eLIFE elifesciences.org

RESEARCH ARTICLE

Figures and figure supplements

Stochastic tuning of gene expression enables cellular adaptation in the absence of pre-existing regulatory circuitry

Peter L Freddolino *et al*

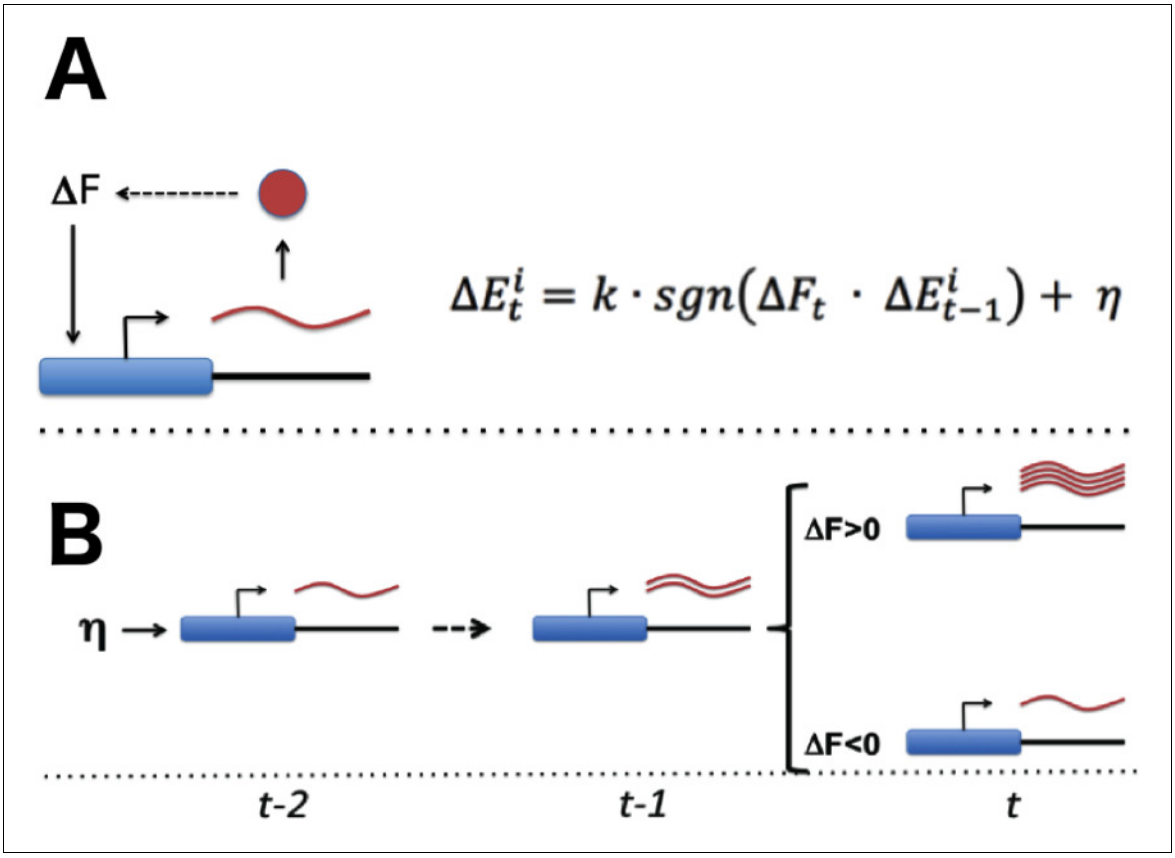

Figure 1. Stochastic tuning of gene expression by fitness optimization at gene promoters. (**A**) Each gene contains a noisy expression apparatus with noise amplitude η that allows exploration of a range of transcriptional activities. Each transcription apparatus also maintains a record of its previous change in transcriptional activity (ΔE_{t-1}). The change in transcriptional activity has the potential to contribute to a change in global health (ΔF_t) through the downstream effect of the gene product's activity (likely through a multi-step pathway; for example, the biosynthesis of a metabolite that is limiting for growth). A global metabolic integrator can transduce this change in health/fitness to every gene's expression apparatus. At any point in time, the expression apparatus executes a change in transcriptional activity (ΔE_t) proportional (k) to the sign (*sgn*) of the product of ΔE_{t-1} and ΔF_t plus noise (η). (**B**) A simple example of this can be seen for a gene that experiences a random burst in transcriptional activity. If this leads to an increase in fitness the expression apparatus further increases transcriptional activity. Conversely, if there is a decrease in fitness, the expression apparatus decreases transcriptional activity.

