

SUPPLEMENTARY INFORMATION

HERITABLE STOCHASTIC SWITCHING REVEALED BY SINGLE CELL GENEALOGY

Benjamin B. Kaufmann^{1,2}, Qiong Yang¹, Jay Mettetal¹, Alexander van Oudenaarden¹

¹*Department of Physics, Massachusetts Institute of Technology,
Cambridge, Massachusetts 02139, USA*

²*Division of Engineering and Applied Sciences, Harvard University,
Cambridge, Massachusetts 02138, USA*

Contents

<u>S1 Network Description</u>	2
<u>S2 Additional Measurements</u>	2
<u>S2.A Growth rate for colonies</u>	2
<u>S2.B Doubling times for individual measured cells</u>	3
<u>S2.C Demonstration of rapid fluorophore maturation</u>	4
<u>S3 Data Analysis</u>	6
<u>S3.A Corrections to the 1-D data required from finite viewing window</u>	6
<u>S3.B Corrections to the 2-D scatter data required from finite viewing window</u>	8
<u>S3.C Calculation of the cumulative percent switched</u>	9
<u>S3.D Calculation of the conditional percent switched</u>	9
<u>S3.E Calculation of mean squared deviation</u>	9
<u>S3.F Generation of the Poisson model</u>	10
<u>S3.F Confidence intervals for Monte Carlo model</u>	12
<u>Supplementary References</u>	13
<u>Supplementary Movie 1</u>	14

S1 Network Description

Elaborate regulatory machinery has evolved to help cells focus on the carbon sources that maximize their growth rate in a particular environment. The much studied galactose utilization network (GAL) is a model for such decision/metabolic pathways. The network, comprising roughly a dozen genes, contains two positive and one negative transcriptionally-mediated feedback loops nested one within the next. The positive loops make possible a two-state ‘all-or-nothing’ arrangement, while the negative loop is thought to stabilize cells in one of these two states, called ON and OFF.

The ON state’s defining characteristic is high activity of Gal4p, a transcriptional activator constitutively bound to the promoter of many GAL genes¹. Gal4p activity occurs only in the absence of a dominant repressor Gal80p, which becomes sequestered to the cytoplasm in the presence of active Gal3p². *GAL3*, in turn, is positively regulated by the level of active Gal4p, closing the first positive feedback. Thus Gal3 presents something of a chicken-or-the-egg situation: high expression of Gal3p leads to the activation of the GAL network, and activation of the GAL network leads to high expression of Gal3p.

A membrane protein Gal2p, whose expression is also regulated by Gal4p, forms a second positive feedback loop by importing galactose into the cell and consequently activating Gal3p. Finally, *GAL80* negatively regulates its own production, again by suppressing the activity of Gal4p. For cells in the OFF state, the situation is reversed with low Gal4p activity, presence of repressor Gal80p at the GAL promoters, and a depletion of (active) Gal3p and Gal2p in the cytoplasm.

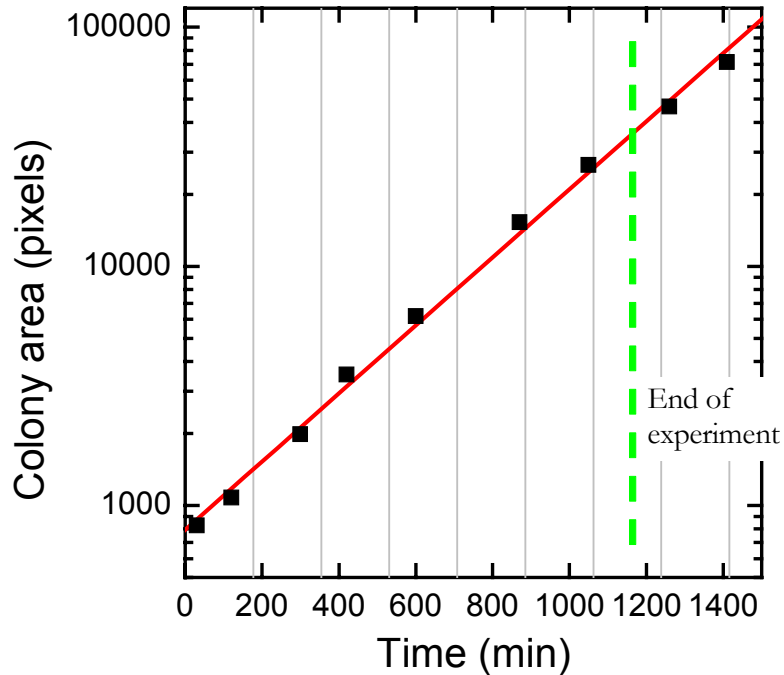
In wild-type cells, transitions between the OFF and ON states can be forced by changing levels of inducers (e.g., galactose) or repressors (e.g., glucose) in the surrounding environment. At intermediate levels of inducer, the OFF or ON status of the cell depends on the history of the media in which their ancestors grew, indicating significant hysteresis. This hysteretic behavior buffers against switching too rapidly between states, perhaps to avoid the metabolic cost incurred.

