Supplementary Materials for

A Living Vector Field Reveals Constraints On Galactose Network Induction In Yeast

Sarah R. Stockwell and Scott A. Rifkin

correspondence to: sarifkin@ucsd.edu

This PDF file includes:

Other Supplementary Materials for this manuscript includes the following:

Supplementary Materials:

Materials and Methods:

The two yeast strains used in this paper were derived from BY4741(*33*) (*his3Δ, leu2Δ, met15Δ, ura3Δ, MATa*), which is a derivative of S288C. YSR0145 had a Gal1p C-terminal fusion to yECerulean and a Gal3p C-terminal fusion to 2x-yECitrine.

To make YSR0145, we replaced the GAL1 stop codon with yECerulean-ADH1t-pTEF-SpHIS5-TEFt from pKT101a (from Natalie Cookson) which is identical to pKT101 (*34*) except with CFP replaced by yECerulean. We used the oligos CGTCTCTAAACCAGCATTGGGCAGCTGTCTATATGAATTAggtgacggtgctggttta and GTTATTATTGCGTATTTTGTGATGCTAAAGTTATGAGTAGtcgatgaattcgagctcg to insert this translational reporter into the yeast genome (the upper case pairs with the yeast genome and the lower case pairs with the plasmid). We then made a translational reporter for Gal3p by inserting 2x-yECitrine from plasmid pSRS001 (Fig. S3) before the stop codon of GAL3 using the oligos AGTTTCGAAGCCTGCCTTGGGTACTTGTTTGTACGAACAAtacgctgcaggtcgacggat and CTTTTAATATTTAAAGGTTGTTCCAAGAAGGTGTTTAGTGttataattggccagtctttt. 2x-yECitrine consists of two copies of yECitrine connected by a linker sequence, which approximately doubles the fluorescence intensity of a single yECitrine reporter and allows detection of Gal3p even at low expression levels.

As an internal control, we constructed YSR0140, an isogenic strain to YSR0145 without Gal reporter proteins but with an mCherry (*35*) nuclear marker to enable us to distinguish these cells from YSR0145. We constructed strain YSR0140 by replacing the stop codon of the nuclear-localized NRD1 in BY4741 with mCherry-pTEF-NAT-TEFt (*35*) using the oligos GAATATGCTTAACCAACAGCAGCAGCAACAACAACAAAGCatggtgagcaagggcgagga and TTTTATGTACTATGAGCAAATAAAGGGTGGAGTAAAGATCatcgatgaattcgagctcg with plasmid pBS35+25. This plasmid is identical to plasmid pBS35 from the Yeast Resource Center (http://depts.washington.edu/yeastrc/pages/pBS35.html) except with hph replaced by NAT.

Growth conditions and media

Cells were grown at 30C in low-fluorescence synthetic dropout media with complete amino acids and the appropriate carbon source. For microfluidics experiments, cells were cultured in *history media* (2% glucose for LTGR, 2% galactose for STGR, or 3% glycerol) for at least 10 generations or 24 hours, whichever was longer. For the final 10 generations, cell density was kept at OD600 < 0.15 to maintain effectively constant sugar concentrations. Following this, cells were loaded into the microfluidics device according to a published protocol (*17*) and grown in history medium until approximately confluent. We aimed for the majority of cells to be experimental cells with enough control cells to be able to estimate background fluorescence well. The microchemostat media was then switched to an induction medium as described in the text and imaged for ≥25 hours. For STGR experiments, the microfluidics medium was instead switched to 12 hours 2% glucose then ≥25 hours 2% galactose. The media took less than 1 minute to completely switch in the cell chamber.

The cells were imaged using a Nikon TiE inverted microscope with the Nikon Perfect Focus System and a 60X oil immersion objective (CFI Plan Apo VC 60X Oil, 1.4NA, 0.13WD). A Prior Lumen 200 with a mercury lamp was used with a Sutter Lambda SC shutter and three filter cubes for fluorescence imaging of YFP (Semrock YFP-2427B), CFP (Semrock CFP-

2432C), and mCherry (Nikon Y-2E/C). The light source was limited to 25% maximum intensity. Bright-field images were taken with a Nikon halogen lamp with exposure controlled by a Sutter shutter. 16-bit greyscale images were captured using a Princeton Instruments PIXIS1024b cooled-CCD camera with a 1024x1024 pixel, (1.3cm x 1.3cm) chip. The temperature on the microscope stage during imaging was maintained at 30C with a Nevtek ASI400 air stream incubator and a custom-rigged microscope enclosure.

We imaged cells briefly and infrequently to minimize photostress and photobleaching (Fig. S4): bright-field 50msec every 2 min.; mCherry 50 msec every 40 min; 2x-yECitrine and yECerulean for 600 and 200 msec respectively every 20 min. Automated time-lapse imaging was controlled by a PC running custom routines in μManager (*36*) and Matlab (*37*). In order to facilitate the image processing, we also took a series of 50 msec bright-field images above and below the target focal plane. Because Gal1p-yECerulean expression varies over a huge range, a single exposure time would not capture the entire dynamic range without saturating. To circumvent this, we took a series of four successive 50 msec exposures and then summed them to get a total yECerulean image that could accommodate the yECerulean dynamic range. We used a similar method for Gal3p-2x-yECitrine taking a series of twelve 50 msec exposures.

Cell segmentation.

We developed a Matlab (*37*) based cell segmentation and tracking pipeline based on a published pipeline (*38*) to extract data from the images (Figs. S5-9).