DOI: https://doi.org/10.7554/eLife.31867.003

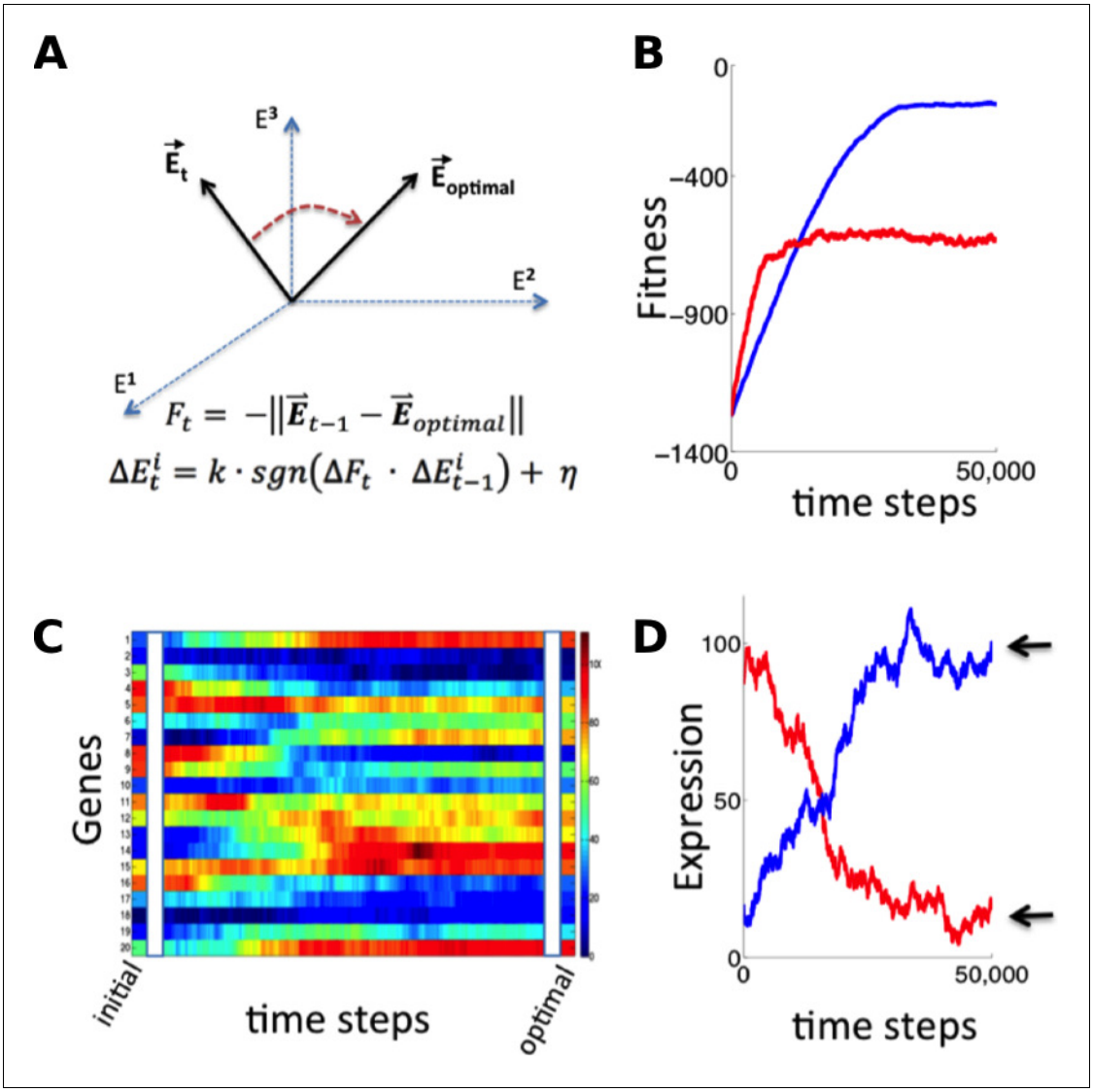

Figure 2. Simulation of fitness-directed stochastic tuning for a thousand-gene system. (**A**) Quantitative framework describing stochastic tuning. The transcriptional activity state of the genome is represented by the vector E, here schematically represented for a three-gene system. In any environment, there is an optimal transcriptional state vector ($E_{optimal}$) that yields maximum fitness. At any time (t), a cell with transcriptional activity state E_t has global health/fitness (F_t) defined as the negative of the Euclidean distance between the immediately preceding transcriptional activity state E_{t-1} and $E_{optimal}$. Each gene promoter (i) executes a change in transcriptional activity ΔE^i_t which has two components: (1) a step with magnitude of *k* and sign (*sgn*) matching that of the product of the global change in fitness (ΔF_t) experienced at time *t* and the preceding change in transcriptional activity ΔE^i_{t-1}, and (2) a noise component with a magnitude of η and a random sign (+/-). (**B**) The stochastic tuning process moves the transcriptional activity state towards the optimum, resulting in increasing health/fitness over time. Simulated trajectories are shown for a 1,000-gene system with $k = 0.1$, $\eta = 0.1$ (blue); $k = 0.5$, $\eta = 0.5$ (red). (**C**) The time evolution of the transcriptional activity state vector as a system containing 1000 genes converges to optimal transcriptional activities through stochastic tuning. The temporal profiles of 20 representative genes are shown, starting from randomly assigned initial activities, and gradually converging to activities that are near optimal for fitness (using parameters corresponding to the blue curve in panel **B**). (**D**) Trajectories of two representative genes are shown for the same simulation as in panel C). Transcriptional activities start at randomly assigned initial values and gradually converge to near the optimum (arrows).

DOI: https://doi.org/10.7554/eLife.31867.004

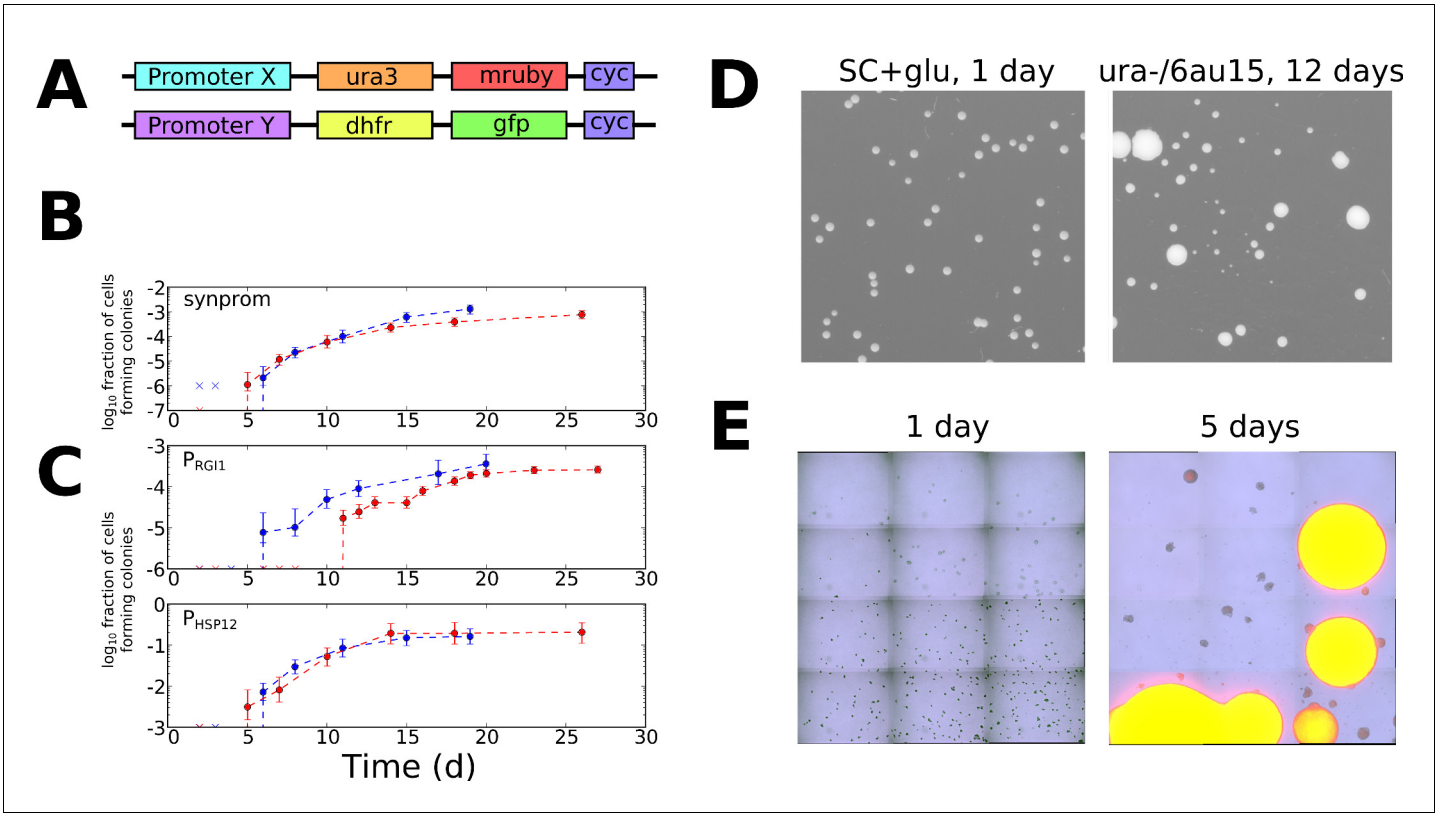

Figure 3. Stochastic tuning of yeast cells under uracil starvation. (**A**) Schematic of the constructs used in this study. All strains are diploid, containing similar insertions at the LEU2 locus of both copies of chromosome III. X is either a synthetic promoter (synprom) or a natural promoter (P_{RGI1} or P_{HSP12}) unless otherwise noted, and Y is either the same promoter as X or is the strong constitutive promoter P_{ADH1}. 'cyc' indicates the well-characterized CYC1 transcriptional terminator (***Russo and Sherman, 1989***). (**B**) Stochastic colony formation on ura-/6AU15 plates for cells containing URA3-mRuby under control of synprom and DHFR-GFP under control of P_{ADH1}. Error bars show central 95% credible intervals; colors show biological replicates performed on different days. 'x' marks are shown at the bottom of the axis for days where zero visible colonies were present at all plated dilutions. Cells plated on SC+glu uniformly form visible colonies within 1–2 days. (**C**) As in panel B, but with URA3-mRuby controlled by P_{RGI1} or P_{HSP12} as indicated. (**D**) Images of colony growth on SC+glu and ura-/6AU15 plates taken at the specified number of days after plating (1 day for SC+glu, 12 days for ura-/6AU15). Growth of colonies is nearly uniform on SC+glu plates but shows non-uniform stochastic emergence on ura-/6AU15. *N.b.* the plated dilutions for the two plate types are not the same. URA3 expression for the experiment shown is controlled by P_{HSP12}, but similar behavior was observed for all promoters discussed here. (**E**) Early colony formation on ura-/6AU15 plates imaged by superimposed differential interference contrast and fluorescence microscopy. Cells contain P_{HSP12}-URA3-mRuby/P_{ADH1}-DHFR-GFP. Left panel: One day after plating. By this timepoint small, macroscopic colonies would have formed on SC+glu plates, but instead cells remain in microcolonies having undergone no more than three doublings. Right panel: Same plate as left, five days after plating. While most cells have not grown since the one-day timepoint, other cells having undergone successful tuning instead form larger colonies with URA3 expression sustained throughout them.