S2 Additional Measurements

S2.A Growth rate for colonies

To calculate the effective bulk growth rate for the cell population, we measured the total area of micro-colonies as they grew from one cell to several hundred. This area grew exponentially with an average doubling time of 177 ± 24 min and remained constant over 24 hours (Supplementary Figure 1), although slight day-to-day variations were observed (CV=0.14). For times longer than about 30-35 hours, the slide becomes confluent with

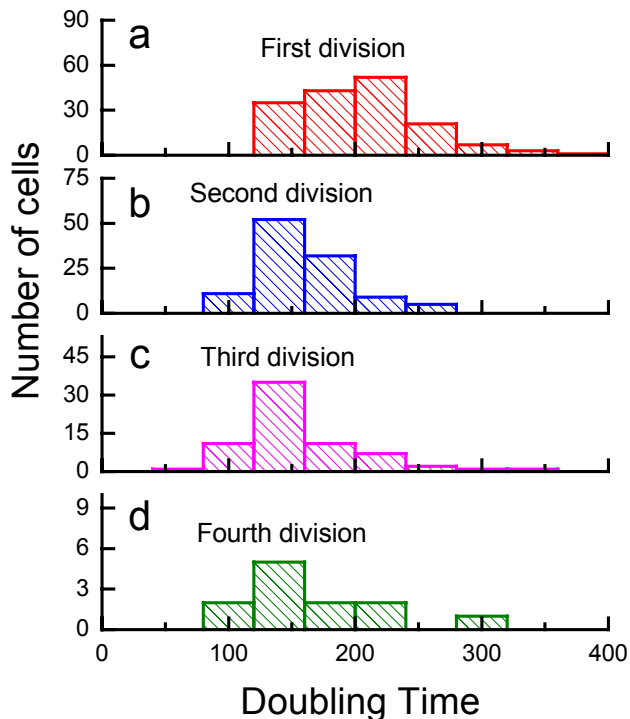
cells. Typically we cease experiments after about 17 hours when cells become densely packed or begin to get pushed out of the focal plane and into the agar pad.



Supplementary Figure 1. Growth for selected cell colony in measurement chamber remains constant for longer than our typical measurement period. Black points are measured areas. Red curve is a linear fit to the data. The quoted figure 177 minutes comes from averaging four such curves measured in different days and/or different points on the same day. The green dashed line is the typical point at which the experiment was terminated.

S2.B Doubling times for individual measured cells

As cells grew and divided, we recorded the times when each cell septated from its daughter cell. Newly born cells took on average 205 ± 58 minutes to give birth to their first daughter cell. Subsequent buddings, however, require only 157 ± 51 minutes for the subsequent 2-4 divisions, consistent with measurements others have made³.



Supplementary Figure 2 Histograms of doubling times as a function of previous cell divisions. First division time (**a**) is measured as from the moment a new daughter septates until it septates in turn with its first daughter. The second, third, and fourth division times (**b-d**) are measured from the previous division time.