The cell segmentation pipeline has a number of free parameters, some of which are more influential than others. It was common for a parameter set to work for most but not all of the cells in an image and for another parameter set to pick up the missed ones. We wrote a Matlab GUI to help pick sets of parameters that would collectively identify and segment the cells in a dataset and used three parameter sets for each dataset. We segmented the cells based on each parameter set and then merged them to make a final cell segmentation (Fig. S5).

We took a series of images above and below the focal plane and then max-merged subsets of these to make above-plane and below-plane images. When we subtracted the below from the above images, constructive interference from the diffraction patterns around the cells roughly outlined the cell boundaries (*38*). We then performed a series of image processing steps on this image. We (a) thresholded this image using Otsu's method (*39*), (b) removed small objects, (c) morphologically closed the image to fill holes, identified the insides of cells through a combination of (d) dilation and erosion to identify and remove cell membranes and (e) inverting the segmentation mask after (c), (f) smoothed the edges of the objects, (g) separated clustered cells by the watershed method, (h) filled holes, (i) removed objects that were too small, too large, or touched the edges, and (j) ensured that all segmented objects were 8-separated. Vacuoles not infrequently interfered with proper segmentation, leaving us with objects that looked like Cs or doughnuts. We wrote a script to assess these segmented objects based on area and statistics relating to circularity, merge them into a single cell if appropriate, or fill them in otherwise. This gave us three cell segmentation masks for a frame. We merged them using a pipeline similar our tracking pipeline described below that matches cells between frames.

Cell tracking.

Once a segmentation mask was created for each frame, we tracked cells across frames. This involves determining how segmented objects in one frame correspond to the segmented objects in the next. The tracking algorithm of Ricicova et al. (*38*) matches cells across frames using the Hungarian algorithm (*40*) to minimize the distances between the centroids of

segmented objects in one frame and the predicted locations of objects from the previous frame based on their previous location and velocity. Unlike Ricicova et al., our interval between frames was only two minutes, and so there was limited movement between frames. For our images, their assignment procedure mismatched a large fraction of cells. However, because the cells did not move much over the two minutes, we were able to rely primarily on overlap between segmented objects to track cells between frames. We devised a pipeline (Figs. S6-7) to make these assignments and to cope with both cell movement and improper segmentation due to splitting a single cell into two objects, combining two cells into one object, and filling in a missing cell in one or two frames.

When confluent, approximately 2000-3000 cells were present in a field of view in the microfluidic device. We were able to reliably track a large fraction of cells across hours and, crucially, measure many cells entirely through their induction of the GAL network.

Classification of cells as alive or dead.

Dead cells autofluoresce for a time but also take on a characteristic appearance under bright-field illumination (Fig. S9). We trained a random forest classifier to distinguish between alive and dead cells based on the bright-field images and classified all segmented objects as alive or dead. We also used this classification to improve the cell tracking by splitting 'cells' that were classified as dead for a number of frames but then were resurrected.

Estimation of protein concentration.

We estimated protein concentrations within the cells based on Gal1p-yECerulean or Gal3p-2x-yECitrine fluorescence intensities. To measure these, we first added the separate 16-bit images to make a composite image where the range could be larger than 2^16. To remove the effects of variable background signal, we calculated the median background fluorescence level for non-cell areas of each image and subtracted that from the image. This gave all images a comparable baseline. During the experiments, some cells developed puncta which showed up as small, high-intensity fluorescent spots, particularly in the 2x-yECitrine channel. To prevent these from distorting the fluorescence intensity concentrations, we removed them in a two step process. After scaling the intensities of fluorescence image to be between 0 and 1, identified cells where the skewness was greater than 1.5, and then removed any pixels in that cell with intensity values greater than seven median absolute deviations above the median. This procedure reliably identified pixels in puncta, particularly in cells with low expression where the excess fluorescence would have the most effect. The mean of the remaining pixel values gave a much better estimate of the protein concentration in the cell.

We distinguished control from experimental cells using the nuclear localized mCherry marker. We calculated several statistics on the magnitude and spatial organization of the mCherry signal as well as the yECerulean signal in each cell and used a custom Matlab GUI to train a random forest (*41*) to distinguish control cells from experimental cells from bad cells (i.e. cells that had likely died and had intense autofluorescence in all channels). We used a cost function that was conservative in classifying a cell as control so that our estimated normalization factors would not be skewed high. The false identification rates based on out-of-bag estimates were near zero for control cells and at most a few percent for experimental cells – not high enough to affect our results. We used the cell type classifications to do a final curation of the cell tracking by separating cells that had been classified as experimental in some frames and control

in others into two different cells. If a cell were marked as bad in a frame, it was classified as bad for all subsequent frames (Fig. S9).

Based on the cell masks and the background-subtracted fluorescent images, we measured total pixel intensity within a cell and the area of the cell, and from these calculated the average pixel intensity which we took as proportional to the protein concentration in the cell and used as the cell's fluorescence level. Because the control cells had neither 2x-yECitrine nor yECerulean, we could use them as an internal control for autofluorescence and any bleedover in fluorescence due to out-of-focus light from the experimental cells. To remove any image artifacts that were segmented and tracked as cells, we restricted the dataset to cells that were present for at least 5 bright-field frames. We also ensured that there were at least 50 control cells per frame. For each frame we estimated the median fluorescence level across the control cells and subtracted this from all the cells in the frame. In that way, zero for each frame was set to the median control cell fluorescence level.

Trajectory smoothing for fluorescence and scaling each experiment to be between 0% and 100% expression.