DOI: https://doi.org/10.7554/eLife.31867.005

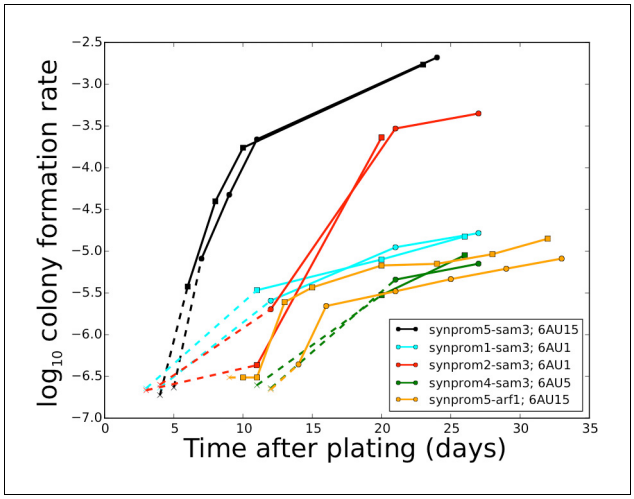

Figure 3—figure supplement 1. Stochastic colony formation rates for cells with URA3 driven by a variety of synthetic promoters. The 'synprom' discussed in the main text is synprom5-sam3. Data are taken from ura-/6AU agar plates containing the listed 6AU concentration. An 'x' followed by a dashed line indicates the threshold of detection from the experiment, marked at the last data point prior to any colonies being observed.
DOI: https://doi.org/10.7554/eLife.31867.006

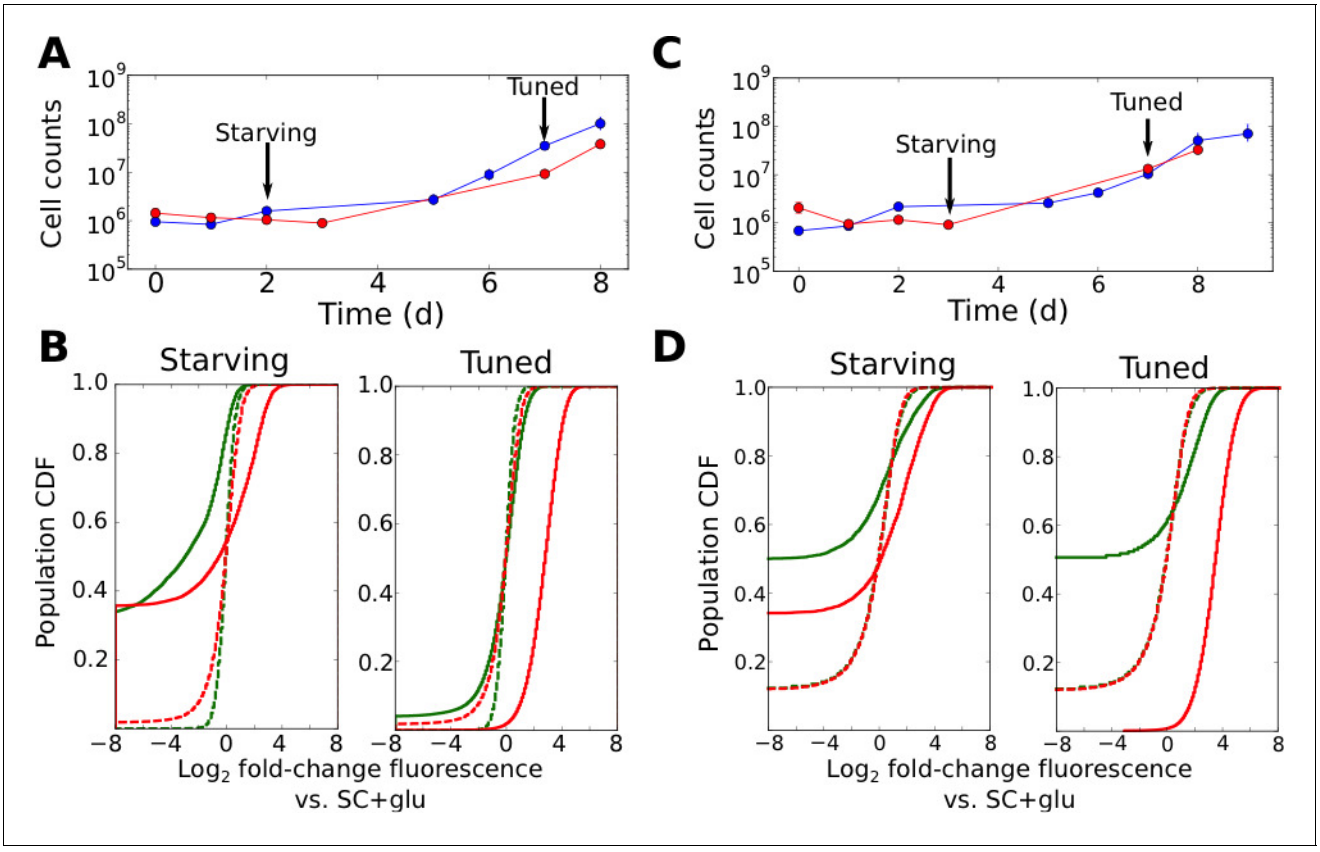

Figure 4. Tuning is both promoter- and allele-specific. (**A**) Cell counts for synprom-URA3-mRuby/P_{ADH1}-DHFR-GFP cells in liquid ura-/6AU5 media. Colors correspond to different biological replicates started on different days. Arrows indicate two timepoints from each strain for which fluorescence cumulative distribution functions (CDFs) are shown below. Error bars for cell counts show central 95% credible intervals. (**B**) Flow cytometry cumulative distributions of fluorescence levels for URA3-mRuby and DHFR-GFP during uracil starvation. In each CDF a given timepoint (solid line) is compared to the distribution present for cells in logarithmic growth in SC+glu (rich) media (dashed lines). The values shown are $\log_2$ ratios to the median value of cells growing exponentially in SC+glu. GFP signals are shown in green and mRuby signals in red. (**C**) Analogous to A, but we consider cells where synprom drives both URA3-mRuby and DHFR-GFP. (**D**) Analogous to B, but for cells with synprom driving both URA3-mRuby and DHFR-GFP.
DOI: https://doi.org/10.7554/eLife.31867.007

