oudenaarden, A. Optimality in the development of intestinal crypts. *Cell* **148,** 608–619 (2012).

27. Hormoz, S. *et al.* Inferring Cell-State Transition Dynamics from Lineage Trees and Endpoint Single-Cell Measurements. *Cell Syst.* **3,** 419–433.e8 (2016).

28. Robert, L. *et al.* Pre-dispositions and epigenetic inheritance in the Escherichia coli lactose operon bistable switch. *Mol. Syst. Biol.* **6,** (2010).

bioRxiv preprint doi: https://doi.org/10.1101/411447; this version posted September 7, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

29. Veening, J.-W. *et al.* Bet-hedging and epigenetic inheritance in bacterial cell development. *Proc. Natl. Acad. Sci. USA* **105,** 4393–4398 (2008).

30. van Vliet, S. *et al.* Spatially Correlated Gene Expression in Bacterial Groups: The Role of Lineage History, Spatial Gradients, and Cell-Cell Interactions. *Cell Syst.* **6,** 496–507.e6 (2018).

31. Kuchen, E. E., Becker, N., Claudino, N. & Thomas, H. Long-range memory of growth and cycle progression correlates cell cycles in lineage trees. *bioRxiv* 1–16 (2018). doi:https://doi.org/10.1101/373258

32. Sandler, O. *et al.* Lineage correlations of single cell division time as a probe of cell-cycle dynamics. *Nature* **519,** 468–471 (2015).

33. Feigelman, J. *et al.* Analysis of Cell Lineage Trees by Exact Bayesian Inference Identifies Negative Autoregulation of Nanog in Mouse Embryonic Stem Cells. *Cell Syst.* **3,** 480–490.e13 (2016).

34. Kuzmanovska, I., Milias-Argeitis, A., Zechner, C. & Khammash, M. Parameter inference method for stochastic single-cell dynamics from tree-structured data. 1–14 (2017).

35. Bintu, L. *et al.* Dynamics of epigenetic regulation at the single cell level. *Science (80-.).* **351,** 720–724 (2016).

36. Deluz, C. *et al.* A role for mitotic bookmarking of SOX2 in pluripotency and differentiation. *Genes Dev.* **30,** 2538–2550 (2016).

37. Sakaue-Sawano, A. *et al.* Visualizing Spatiotemporal Dynamics of Multicellular Cell-Cycle Progression. *Cell* **132,** 487–498 (2008).

38. Kent, W. J. BLAT — The BLAST -Like Alignment Tool. *Genome Res.* **12,** 656–664 (2002).

39. Nagaoka, M. *et al.* E-cadherin-coated plates maintain pluripotent ES cells without colony formation. *PLoS One* **1,** 1–7 (2006).

bioRxiv preprint doi: https://doi.org/10.1101/411447; this version posted September 7, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

40. Bensaude, O. Inhibiting eukaryotic transcription: Which compound to choose? How to evaluate its activity? *Transcription* **2,** 103–108 (2011).

41. Newman, R. H. & Zhang, J. Fucci: Street Lights on the Road to Mitosis. *Chem. Biol.* **15,** 97–98 (2008).

Acknowledgements

Work in the Naef lab was supported by the EPFL. Work in the Suter lab was supported by the Swiss National Science Foundation (grant #PP00P3_144828).

Author Contributions

Conceptualization: A.M., D.M.S and F.N.; Methodology: A.M., N.E.P, S.O., F.N. and D.M.S.; Software, N.P and F.N.; Formal Analysis, N.P and F.N; Investigation, A.M and D.M.S; Resources, D.M.S. and F.N., Writing – Original Draft, N.E.P, D.M.S and F.N.; Writing – Review & Editing, N.E.P, A.M., F.N., D.M.S; Funding Acquisition, D.M.S., and F.N.; Supervision, D.M.S. and F.N.

Competing Interests

The authors declare no competing financial interests.

Figure Legends

Fig. 1. Monitoring the propagation of transcriptional activity in proliferating cells

(a) A cell from the *Rbpj* reporter line progressing through two cell cycles. Luminescent cell nuclei are tracked manually.

(b) Representation of events in (a).

(c) Schematic of the transcriptional activity profiles over cell lineages in genes with short or long memory.

bioRxiv preprint doi: https://doi.org/10.1101/411447; this version posted September 7, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Fig. 2. Single-cell transcriptional reporter profiles and their correlation between sister cells.