To remove high frequency noise from the individual cell fluorescence trajectories, we smoothed them by first convolving them with a gaussian kernel and then smoothing the result using lowess with a short span. Upon dying, fully induced cells decline in expression. Our classification of cells as alive or dead (above) allowed us to identify most of these cases and remove the drops from the analysis. To catch any others, we also eliminated parts of a cell's trajectory after it reached its peak where expression persistently (80% of post-peak frames) dropped in expression in either fluorescence channel (indicating possible photobleaching or other problem) or dropped monotonically over 5 frames by 50% as there were likely to represent tracking errors or cell dying. To normalize across experiments, we scaled the Gal3p-2x-yECitrine and Gal1p-yECerulean data between uninduced (0% expression) and fully induced (100% expression). 0% expression was the control cell median baseline described above. For 100% expression, we estimated where the fluorescence level plateaued. After examining the fluorescence trajectories for each experiment, we identified a period of time where the cells appeared to plateau, took the maximum fluorescence intensity for each cell in that time period, and took the median of these maxima to be our estimate of full induction.

Vector field estimation

To estimate the vector field for an experiment, we restricted our focus to between 0 (galactose exposure) and 1200 minutes. This is well after the time for GAL network induction in all experiments. We divided the Gal3p-Gal1p state space into bins of 6% plateau expression on each side. Within each bin, we identified all cells with Gal3p and Gal1p expression in that bin which were also tracked into the next fluorescent frame. Data from these two timepoints gave us a velocity for each cell. The mean velocity within the bins is represented in the vector field plot by the length of the arrows (scaled by 1/2 so that the arrows do not overlap). The lightness of the arrows indicates the circular variance of the vector directions, a statistic that ranges from 0 (low variation in directions) to 1 (high variation) (*42*).

To estimate the consistency of the vector fields during induction, we filtered the data so as to remove much of the plateau period. For each experiment we determined the frame at which 95% of the inducing cells had induced (Fig. 1b). For each cell we assigned a cutoff frame as follows: for cell that spent at least 5 frames above 85% plateau level for Gal1p before the

experiment-specific frame determined above, the cutoff frame was the fifth frame; for all other cells, the cutoff frame was the experiment-specific frame. Only data at or before this cutoff frame was used in the consistency calculations.

The consistency statistic is an estimate of the average circular standard deviation of mean vector directions in each of the experiments LTGR, reinduction, and glycerol. We divided the data into bins and only used bins for which there were at least 10 cells in each of the experiments. For each experiment, we sampled the data 500 times with replacement with a sample size of 10 and calculated the mean vector direction for each replicate. For each of the 500 replicates, we took one of the resampled mean vector directions for each experiment and calculated their standard deviation. Our statistic is the mean of these 500 circular standard deviations. In Fig. 2d, we estimated the statistic at a fine spacing by linear interpolation.

Variance decomposition

To estimate the fraction of cells 'steady-state' variation due to consistent inter-cell differences, we subtracted the mean cell fluorescence for each timepoint between 15 and 25 hours after galactose exposure (when the cells were at their plateau level), then fit linear mixed-models to the fluorescence levels, with experiment as a fixed effect and cell ID as a random effect.

Software

Data processing and analysis were done using custom scripts in Matlab (*37*), MySQL (*43*), and R (*44–49*), and all figures were made using ggplot2 (*50*) or Matlab (*37*).

	Gal3p (hours)					Gal1p (hours)			
	25%	50%	75%	IQR		25%	50%	75%	IQR
Glycerol	0.5	0.8	1.2	0.7		1.2	1.4	1.7	0.5
Reinduction	1.7	2.2	2.8	1.1		2.0	2.5	3.1	1.1
LTGR	5.4	6.4	7.5	2.1		6.3	7.4	8.5	2.2

Table S1

Length and variability of lag times until induction upon galactose exposure. Hours until 25%, 50%, or 75% of inducing cells reached 10% plateau expression level for each inducer, Gal3p and Gal1p (Fig. 1b). The IQR is the number of hours between 25% and 75% of cells inducing.