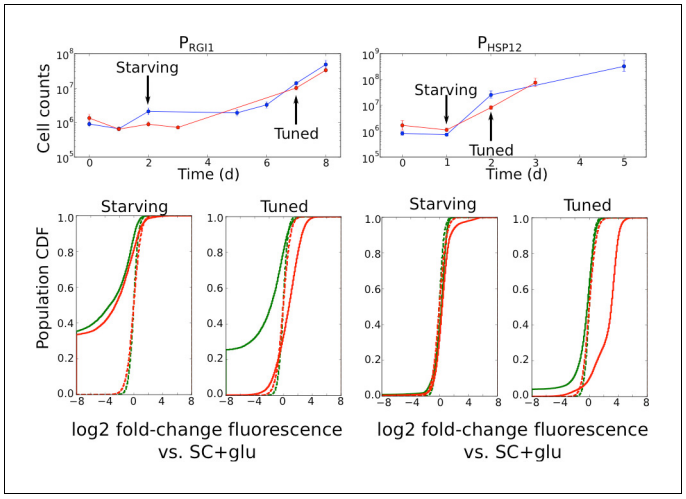

Figure 4—figure supplement 1. Promoter-specific stochastic tuning of URA3 expression by native promoters in *S. cerevisiae*. Flow cytometry data on counts and fluorescence distributions for cells containing URA3-mRuby under control of the noted promoter (P_{RGI1} or P_{HSP12}) and DHFR-GFP under control of P_{ADH1}. Cells were grown in liquid ura-/6AU5 media. For the cell counts (top), colors correspond to different biological replicates started on different days. Arrows indicate two timepoints from each strain for which fluorescence cumulative distribution functions (CDFs) are shown below; in each CDF a given time-point (solid line) is compared to the distribution present for cells in logarithmic growth in SC+glu media (dashed lines). The values shown are $\log_2$ ratios to the median value of cells growing exponentially in SC+glu. GFP signals are shown in green and mRuby signals in red. Error bars for cell counts show central 95% credible intervals.

DOI: https://doi.org/10.7554/eLife.31867.008

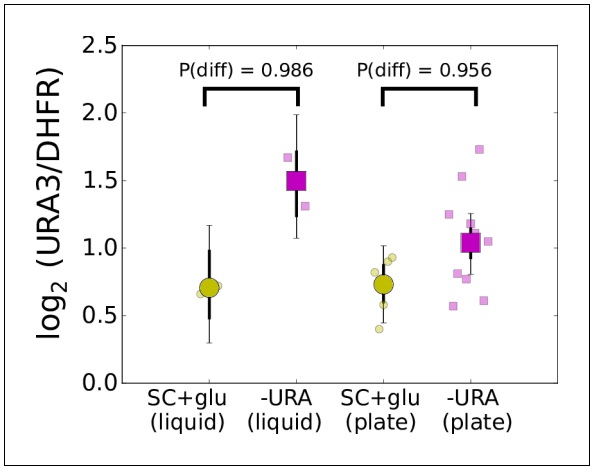

Figure 4—figure supplement 2. Local tuning of URA3 expression. RT-qPCR based quantification of URA3:DHFR ratio for synprom-URA3-mRuby/synprom-DHFR-GFP cells either in in SC+glu media (liquid or plates), from tuned liquid cultures in ura-/6AU5 media, or for tuned colonies on ura-/6AU15 plates. Large points show estimated averages obtained via a Bayesian hierarchical model (see Methods), small points show values for biological replicates, and error bars indicate 75% (strong) or 95% (thin) credible intervals. P(diff) in each case gives the posterior probability that the –URA value is greater than the corresponding SC+glu value. See ***Supplementary file 7*** for primer sequences.
DOI: https://doi.org/10.7554/eLife.31867.009

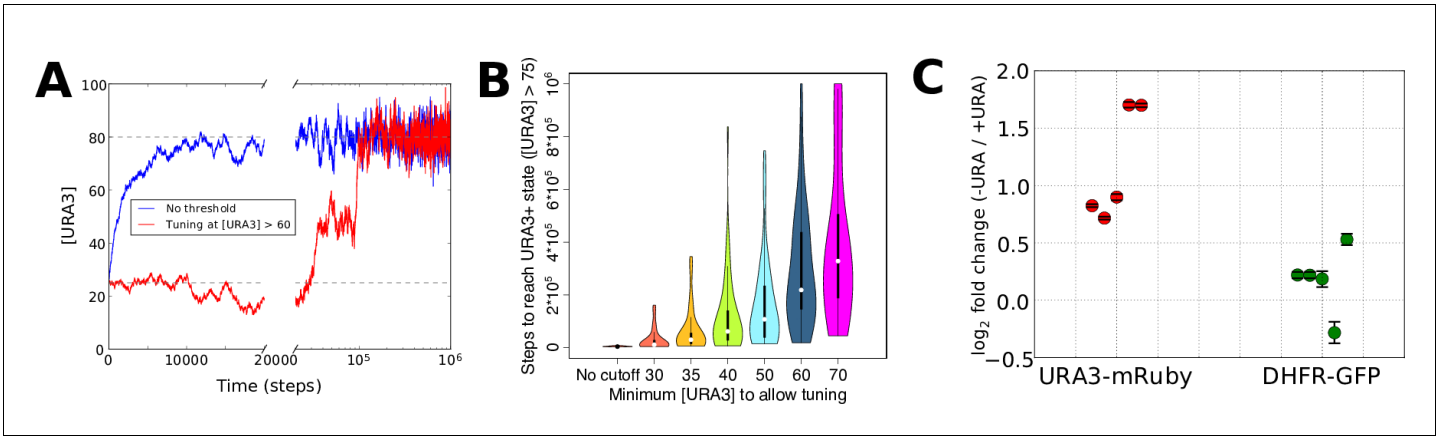

Figure 5. Numerical modeling and experimental validation of changes in tuning behavior as a function of 6AU concentration. We simulated the gene expression dynamics of cells containing URA3 under the control of a non-native promoter, when exposed to uracil-depletion stress with varying concentrations of the URA3 inhibitor 6AU. The model employed is equivalent to that in ***Figure 2A***, with k = 0.1, η = 0.1, and the target expression profile equal to that for the case shown in ***Figure 2B*** except for the case of the gene corresponding to URA3, whose optimal value was set to 80. (**A**) Typical trajectories of URA3 expression levels for a cell in the presence of low (blue) or high (red) 6AU concentrations, which alter the minimum URA3 expression level at which fitness-directed stochastic tuning can occur. We show results for a starting URA3 level [URA3]=25, with optimal fitness occurring at [URA3]=80. The initial and optimal URA3 levels are shown as gray lines. (**B**) Violin plots of the distributions of the minimum time required to reach a URA3+state (defined as [URA3]>75) in the presence of increasing concentrations of 6AU (implemented as higher thresholds of URA3 required for stochastic tuning to become active). In each case distributions reflect 50 independent trajectories simulated at each 6AU level. (**C**) Experimental validation of model predictions. Cells were grown in liquid ura-/6AU1 media (-URA) for 3–4 hr and then had the expression of fluorescent reporter proteins compared (using flow cytometry) with those of the equivalent cells grown in SC+glu (+URA) over the same time period. Values show $\log_2$ fold changes from SC+glu to ura-/6AU1; error bars show bootstrap-based 95% confidence intervals. Biological replicates performed on different days are shown side by side; the order of replicates is matched for URA3-mRuby and DHFR-GFP.