(a) Single-cell transcriptional reporter time series (total intensity per cell) for two genes (top: *Pgk*, 59 pair of cells; bottom: *Dstn*, 50 pairs), measured from one cell division to the next (time is expressed in % of cell cycle time). Cells are colour-coded according to the ranking of the initial reporter level within the population. Dotted black line: population mean.

(b) The top three cells with the highest/lowest initial reporter levels.

(c) Examples of three pairs of sister cells (sister cells have the same colour).

(d) The decrease in correlation between sister cells over the cell cycle. Green: correlation between sister cells; red: correlation between random cells, where each cell is matched with a non-sister with the nearest initial values. Error bars denote standard deviations obtained with bootstrap sampling.

Figure 3. A hierarchical model of single-cell reporter dynamics in pairs of sister cells.

(a) Single-cell dynamics are modelled probabilistically using stochastic differential equations (Methods). Each cell has a transcriptional activity (*S*) and a bioluminescent reporter (*R*) variable, where *S* controls the production of *R*.

(b) To account for stochastic fluctuation, both *S* and *R* are perturbed by noise terms ε and η, respectively. The transcriptional noise (ϵ) experienced in the two sister cells is correlated with parameter ρ, which describes whether sister cell dynamics are independent ($\rho = 0$) or if they share a similar shape over the cell cycle ($\rho > 0$).

(c) The mean level of *S* is cell-specific and denoted by μ_i for cell *i*, and the strength of the noise terms for *S* and *R* are also cell-specific and are denoted by $\sigma_{S,i}$ and $\sigma_{R,i}$, respectively. The distribution of cell-specific parameters μ_i, $\sigma_{S,i}$ and $\sigma_{R,i}$ are described at the population level with log-normal distributions. *s* describes the population level variability in cell-specific means.

bioRxiv preprint doi: https://doi.org/10.1101/411447; this version posted September 7, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

(d) The correlation of mean transcriptional levels between sister cells is quantified with λ.

Figure 4. Sister cell correlation over the cell cycle is controlled by gene-dependent variability of cell-specific means and similarity in dynamics.

(a) Posterior distributions (shown as boxplots) of the coefficient of variation (CV) for cell-specific means, calculated from the posterior distributions of *s* and *m*. The boxplots represent the 25th, median (50th) and 75th percentiles of the posterior distribution and the whiskers represent the 5th and 95th percentiles.

(b) Posterior distributions of the correlation of mean transcriptional activity between sister cells (λ).

(c) λ correlates with CV of cell-specific means (crosses denote mean posterior values for each gene).

(d) The inferred posterior probability distribution of the similarity in dynamics (ρ) between sister cells.

(e) The inferred posterior probability distribution of the similarity in dynamics (ρ) between randomised cells, where the randomisation ensures that cells have the same correlation in cell-cycle lengths as sister cells.

(f) The inferred posterior probability distribution of the similarity in dynamics ρ for both sister cells and non-sister cells with the same average distance as non-sister cells.

(g-i) Decrease in correlation between sister cells over the cell cycle. Green - the evolution of the sister-cell correlation over the cell cycle from the data, where time is expressed in % of cell cycle time. Red - the parameter posterior means for each gene are used to predict the evolution of sister-sister correlation over the cell cycle from the model, which is normalised to the average cell cycle length (13.5 hours). Yellow - the correlation between sisters is recalculated with ρ=0. Blue - the correlation between sisters is recalculated with *s*=0, which removes cell-specific means from the model.

bioRxiv preprint doi: https://doi.org/10.1101/411447; this version posted September 7, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Figure 5. The mean transcriptional activities are highly correlated between mother and daughter cells while the similarity in dynamics is weak

(a) Examples of two pairs of mother and daughter cells from the *pGK* gene. Red and blue represent different pairs of cells.

(b) Posterior distribution (show as boxplots) of the correlation of mean transcriptional levels (λ) between mother and daughter cells. The box represents the 25th, median (50th) and 75th percentiles of the posterior distributions and the whiskers represent the 5th and 95th percentiles.

(c) The inferred posterior probability distribution of the similarity in dynamics (ρ) between mother and daughter cells.

Figure 6. Families of related cells have similar transcriptional levels

(a-c) The distribution of transcriptional reporter levels in different families. For each family, crosses represent averaged luminescence levels from three frames preceding nuclear envelope breakdown in individual cells. Circles represent the mean of the family.