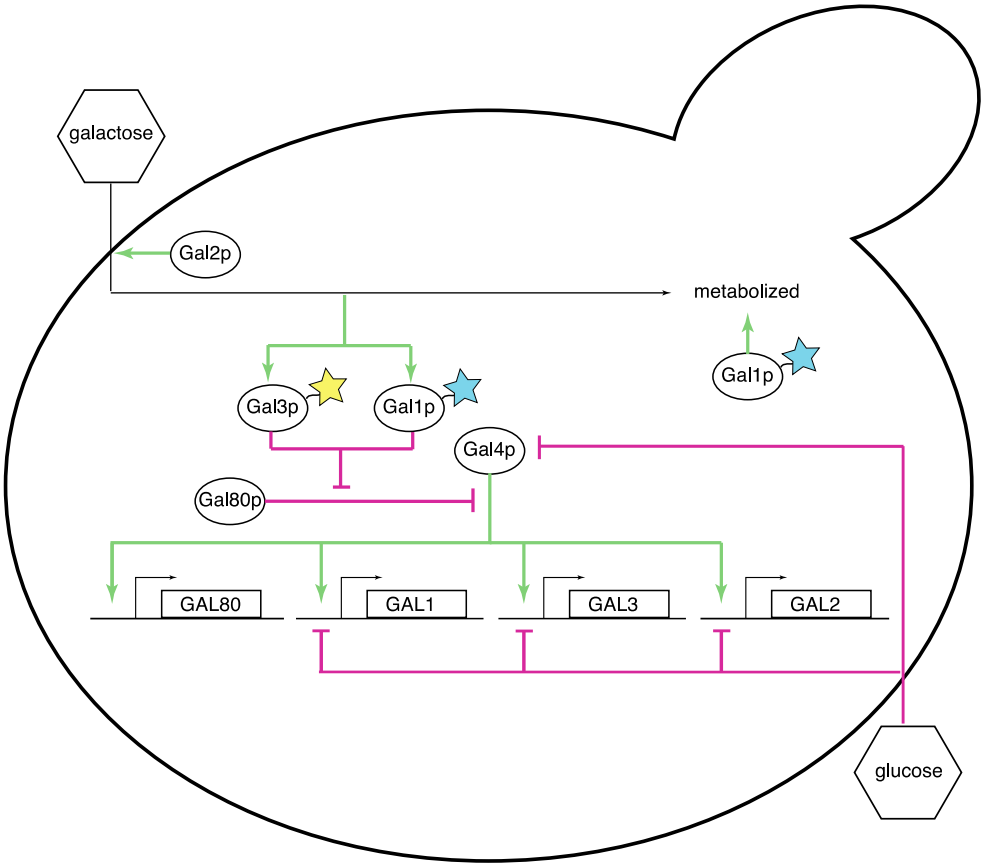

Fig. S1.

A schematic of the regulatory portion of the galactose network in *Saccharomyces cerevisiae* including key targets of glucose repression. Gal7p and Gal10p (not shown) help Gal1p metabolize galactose. Activating and repressing interactions are indicated by green and magenta respectively. Molecular details of glucose repression are not shown.

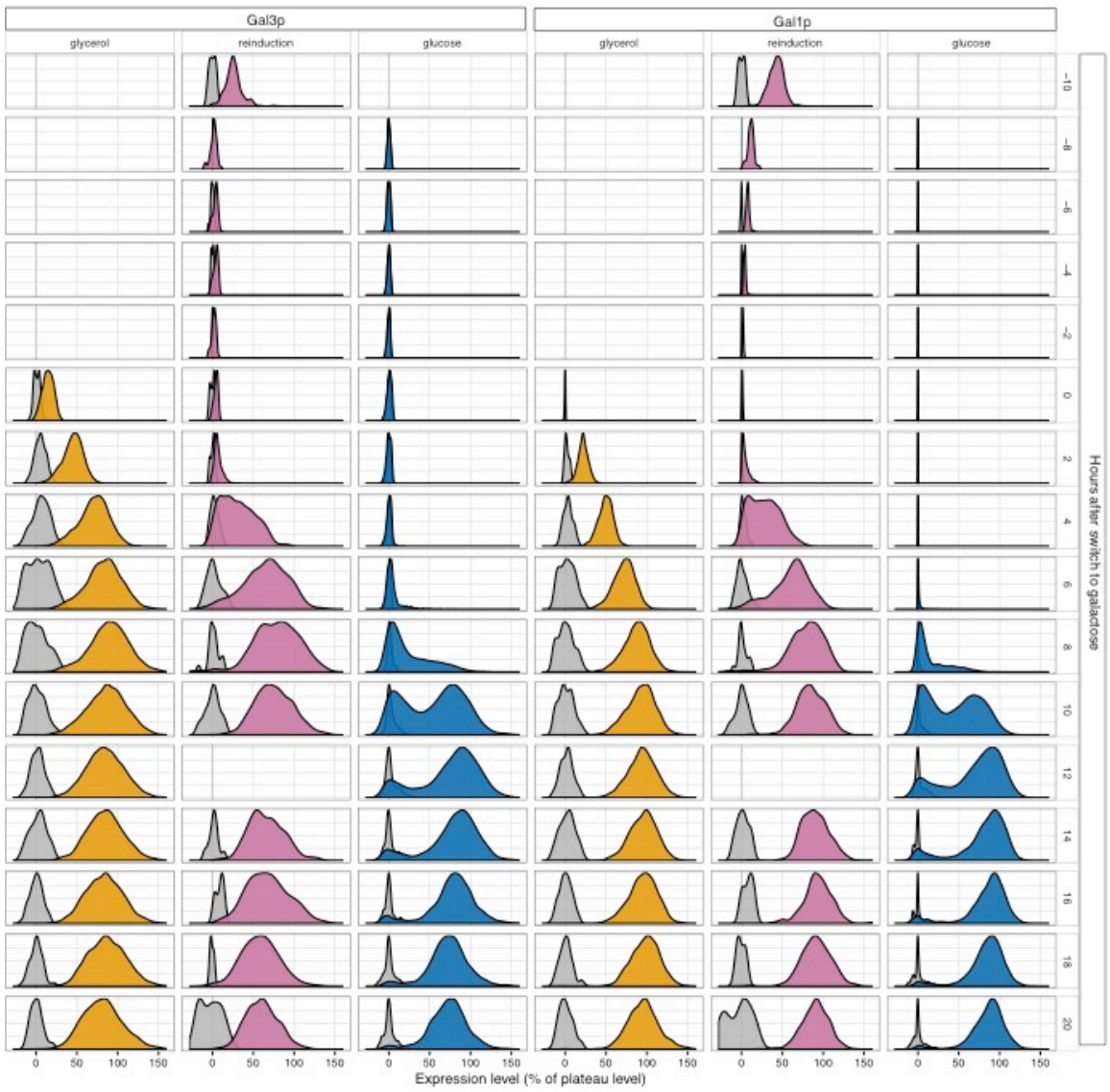

Fig. S2

Population level depiction of galactose network induction. Empirical population densities for Gal3p and Gal1p induction levels in glycerol-history, reinduction, and LTGR conditions. Grey densities represent control cells which have neither yECerulean or 2x-yECitrine. Empty blocks in the reinduction experiemen result from where there were not enough cells to estimate a density reliably. In a few of the frames across the experiments the bright field images did not give a reliable segmentation.