DOI: https://doi.org/10.7554/eLife.31867.010

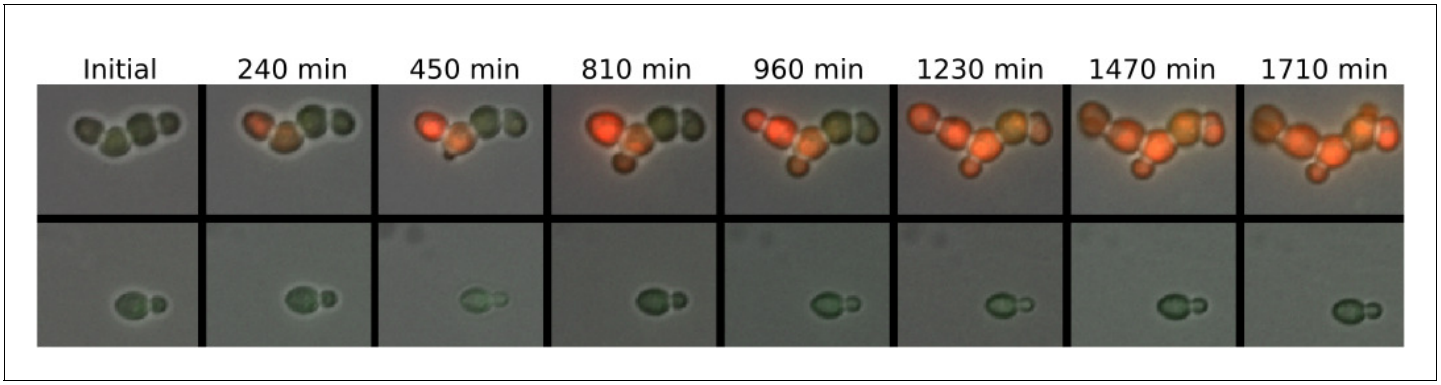

Figure 6. Sustained trans-generational inheritance of URA3-mRuby expression in tuned microcolonies. Shown are fluorescence microscopy time courses of microcolonies beginning after 12 hr of exposure to ura-/6AU5 media. A tuned colony is shown on top and a nearby untuned colony on the bottom. Fluorescence values are uniformly scaled but are not otherwise processed.

DOI: https://doi.org/10.7554/eLife.31867.011

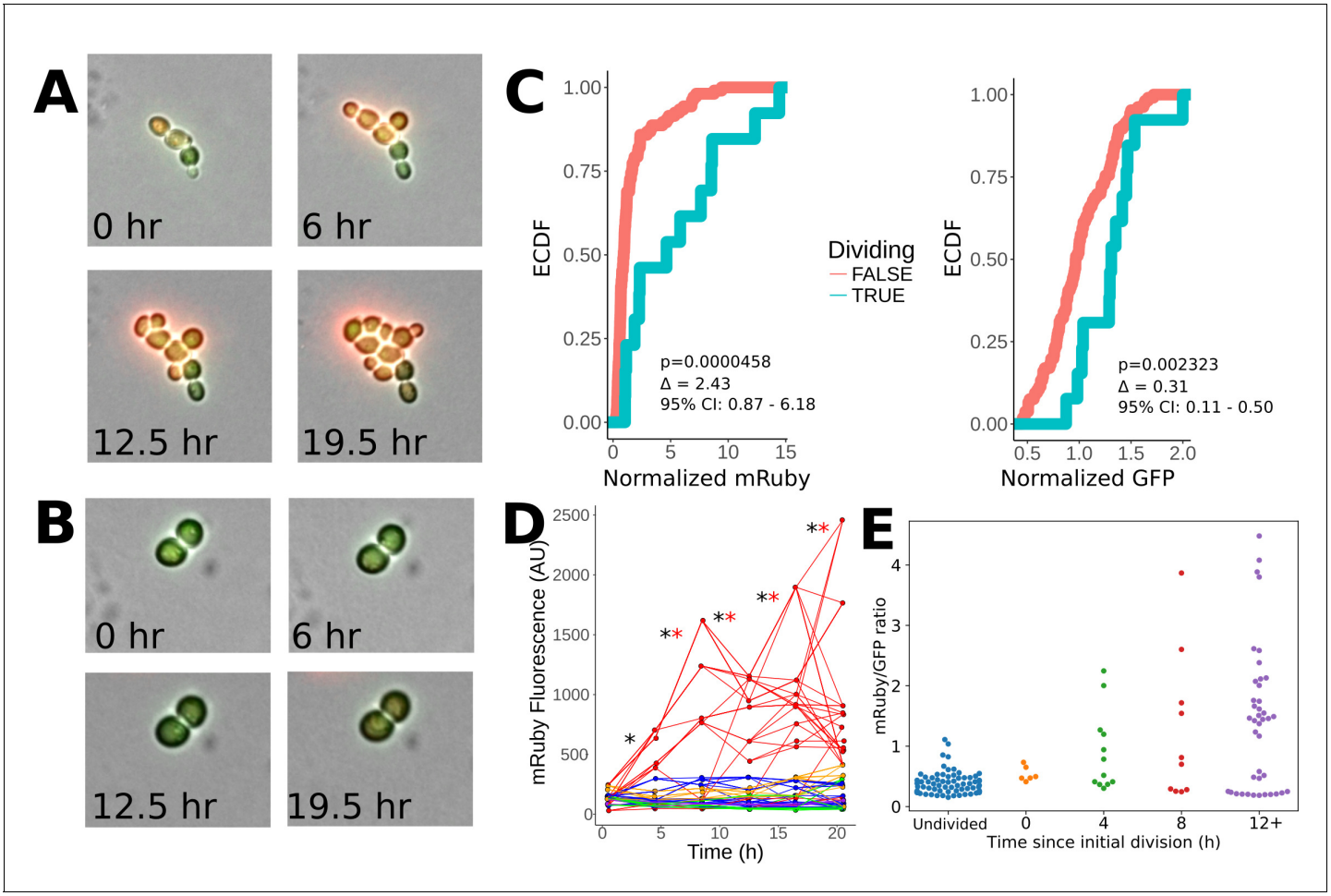

Figure 7. Heritability of elevated mRuby levels during tuning. (**A**) Formation of a microcolony over 24 hr of exposure to ura-/6AU5 media in P_{HSP12}-URA3-mRuby/P_{ADH1}-DHFR-GFP cells. GFP and mRuby are shown as transparent green and red overlays. (**B**) Snapshots equivalent to **A**) for a non-tuned colony in the same field of view. (**C**) Observed cumulative distributions (empirical cumulative distribution function; ECDF) of mRuby (left) and GFP (right) levels for cells that either do or do not divide in the timepoint following the measurement (analyzed in four-hour intervals). Values are pooled over all timepoints except the first, for five colonies growing in a single field of view. p-values arise from a Wilcoxon rank sum test applied to the shift between the non-dividing and dividing cells. Δ indicates a point estimate for the difference in fluorescence of the dividing vs. non-dividing cells, along with a 95% confidence interval (95% CI). Values shown are raw fluorescence normalized by the median value for all observations of each fluorescent protein; note the different x scales for mRuby vs. GFP. (**D**) Lineage traces showing long term propagation/inheritance of URA3-mRuby protein levels. At each specified timepoint, the average fluorescence of each cell is shown on the y axis, with lines connecting each cell to the cell(s) arising from it at the subsequent timepoint; thus, forks in the lines indicate cell division. Colors specify which of five microcolonies a given cell is a part of; only the red microcolony showed notable tuning over the course of the experiment. A black '*' is shown for each transition between adjacent timepoints for which the correlation of ranks between the timepoints in question is significant ($p<0.05$) using a Spearman correlation test, and a red '*' is shown for transitions where the same criterion holds considering only the rank ordering of cells in the red (tuned) colony (the colony shown in panel **A**). (**E**) Observed distribution of mRuby/GFP ratios depending on time elapsed since a lineage of cells began to divide. The x axis divides the cells up by the time (measured in four-hour intervals) that has elapsed since the first observed division event of an ancestor of that cell; 'Undivided' indicates cells in lineages that have not yet divided in the analyzed trajectory, and 0 hr denotes cells that will divide before the next analyzed snapshot. Note that points are plotted for each cell at each analyzed frame relative to its own growth history, and thus not all cells at a given x position necessarily arise from the same time point in the image series.