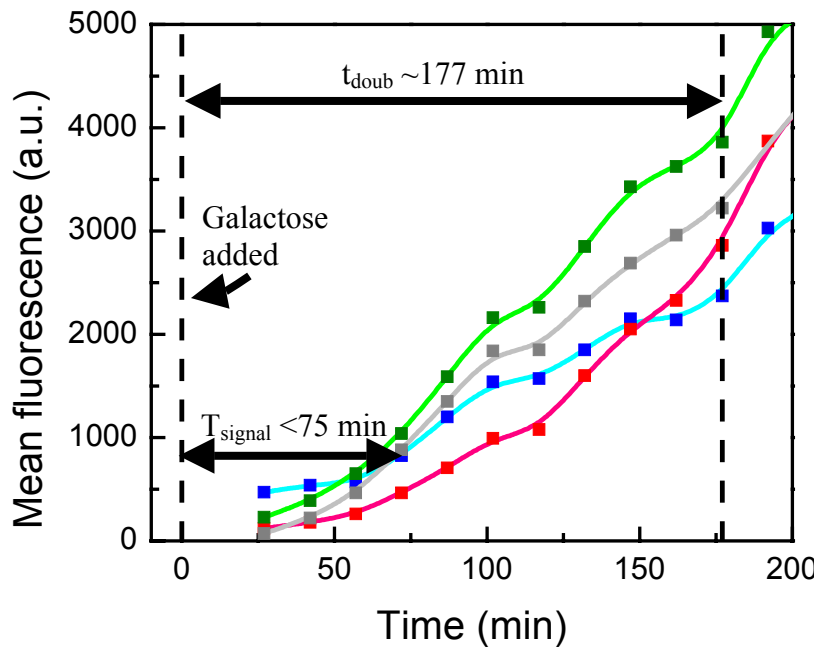
S2.C Demonstration of rapid fluorophore maturation

Once the P_{GALI} -YFP gene becomes active, the delay until fluorescence is detected is limited by rates of transcription, translation and chromophore maturation. To measure the duration of this process, and in particular to be certain that slow YFP maturation was not the reason for the long correlations, we grew cells (MA182) with the endogenous *GAL3* gene removed and replaced by a doxycycline-inducible one, tuned (at 0.05 μ g/ml dox) so that Gal3p levels were approximately the same as those in induced wild type cells. This strain was selected because the constitutive presence of Gal3p in the cytoplasm would enable the rapid activation of the GAL genes when galactose was added.

We first grew these cells overnight in media with doxycycline, 2% raffinose and no galactose. About 20 hours later, we spun down the cells and resuspended them in 2% raffinose and 1% galactose. Immediately we transferred the cells into our chamber and waited until cells settled in place (27 min). Then, at intervals of 15 minutes, we imaged their fluorescence. Between the first two planes, at 27 and 42 min respectively, we already see an increase in fluorescence, which from that point on continues to rise. A

conservative estimate then of the time from activated P_{GALI} -YFP gene to observable signal is about 75 minutes, far less than the 177 minute doubling time of a typical cell.

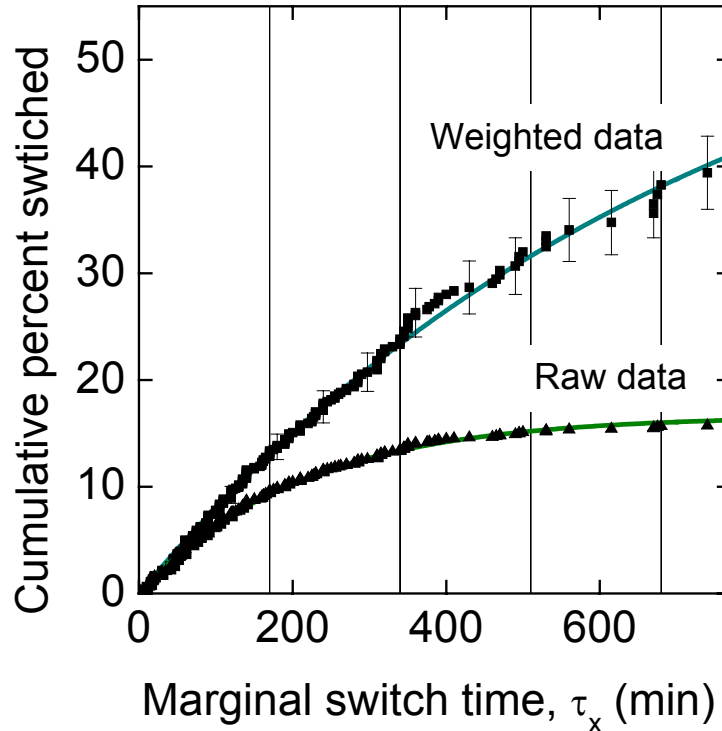
The opposite transitions, from ON to OFF, are difficult to localize temporally because at the instant the cell goes into the OFF state fluorescence levels are still high. Only by diluting or degrading the YFP can the OFF state be ascertained. However, because doubling times, and therefore growth rates, vary from cell to cell, conclusively proving a cell's network is OFF requires repeated measurements over long times. In contrast, switches from OFF to ON become rapidly apparent.