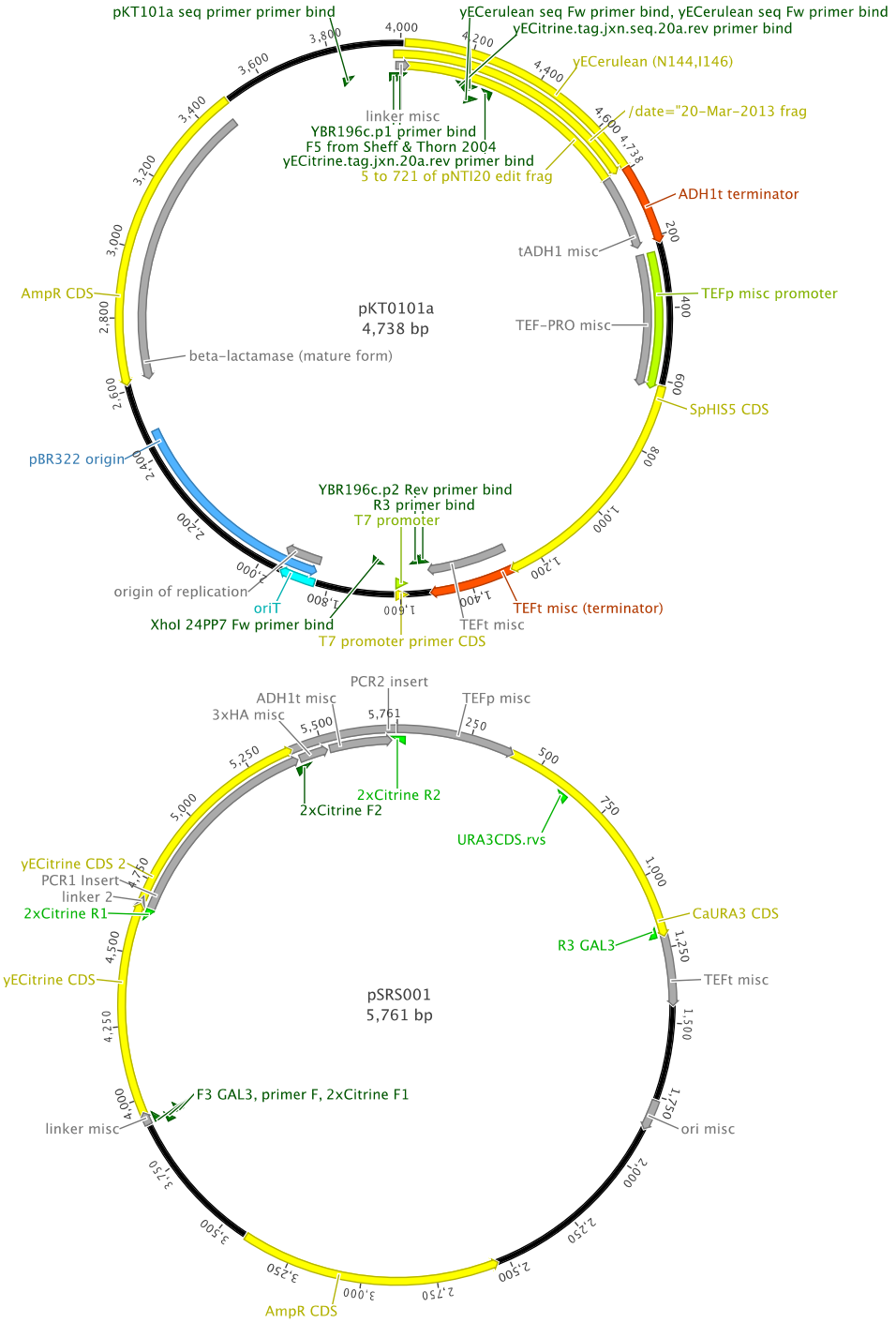

Figure S3

Fig. S3

Plasmids used to generate fusion proteins. pKT0101a was a gift from Natalie Cookson.