(d) Boxplot showing family means (circles in (a-c)). The CV is calculated by dividing the standard deviation of family means by the average of the family means. The box represents the 25th, median (50th) and 75th percentiles of the posterior distributions and the whiskers represent the 5th and 95th percentiles.

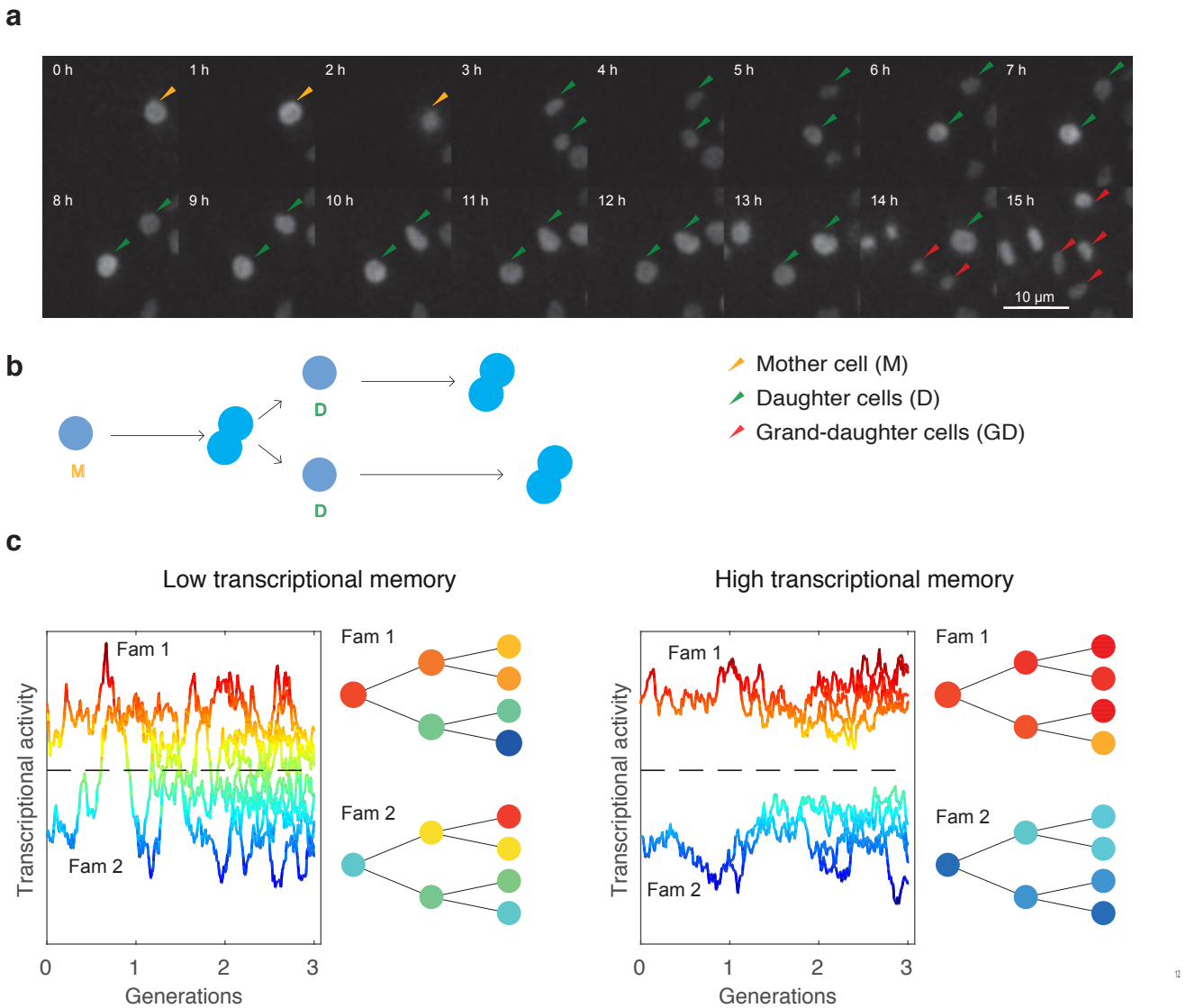
(d) For each family in (a-c) the CV is calculated by dividing the standard deviation of all cells within the family by the family mean.

(c) The strength of the transcriptional noise is quantified by averaging the σ_S over all cells and dividing by the global mean m.