DOI: https://doi.org/10.7554/eLife.31867.012

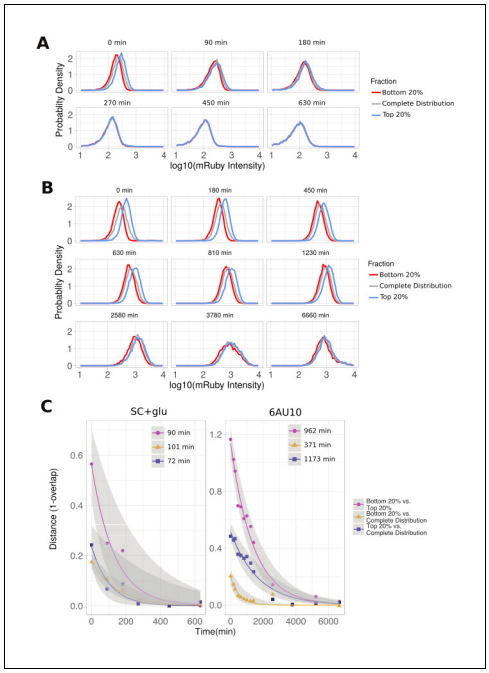

Figure 7—figure supplement 1. Mixing times of mRuby levels for growing (uracil-replete) and uracil-starved cells. (**A**) Time-evolution of distributions of mRuby fluorescence values for cells initially sorted to contain only the top or bottom 20% of cells (Top 20% and Bottom 20%, respectively), compared with the full mock-sorted population (Complete Distribution). Distributions are shown after growth for the specified time in SC +glu media. Data for all panels is for cells with synprom-URA3-mRuby/synprom-DHFR-GFP. (**B**) As in panel **A**), but for cells exposed to uracil starvation via ura-/loflo/6AU10 media. (**C**) Time evolution of overlap of fluorescence distributions between the specified subpopulations for data from panels **A–B**). Plotted points show the distance (1-overlap) of the normalized density estimates. Solid lines show curve fits for simple exponential decay, with grey lines giving 95% confidence intervals. Numbers indicated alongside each point series in the legend give half lives for the decay process.

DOI: https://doi.org/10.7554/eLife.31867.013

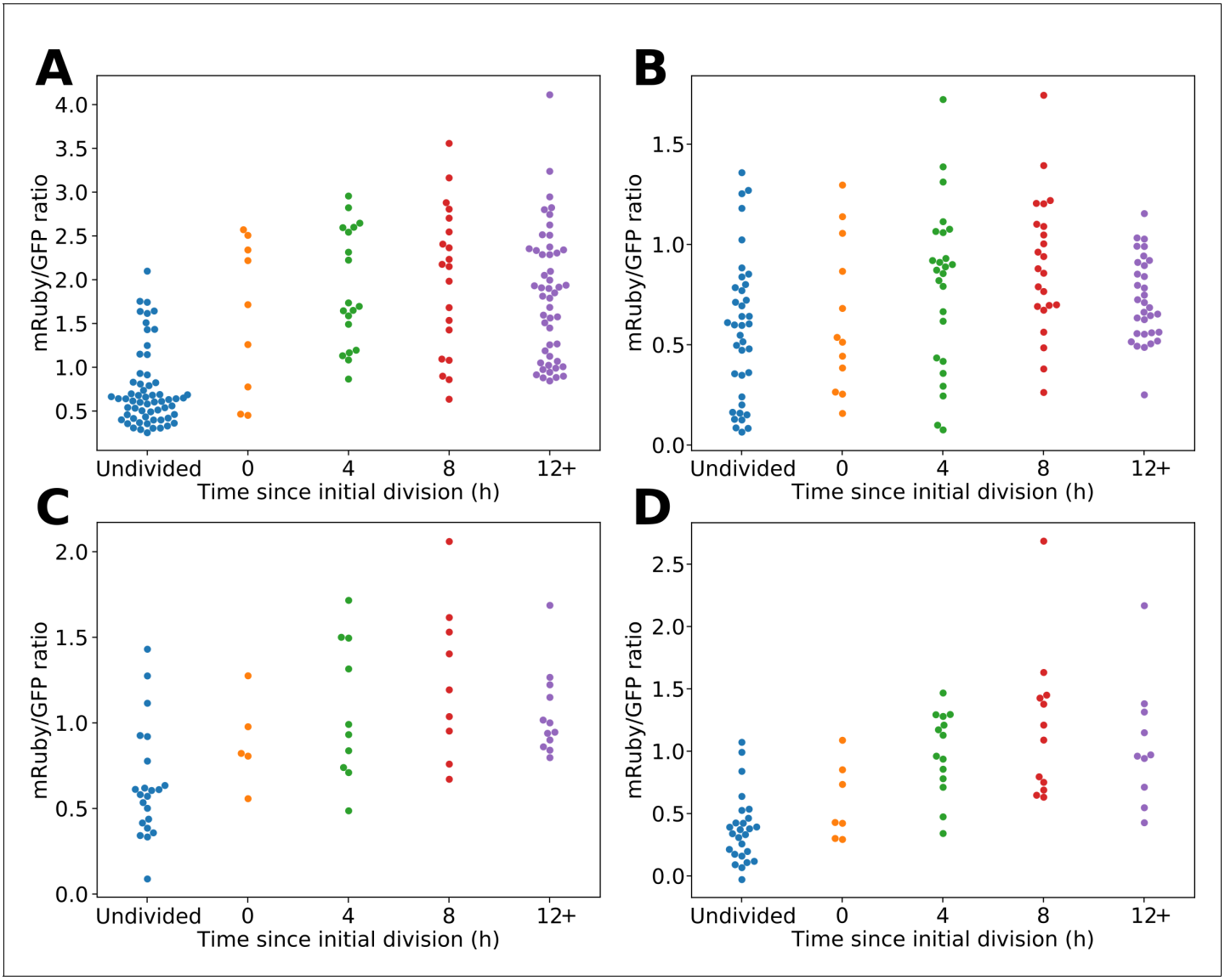

Figure 7—figure supplement 2. Heritability of elevated mRuby levels during early tuning. Analogous to ***Figure 7E***, each panel shows the observed distribution of mRuby/GFP ratios depending on time elapsed since a lineage of cells began to divide. The x axis divides the cells up by the time (measured in four-hour intervals) that have elapsed since the first observed division event of an ancestor of that cell; 'Undivided' indicates cells in lineages that have not yet divided in the analyzed trajectory, and 0 h cells that will divide before the next analyzed snapshot. (**A**) Data from a single field of view plated on ura-/loflo/6AU5 agar after 12 hr of pregrowth in ura-/loflo/6AU5 liquid media. (**B–D**). Data from each of three fields of view where the cells were pregrown in SC+glu media, immobilized on poly-D-lysine coated cover slips, and subsequently covered with ura-/loflo/6AU5 media.
DOI: https://doi.org/10.7554/eLife.31867.014