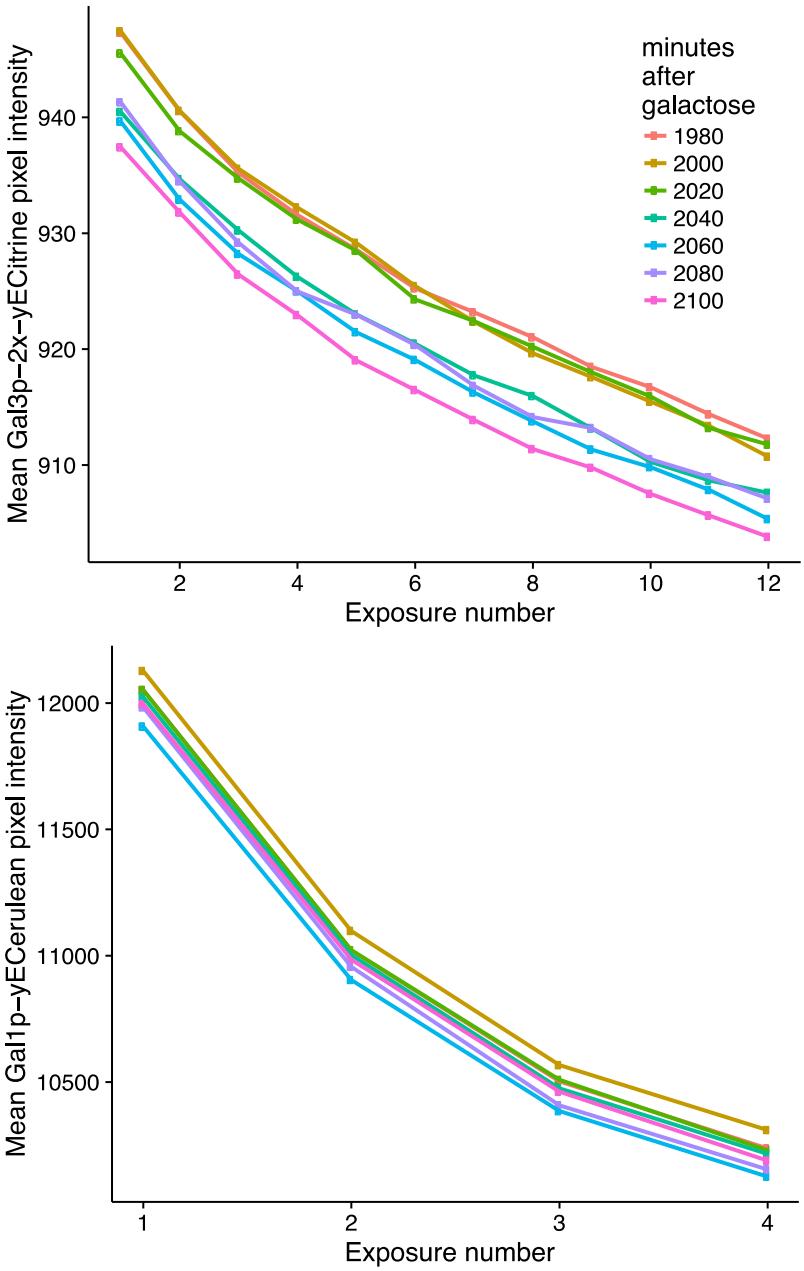

Fig. S4

Photobleaching is not a problem. We took a series of 50-msec fluorescent snapshots at each timepoint. This was needed to capture the full dynamic range of Gal1p-yECerulean but also provided a built-in photobleaching control. Fluorescence levels decline during each series but follow a consistent pattern and recover by the next timepoint 20 minutes later, as shown by the overlapping curves. The curves are mean pixel intensities of seven consecutive frames around 33 hours after galactose exposure in the reinduction condition.

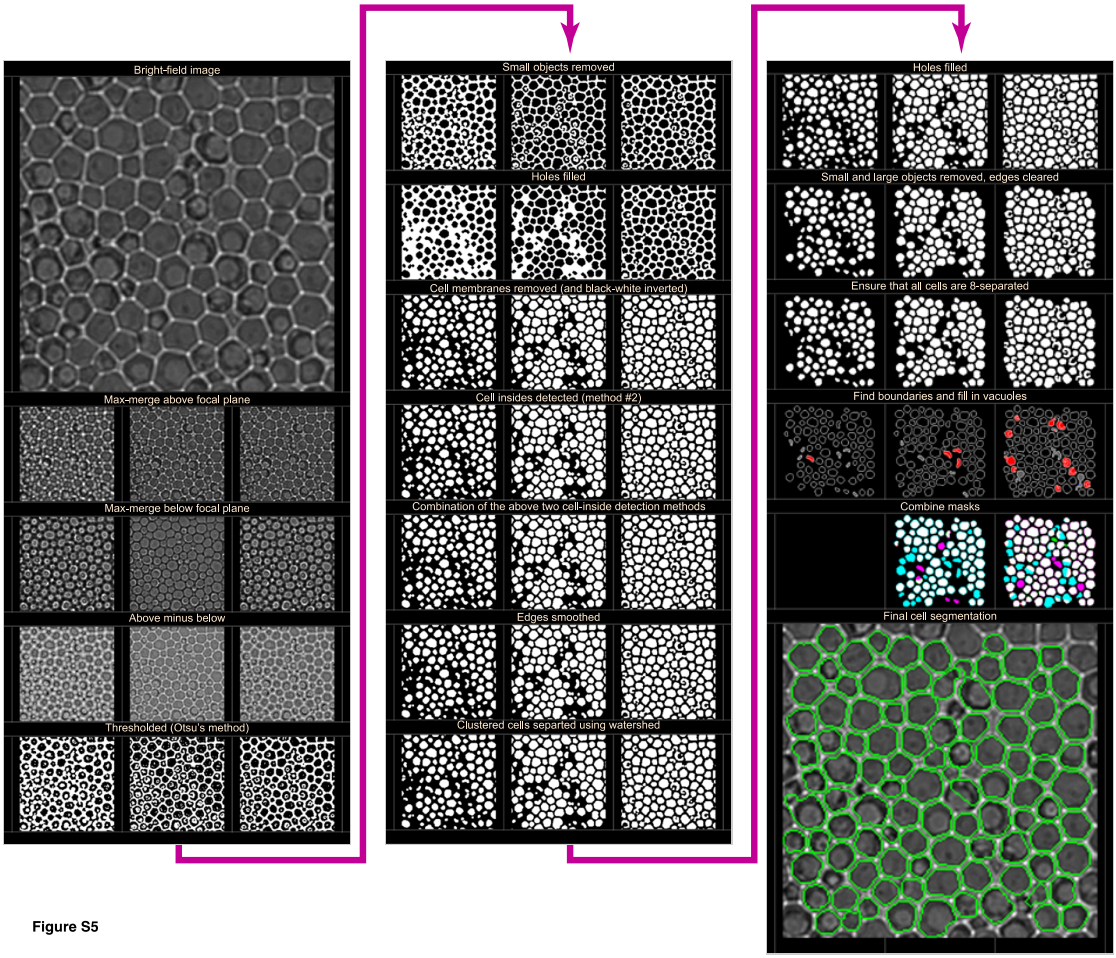

Figure S5

Fig. S5

The cell segmentation pipeline starting from a set of above and below focal plane bright-field images.