(d) Samples of ES cell colonies resulting from ~ five cell divisions for *pGK, Jam2* and *Dstn*. Cells were seeded at low density and colonies were imaged 60 hours later.

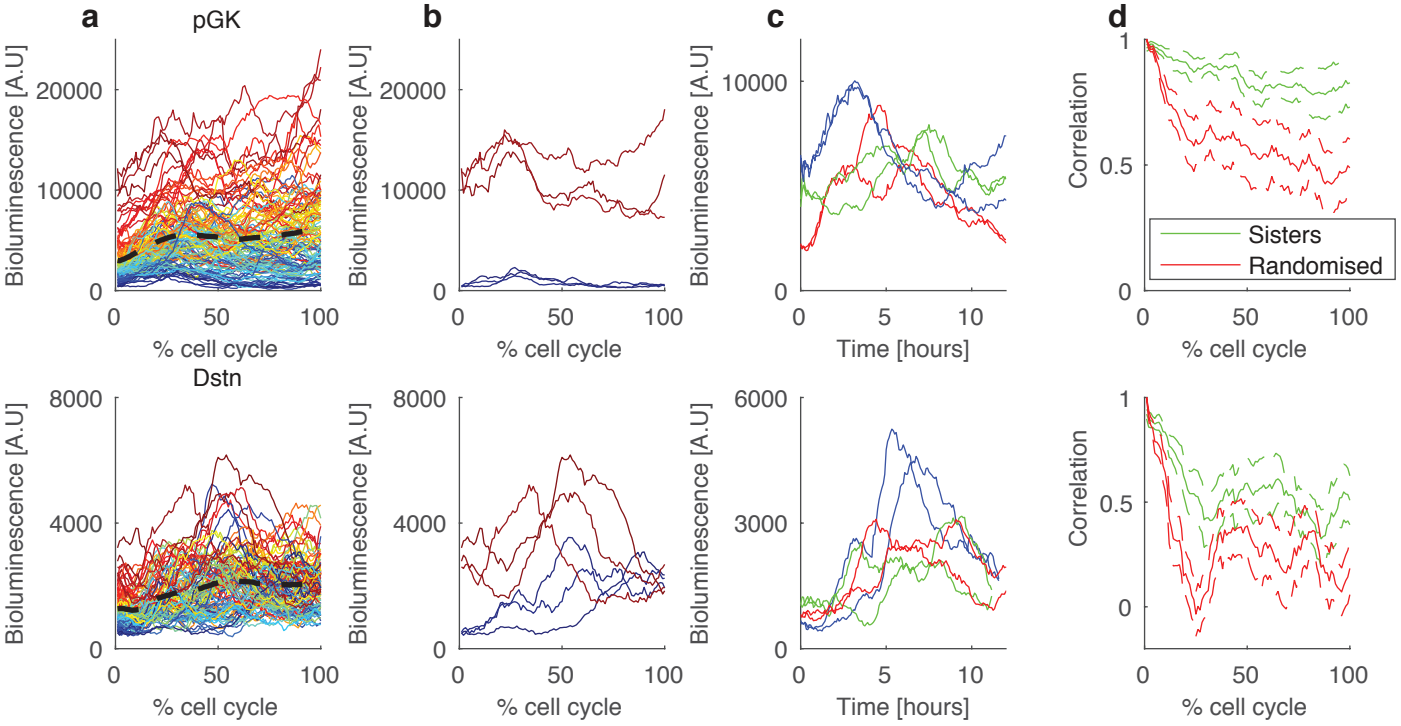
bioRxiv preprint doi: https://doi.org/10.1101/411447; this version posted September 7, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Figure 1