Supplementary Figure 3 Rapid maturation of YFP. The fluorescence trajectories of four randomly chosen control cells (MA182) shown in colored curves. Left dashed line indicates start of experiment when galactose is added to the media, and the right dashed line indicates a typical cell division. T_{sig} , at about 75 min, is the time passed until fluorescence becomes observable. t_{doub} is the mean cell doubling time of 177 min.

S3 Data Analysis

S3.A Corrections to the 1-D data required from finite viewing window

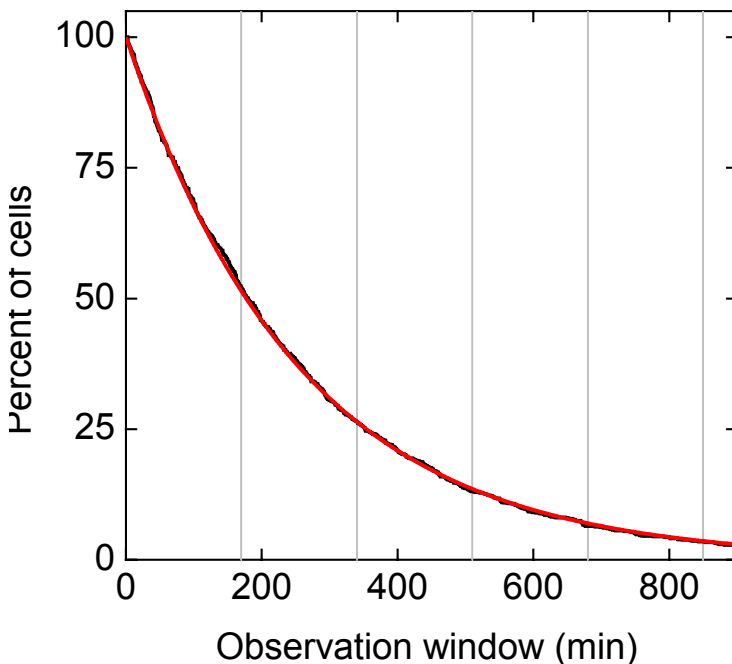


Supplementary Figure 4 The cumulative percent of cells that have switched is plotted against their marginal switch time. The triangles represent 251 switching cells, and the green line is an exponential fit. The asymptote represents the fact that only 16% of cells switch in the approximately 920 minute (>15 hour) duration of the experiment are observed to switch. Adjusting for finite viewing windows (black squares) generates a different picture more representative of what one would see if cells were tracked for an infinite period of time. The violet dashed line is an exponential fit to the data that forces the asymptote to be 100 percent switched, and in orange an unconstrained exponential fit (asymptote is 63.4% switched). Error bars are derived from a bootstrap analysis.

The raw distribution of τ_x over our entire population ($N=251$ switched cells), excluding original progenitor cells because no birth time is known, fits an exponential (Supplementary Figure 4, triangles and green line) with a time constant of 208 min. However, only 16% of cells were observed to switch. To interpret these results, we must account for the fact that cells exponentially divide throughout our measurement period. While this greatly expands the total number of observed cells, it inevitably means that some cells will be observed for longer periods than others. Thus very long switching events that span many cell generations, whenever they are observed, deserve special weight.

There are at least two ways to do this. The simplest way to correct for this variation in cell observation time is to measure the bulk growth rate (Supplementary Figure 1). An exponential curve decaying with that time constant (Supplementary Figure 5, red) approximates the relative likelihood of viewing a cell for a given duration. An exponential curve with that time constant is a reasonable estimate in the limit of a large number of division events (Supplementary Figure 5, red). Applying this corrective factor to the marginal switch times provides an estimate of what the switch time distribution would look like had we collected data for an unlimited amount of time (Fig. 3A, squares). Further assuming that all cells eventually would switch, we fit the data to an exponential and arrive at an effective transition rate of 0.14 switches per generation (Fig. 3A, violet dashed). The slight discrepancy between data and exponential fit is likely the result of some cells growing out of the focal plane.