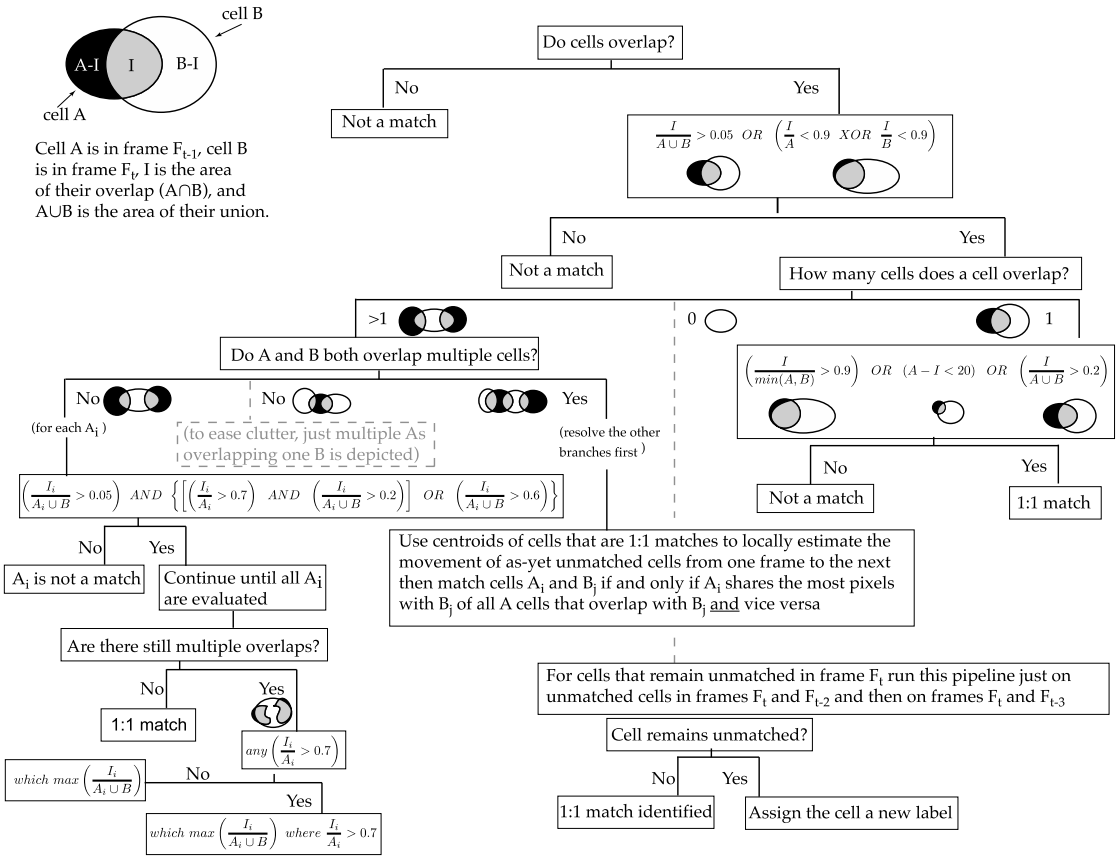

Figure S6

Fig. S6

A flow diagram describing the logic of cell tracking.