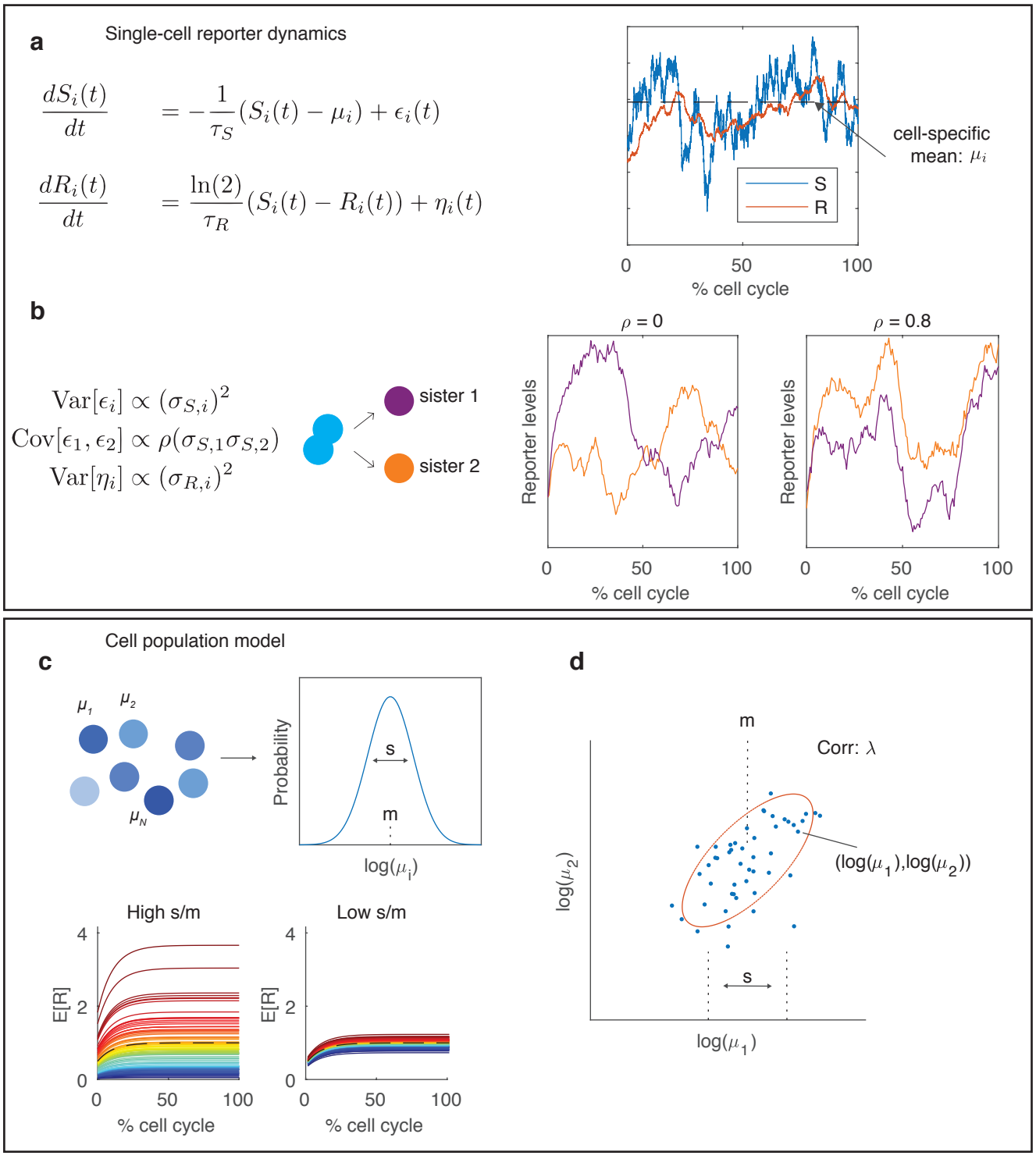
bioRxiv preprint doi: https://doi.org/10.1101/411447; this version posted September 7, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Figure 2