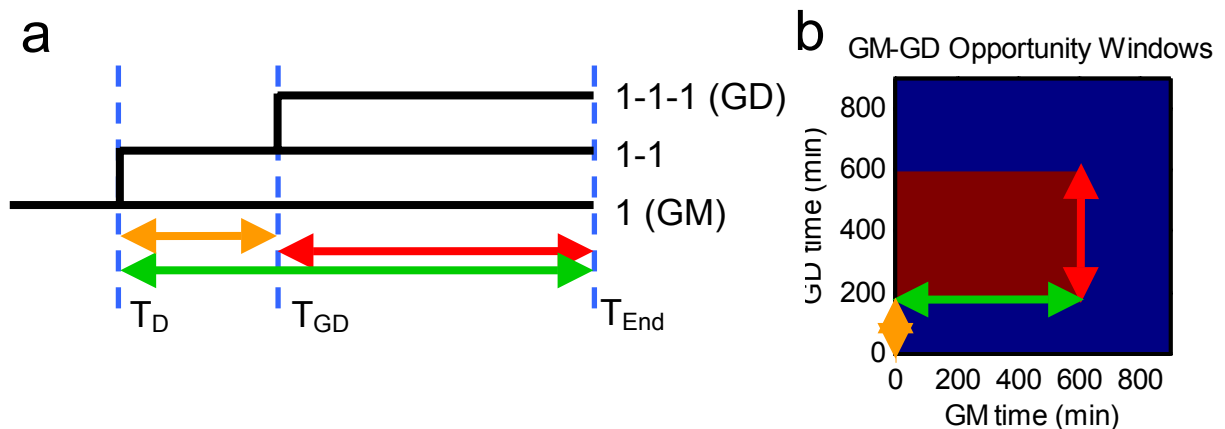
An alternative strategy is to use the actual doubling times for single cells. Drawing randomly from these distributions ($N=361$), we generate an ensemble of simulated family trees. To get the weighting factor for a switching event of length T , we count the number of cells in the simulated family trees that had times equal to or greater than T , and then divide that by the total number of cells. When very numerous (or very large) simulated family trees are used, this curve will closely approximate the exponential curve (Supplementary Figure 5, black). However, as explained in the following section, this latter strategy has the advantage of being easily extended to two dimensions.



Supplementary Figure 5 Corrective factors used to weigh data points according to their significance. In red, an exponential decay using only the bulk doubling time. In black, a similar curve built from the single-cell division times found in Supplementary Fig 2.

S3.B Corrections to the 2-D scatter data required from finite viewing window

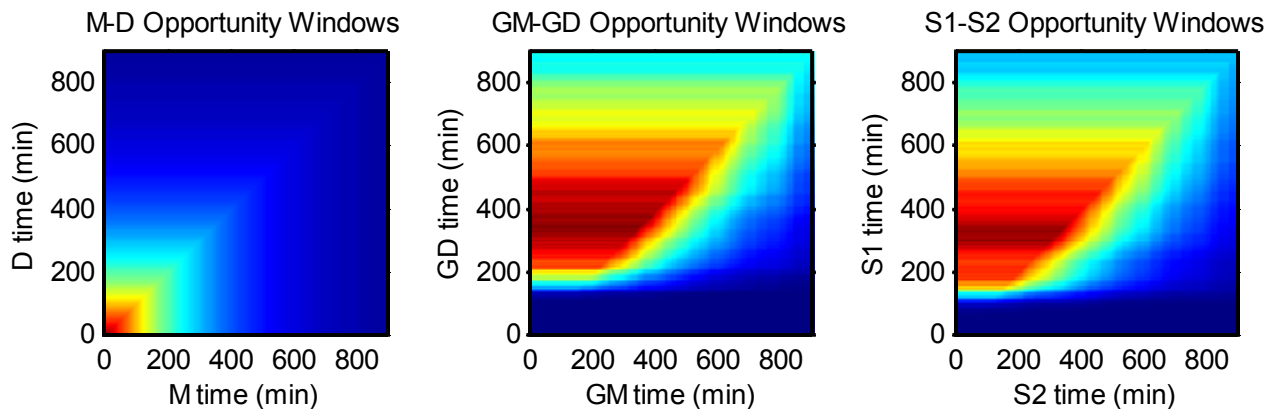
Like the one dimensional marginal switch data, the 2-D scatter plots must also be viewed in the context of finite experimental viewing times. Because not all regions in the x-y plane are sampled with the same frequency, points from infrequently sampled regions must be given special emphasis. Drawing from the measured doubling times ($N=361$), we generated an ensemble of simulated family trees. An example of one of these family trees is shown below (Supplementary Figure 6a).