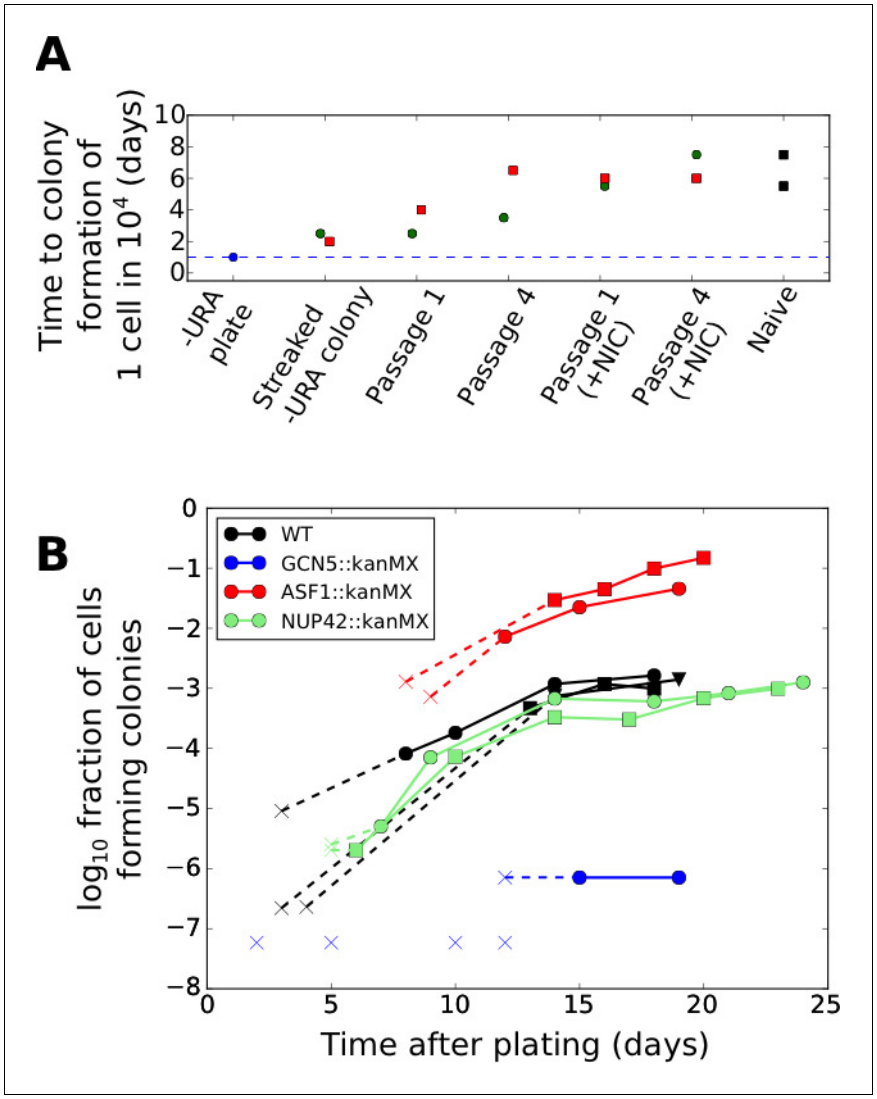

Figure 8. Effects of genetic and chemical perturbations on the efficacy of fitness-directed stochastic tuning and its epigenetic reversion. (**A**) Time courses of recovery back to the naïve state for tuned synprom-URA3-mRuby/P_{ADH1}-DHFR-GFP cells grown in either SC+glu or SC+glu with 25 mM nicotinamide added (+NIC). Extremes are shown for the colony formation times of cells never exposed to –ura conditions (Naïve) and for single colonies isolated after streaking out cells from ura-/6AU15 plates onto SC+glu (Streaked –URA colony). Colors of points indicate a single lineage beginning from a single streaked out colony picked at the first SC+glu plate stage. The cells were then repeatedly passaged in liquid SC+glu media and assessed for colony formation rates on ura-/6AU15 plates on subsequent days, as detailed in ***Figure 8—figure supplement 1***. (**B**) Colony formation rates on ura-/6AU15 plates in the presence of various genetic perturbations, assessed by colony counts from platings of selected dilutions of cells. An 'x' followed by a dashed line indicates no observed colonies and is shown at the threshold of detection from the experiment. All mutations are in a synprom-URA3-mRuby/leu2Δ0 background.

DOI: https://doi.org/10.7554/eLife.31867.015

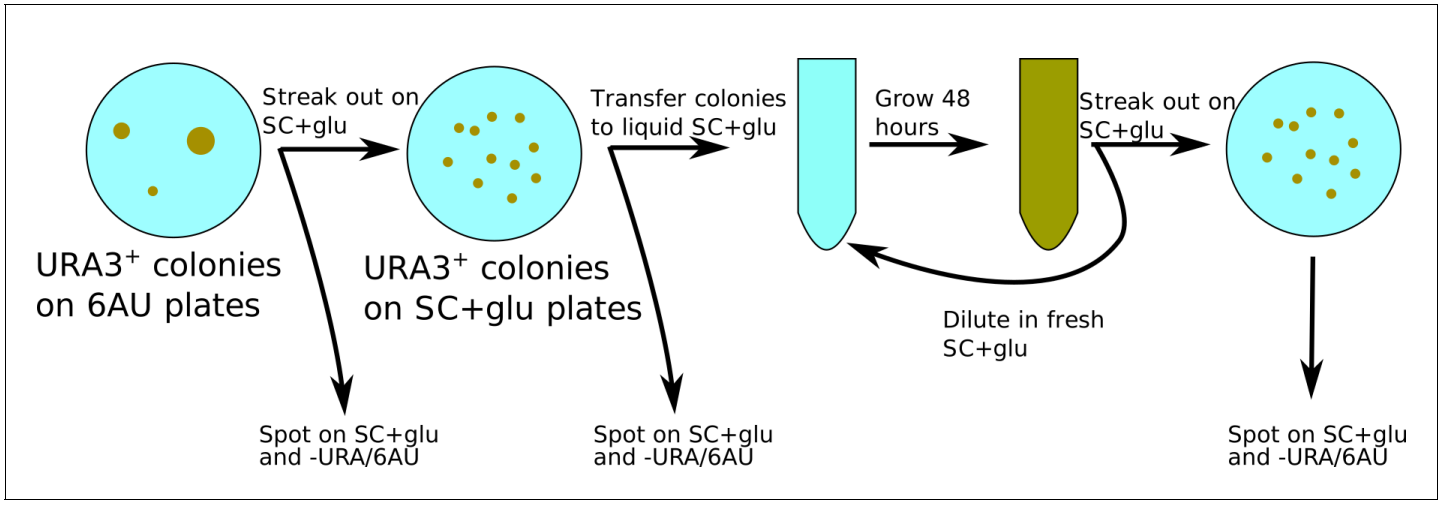

Figure 8—figure supplement 1. Recovery of cells taken from ura-/6AU plates toward a naïve state. Experimental setup for assessing recovery to a naïve state. Tuned colonies growing on a –ura/6AU15 plate were streaked onto an SC+glu plate, colony purified, and then grown for 48 hr cycles in liquid SC+glu. Colony formation rates were assessed after each transfer and quantified according to the amount of time needed for formation of one colony per 10,000 cells plated (see Materials and Methods for details).

DOI: https://doi.org/10.7554/eLife.31867.016

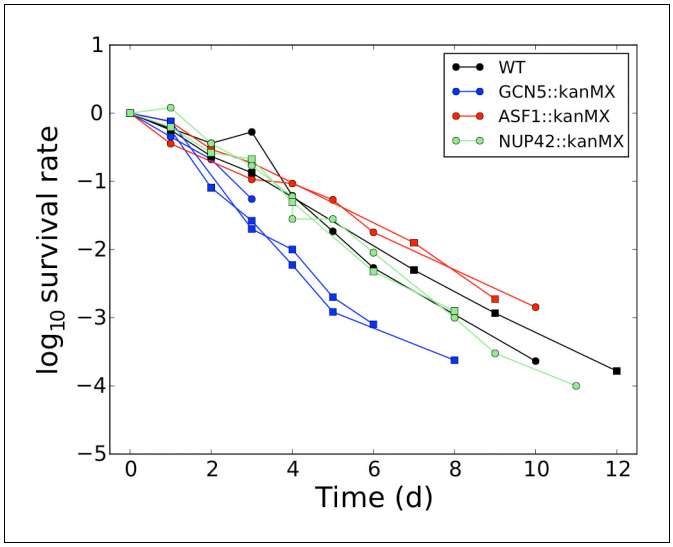

Figure 8—figure supplement 2. Survival of cells in –ura media in the presence of genetic perturbations. Loss of viability of cells with the indicated gene deletion, but not bearing any copy of URA3, in SC-ura+glu media. Separate traces indicate different biological replicates.