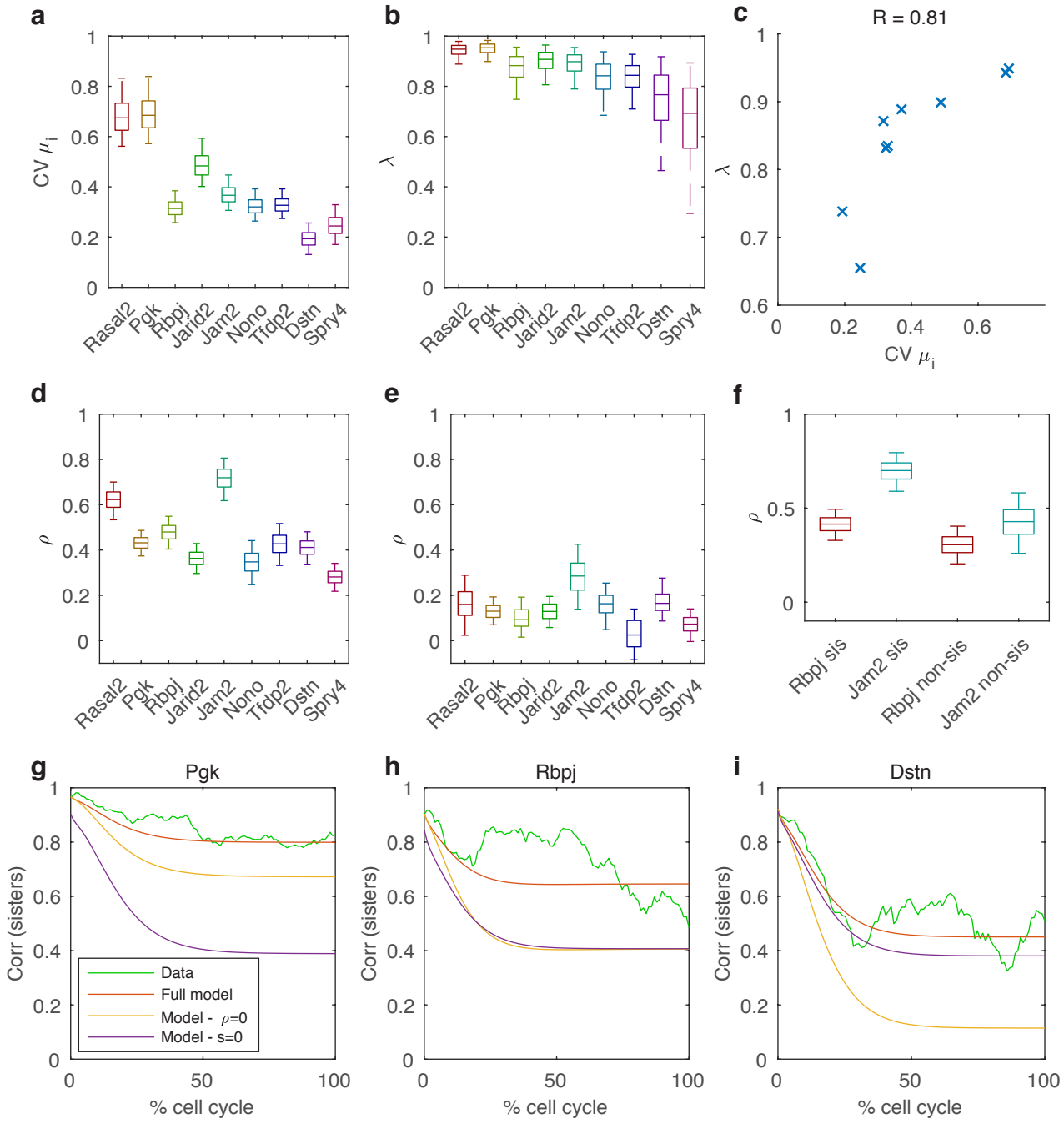
bioRxiv preprint doi: https://doi.org/10.1101/411447; this version posted September 7, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Figure 3



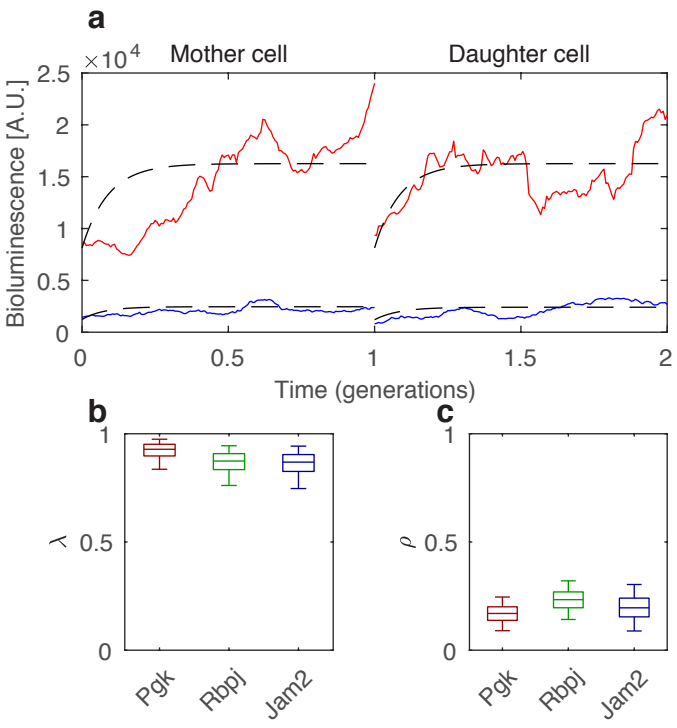
bioRxiv preprint doi: https://doi.org/10.1101/411447; this version posted September 7, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Figure 4



bioRxiv preprint doi: https://doi.org/10.1101/411447; this version posted September 7, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Figure 5



bioRxiv preprint doi: https://doi.org/10.1101/411447; this version posted September 7, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