Supplementary Figure 6 (a) Schematic for weighing the GM-GD opportunity windows. Black lines represent a growing family tree, with time flowing to the right. The orange arrow, running from T_D to T_{GD} , represents the time delay between the birth of the daughter (D) cell 1-1 and the granddaughter (GD) cell 1-1-1. The red arrow, running from T_{GD} to T_{End} , indicates the time when the GD cell had the opportunity to switch. The green arrow, from T_D to T_{End} , indicates when the grandmother (GM) cell had the opportunity. **(b)** The corresponding window of available switch times for the example family tree.

In the figure, cell 1 (the GM) gives birth to cell 1-1 at time T_D . Then, at time T_{GD} , cell 1-1 in turn gives birth to 1-1-1 (the GD). The experiment then ends at some time T_{End} . In order to observe GM-GD switch pair (cells 1 and 1-1-1), each of these cells must switch before T_{End} . Cell 1 can switch anywhere in the range (T_D, T_{End}) [Note: if cell 1 switched before T_D then it gives birth to ON cells which, by our definition, cannot switch.] Cell 1-1-1, on the other hand, can switch in the range from its birth to the end of the experiment (T_{GD}, T_{End}) . These two ranges define the rectangle of Supplementary Figure 6b.

To generate the full weighing distribution for all points, the process is then repeated for all possible GM-GD pairs. The resulting rectangles are added together and the result is Supplementary Figure 7. The inverse of these values is then used as the weighing factor. This strategy ensures that a pair of cells switching simultaneously in the microscope will always receive equal conditional switch times.



Supplementary Figure 7 Opportunity windows for M-D, GM-GD, and S1-S2 cell pairs. Warm (cool) colors represent areas where numerous (few) cells pairs had the opportunity to switch.

S3.C Calculation of the cumulative percent switched

Our measurement of the cumulative percent switched (see Fig 3a main text) requires an estimate of the fraction of cells that switch within our measurement period. However, when the colonies grow to a large size, the cells crowd together and can sometimes be difficult to measure. We therefore focused our attention on all cells born at or before the sixth generation (e.g., cells 1-2, 1-1-1-1-1-1 and 1-3-1-1 were included, but 1-7 was not). In addition, we excluded all original progenitor cells (i.e., cell 1) because these cells may act differently as a result of having been spun down in the centrifuge prior to placement in our chamber (this condition changes our result by less than 2%). From this large subpopulation, which comprises most of the cells, we counted the percentage (15.8%) of observed switching cells.

S3.D Calculation of the conditional percent switched

The conditional percent switched is calculated in the following way: First, we select all daughter cells that were found to switch before some time T . Second, we divide this population of daughter cells in to two groups, one with mothers who also switched at or before time T and another with mothers who switched after time T or did not switch. For each Mother-Daughter switch pair, a weight is assigned (as described above in S3.B) to take into account the likelihood of having viewed it. (In cases where the mother does not switch, the daughter's switch time is used to generate this weight.) Next, the weights of all the cells in the two groups are added together. The conditional percent switched is defined as the sum of the weights of the cell pairs that switched before T , divided by the sum of all the weights of all cell pairs.

S3.E Calculation of mean squared deviation

To generate the mean squared deviation of main text Fig. 4d, we first combined all observed MD, GMGD, and SS switch pairs (Figs. 4a-c) into a single data set. This was

done to increase statistical power. These data sets consisted of two columns, t_1 and t_2 , representing the older and younger cell switch times for 274 cell pairs. From this data set we used the bootstrap technique (draw with replacement) to generate 1000 alternate samples. Then, using each of these generated samples in turn, we transformed the data in the following way:

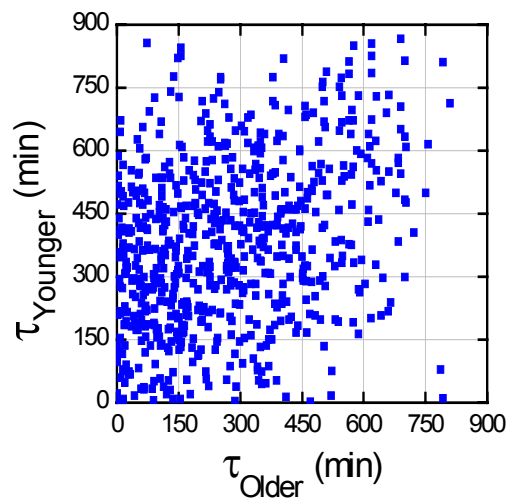
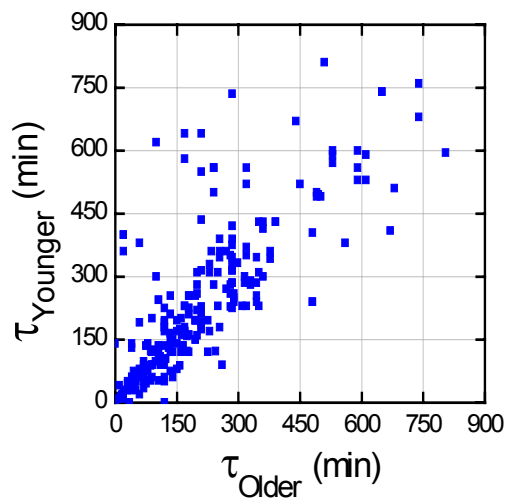
$$S = \frac{(t_1 + t_2)}{2}$$

$$D = |t_1 - t_2|^2$$

We next binned the transformed data according to S , using bins of sizes 25, 50, 100, and 150 min, with centers staggered by 75 min. At each center point, we calculated for all bootstrap samples and all bin sizes the average of the quantity D . Thus for each bin center we had four thousand estimates for the mean squared deviation (4 bin sizes times 1000 bootstraps). The mean and standard deviation of these four thousand estimates gives the mean and error shown on main text Fig. 4d.

S3.F Generation of the Poisson model

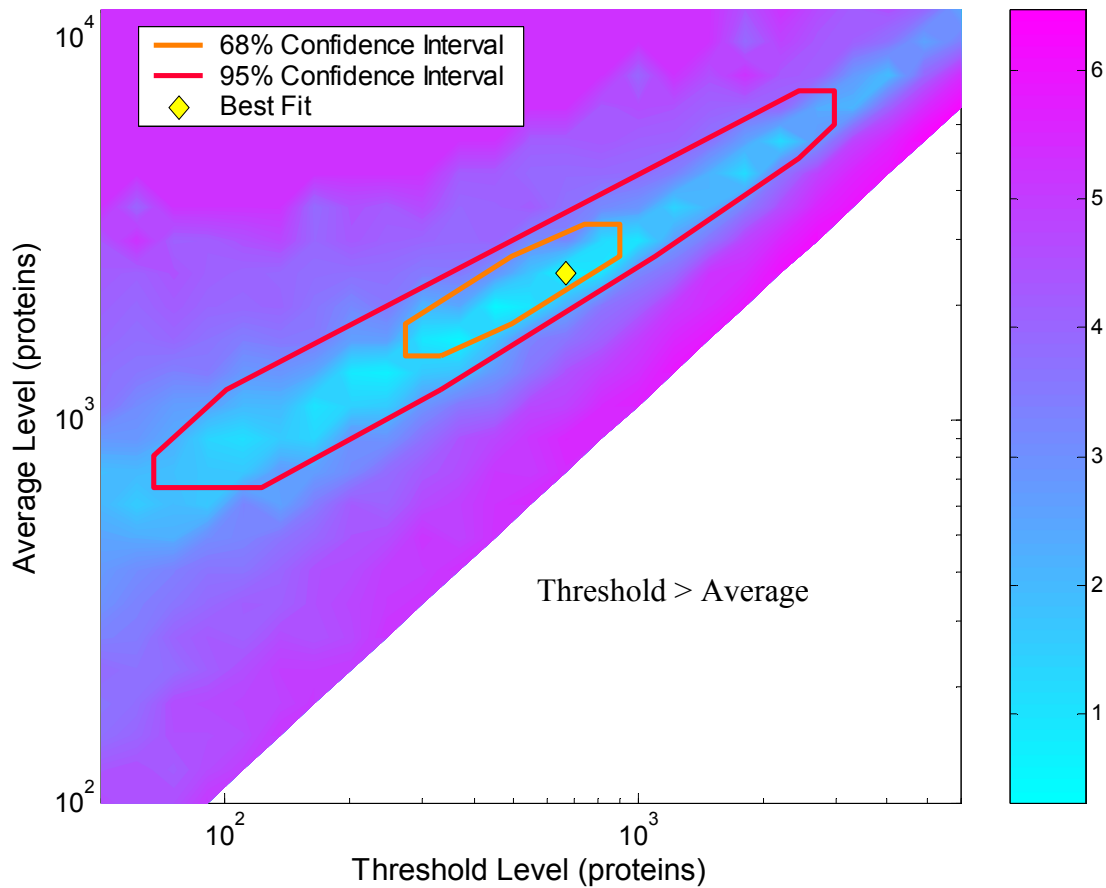
In our MATLAB (MathWorks, Natick, MA) simulation, cells are made to virtually grow and divide according to the measured single-cell doubling times (Supplementary Figure 2). These virtual cells are assumed to begin OFF and then allowed to switch to the ON state with constant probability. To calculate the wait time for these switches, we used the Gillespie algorithm⁴. Once a cell makes a switch, its time is recorded and the remaining OFF cells continue to divide in the simulation. Like data collection at the microscope, we run several hundred of these simulations, each for fixed periods of 920 simulation minutes (the median experimental run time). Our simulated data therefore contain the same ‘sampling artifacts’ (see Supplementary Figure 7) as our measured data. The simulation produces switch data in the same form as the microscopy data, and is then processed identically. A similar scatter plot can be obtained by randomly drawing from the marginal switching distributions and in that way generating an uncorrelated scatter plot. This method is less accurate however because the resulting distribution will not be consistent with the opportunity windows described in S3.B.