DOI: https://doi.org/10.7554/eLife.31867.017

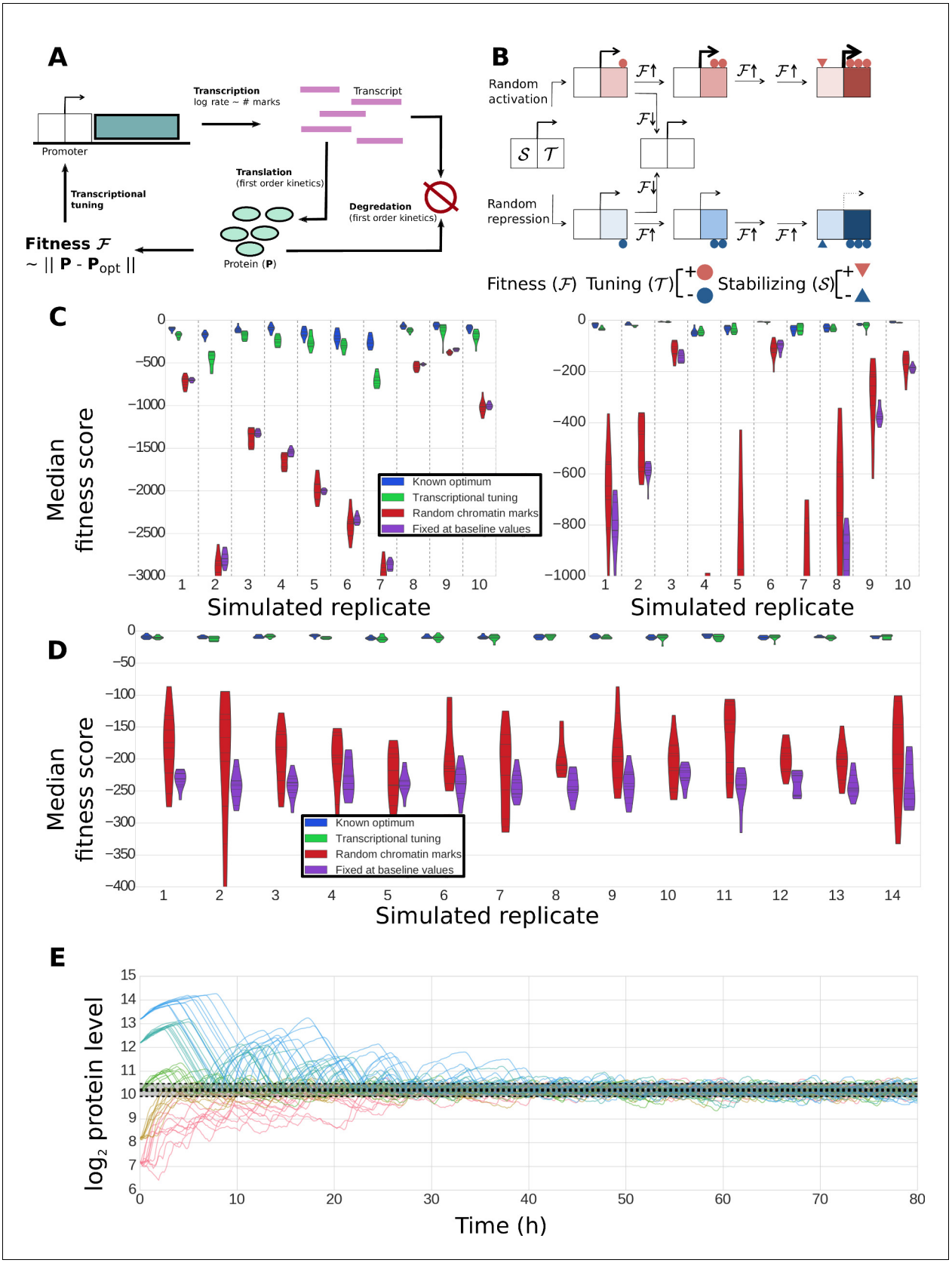

Figure 9. Construction and performance of a biologically feasible model for fitness-directed stochastic tuning. (**A**) Schematic of processes modeled in the simulation. Transcripts are produced at a rate dependent upon the state of chromatin marks at each promoter; each transcript has a fixed, gene-

Figure 9 continued on next page

Figure 9 continued

dependent probability of being translated at each timestep (producing a protein), and may also be degraded (again, with a gene-dependent probability). Similarly, each copy of a protein may be degraded at each timestep with a protein-dependent probability. The fitness of the system is calculated as the Euclidean distance between the current profile of protein counts present in the cell from a target optimum. Chromatin marks may be added or removed at each promoter at each step, as shown in panel **B**). (**B**) Logic underlying changes in tuning and stabilizing mark counts at each step. Tuning marks (T) may be added or removed at each step based on the recent history of changes in fitness, and whether each promoter currently has a net positive (activating) or negative (repressive) T count. Stabilizing marks (S) provide longer term integration by adding activating or repressive marks over time in response to the state of the tuning marks. Thus, if an unmodified promoter undergoes random addition of a positive tuning mark (top path), and that addition proves favorable, it will undergo further addition of positive T marks. If fitness continues to increase, stabilizing marks (S) will be added to stabilize its higher activity. Similar logic holds for the random addition of negative tuning marks (bottom path). In both cases, if the random T-mark perturbation proves unfavorable, the promoter will be modified in the opposite direction, in this case returning it back to its original unmodified state. (**C**) Distributions of fitness scores for a one-gene system obtained in twenty simulations using different randomly sampled biological parameters (e.g., transcript stabilities, translation rates, etc.) – these different parameter sets are the 'simulated replicates' referred to on the x axis. The median scores over the last quarter of the simulation are shown for 10 independent tuning trajectories (differing in their random number seeds). Each simulation proceeded for 300,000 steps (83.3 hr of simulated time). Different colors indicate varying methods used to control transcription rates (as shown in the legend): 'Known optimum' refers to a case where transcription rates are kept fixed at their predefined target values, 'Stochastic tuning' is the full model described in the Methods section, 'Random chromatin marks' is equivalent to the tuning model except that the direction of T chromatin mark addition is random instead of fitness directed, and 'Fixed at baseline values' shows the case where transcription rates are fixed at their initial values (intended to correspond to the environment that the cells were in prior to the onset of stress exposure). Dashed vertical lines group simulations performed with identical parameters. On the left axis we show ten sets of simulations where the target transcription rate was eight-fold higher than the starting rate, and on the right axis simulations where the target transcription rate was eight-fold lower than the starting rate. (**D**) Robustness of tuning against changing model parameters. Violin plots are defined as in panel **C**), but in this case show the distributions of fitness scores observed under variations of the model parameters (*e.g.*, magnitude of individual tuning and stabilizing marks) for a single, randomly chosen set of gene-specific parameters. Plotted are the median fitness scores over the last quarter of each simulation, using either our central 'baseline' parameters for all model parameters (leftmost replicate; see ***Supplementary file 8***), or twofold changes (up or down) of each editable parameter in our model. (**E**) Tuning performance of a single gene matching a wide range of biological challenges. For a fixed set of biological parameters (see Materials and Methods), we performed 10 simulations each where the initial transcription rates were off from the target rate by a factor of 2^3, 2^2, 0, 2^{-2}, and 2^{-3}, running in order from blue to red. A strong dashed black line shows the median obtained from the last quarter of a long (3 million step) simulation with transcription rates fixed at their optimal values; the shaded region shows the extent of a region encompassing 95% of the timepoints observed in that window. Regardless of initial conditions, the protein level approaches the optimal value and then stably oscillates around it, with amplitudes similar to those observed in the control simulation with target transcription rates.

DOI: https://doi.org/10.7554/eLife.31867.019