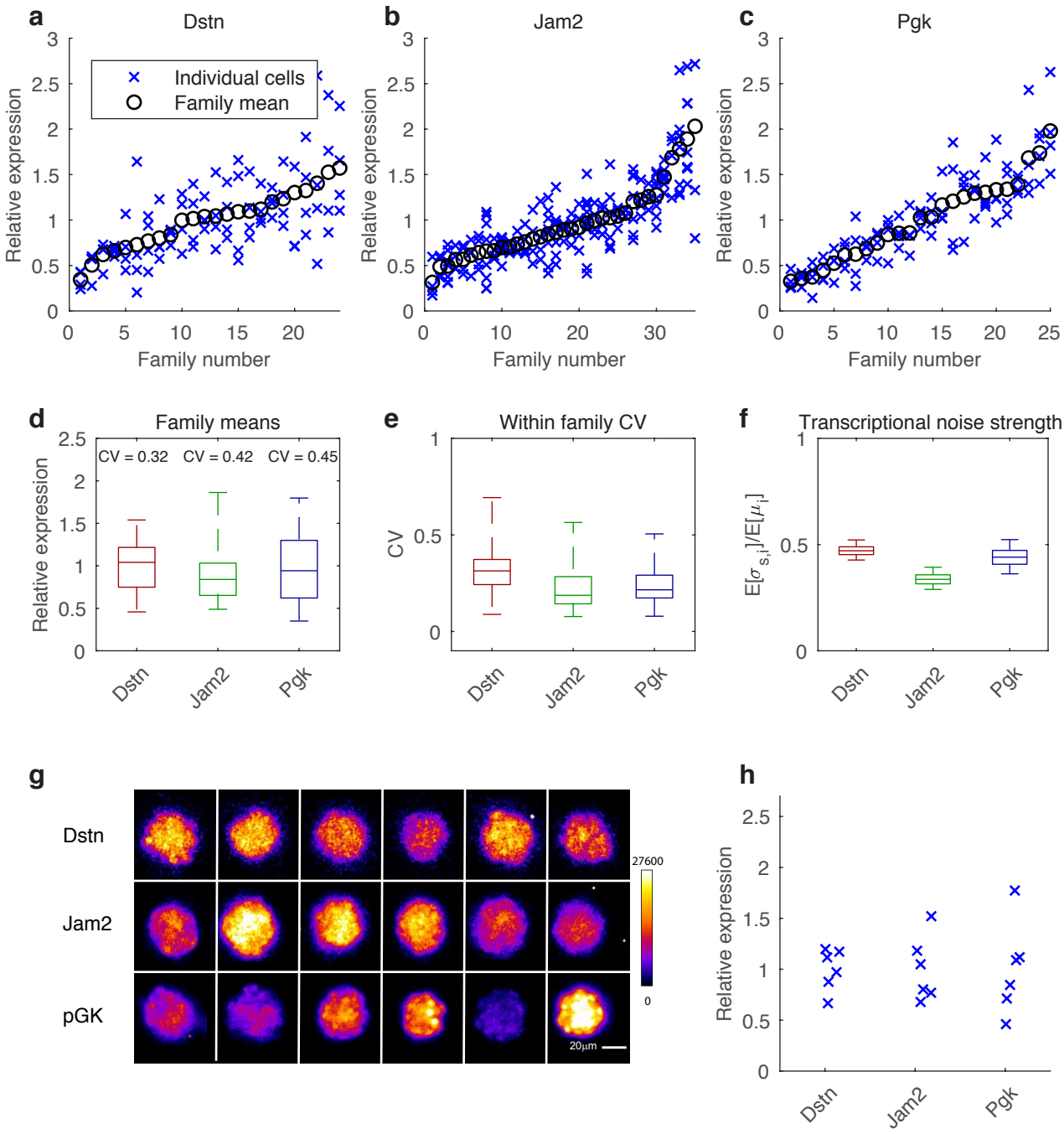



Figure 6