Supplementary Figure 8 (a) Overlay of switching patterns for all M-D, GM-GD, and S1-S2 cell pairs. (b) Overlay of the same in the Gillespie-based model.

S3.F Confidence intervals for Monte Carlo model

There are two free parameters in our stochastic model (see ‘Stochastic Model’ in main text): the average protein level and the switching threshold. Chi-squared values were computed for different combinations of these values, yielding a best-fit value of (average, threshold) \sim (2400 proteins, 670 proteins) and corresponding confidence intervals (see Supplementary Figure 9).



Supplementary Figure 9 Monte-Carlo model confidence intervals. Chi-squared is shown in a heat map as a function of the two free parameters in our model, the average level of proteins and the switching threshold. Colorbar at right indicates chi-squared value on a logarithmic axis. Burst size is fixed to 1,200 proteins/mRNA, consistent with Bar-Even *et al*⁵. Yellow diamond indicates the best fit of all tested values and occurs at (average, threshold) \sim (2400 proteins, 670 proteins). Orange and red contours represent 68% and 95% confidence intervals, respectively. Points where the threshold exceeds the average have been excluded.

Supplementary References

1. Keegan, L., Gill, G. & Ptashne, M. Separation of DNA binding from the transcription-activating function of a eukaryotic regulatory protein. *Science* **231**, 699-704 (1986).
2. Peng, G. & Hopper, J. E. Gene activation by interaction of an inhibitor with a cytoplasmic signaling protein. *Proc Natl Acad Sci U S A* **99**, 8548-53 (2002).
3. Hartwell, L. H. & Unger, M. W. Unequal division in *Saccharomyces cerevisiae* and its implications for the control of cell division. *J Cell Biol* **75**, 422-35 (1977).
4. Gillespie, D. T. Exact Stochastic Simulation of Coupled Chemical Reactions. *The Journal of Physical Chemistry* **81**, 2340-2361 (1977).
5. Bar-Even, A. et al. Noise in protein expression scales with natural protein abundance. *Nat Genet* **38**, 636-43 (2006).

Supplementary Movie 1

This movie shows a single OFF cell as it grows over 690 minutes into a small colony of 16 cells. Images show phase contrast image overlaid with fluorescence in purple. Colored circles and zoomed panels (below) highlight the dynamics of four individual cells: the original progenitor cell (1), its first two children (1-1 and 1-2), and its first granddaughter (1-1-1). Before 600 minutes, no cells have fluoresced. In the last 90 minutes, a group of cells all switch nearly synchronously from the OFF to the ON state. Time between images is 30 minutes.