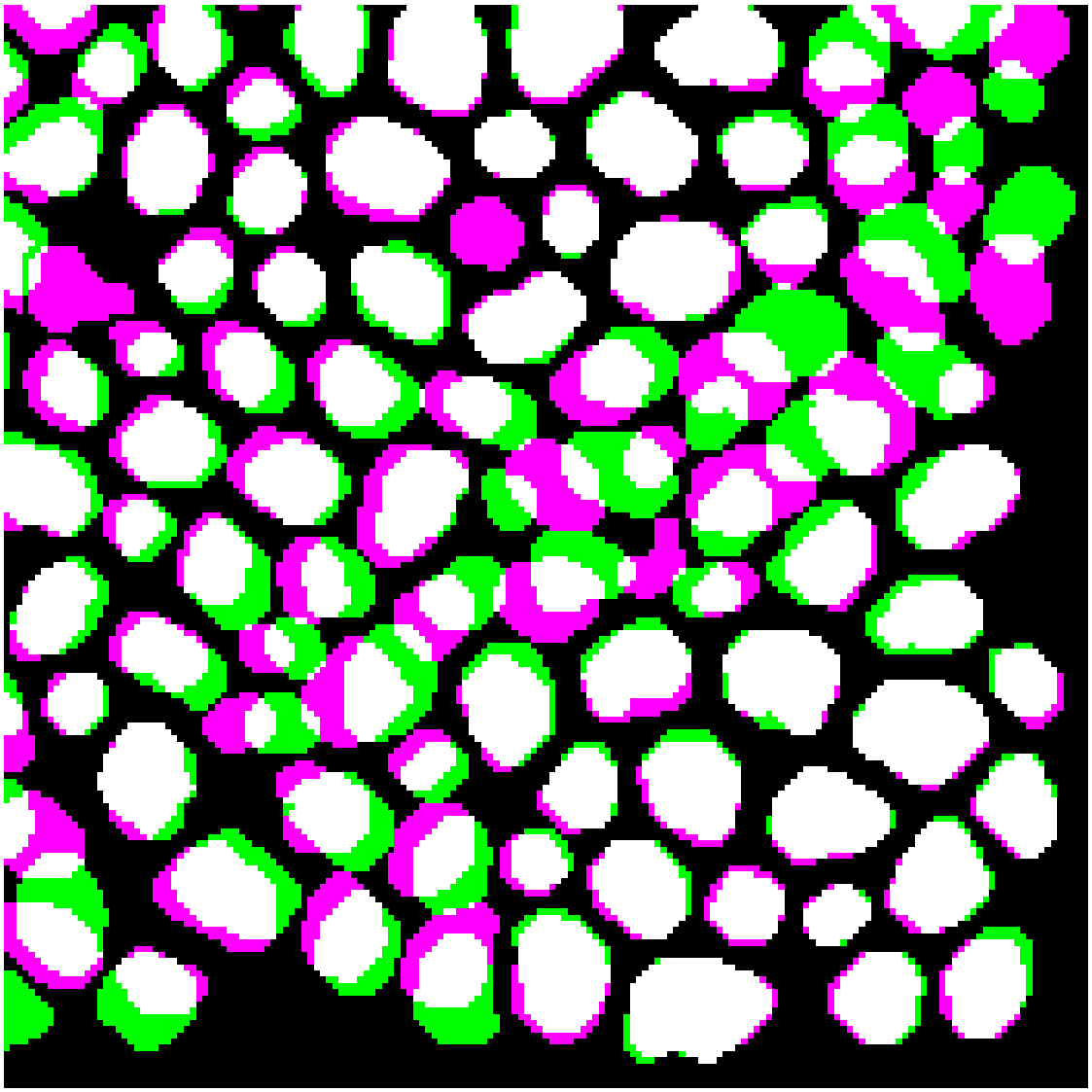

Figure S7

Fig. S7

An example of cell tracking. The green and magenta indicate two successive segmented frames. White represents regions of overlap. This field of view contains many of the kinds of assignment situations resolved by the tracking logic (Fig. S6).

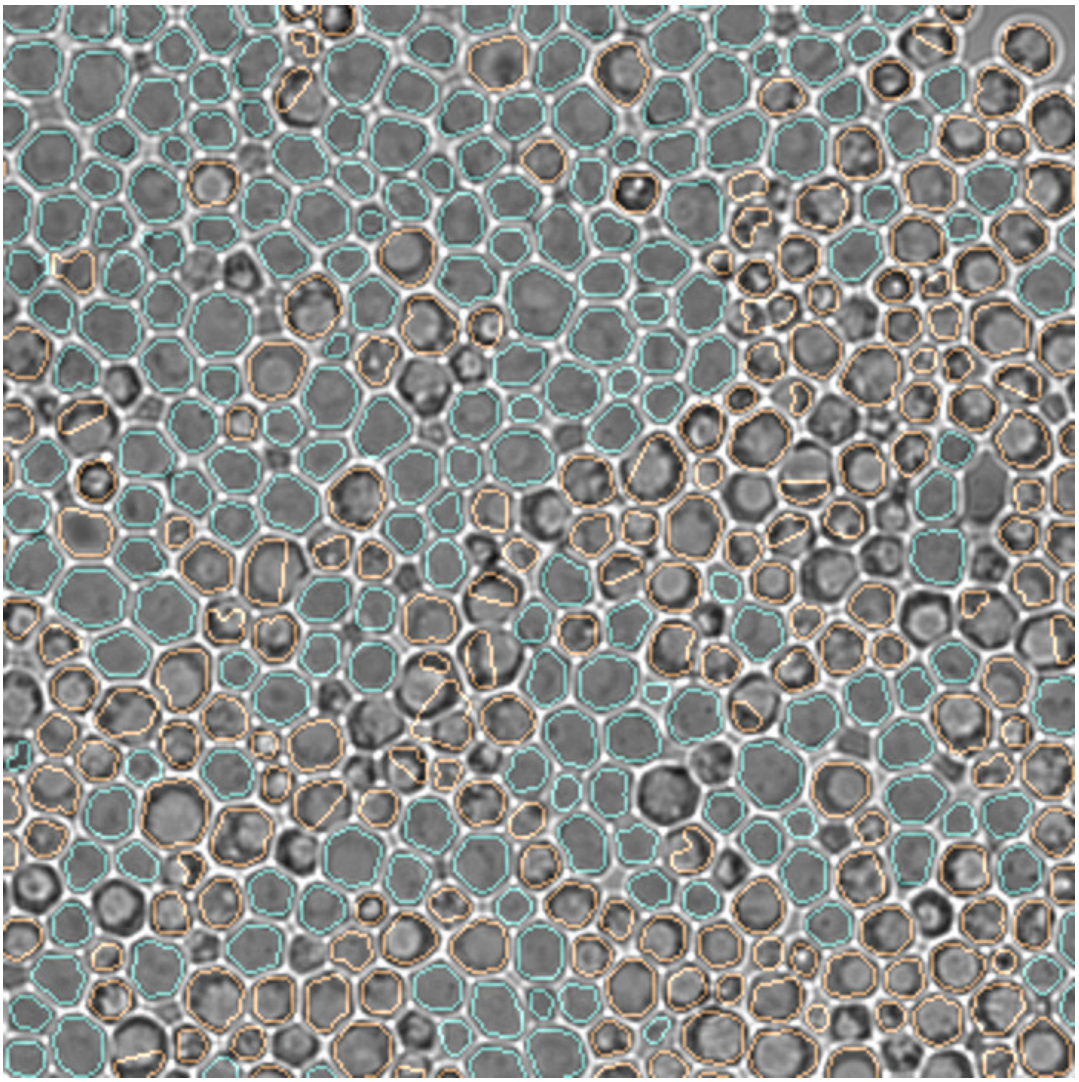

Figure S8

Fig. S8

Alive and dead cells are noticeably different in bright field images. Random forest classification of cells as alive (cyan outlines) or dead (brown outlines) in an image around 6 hours after galactose exposure after LTGR.

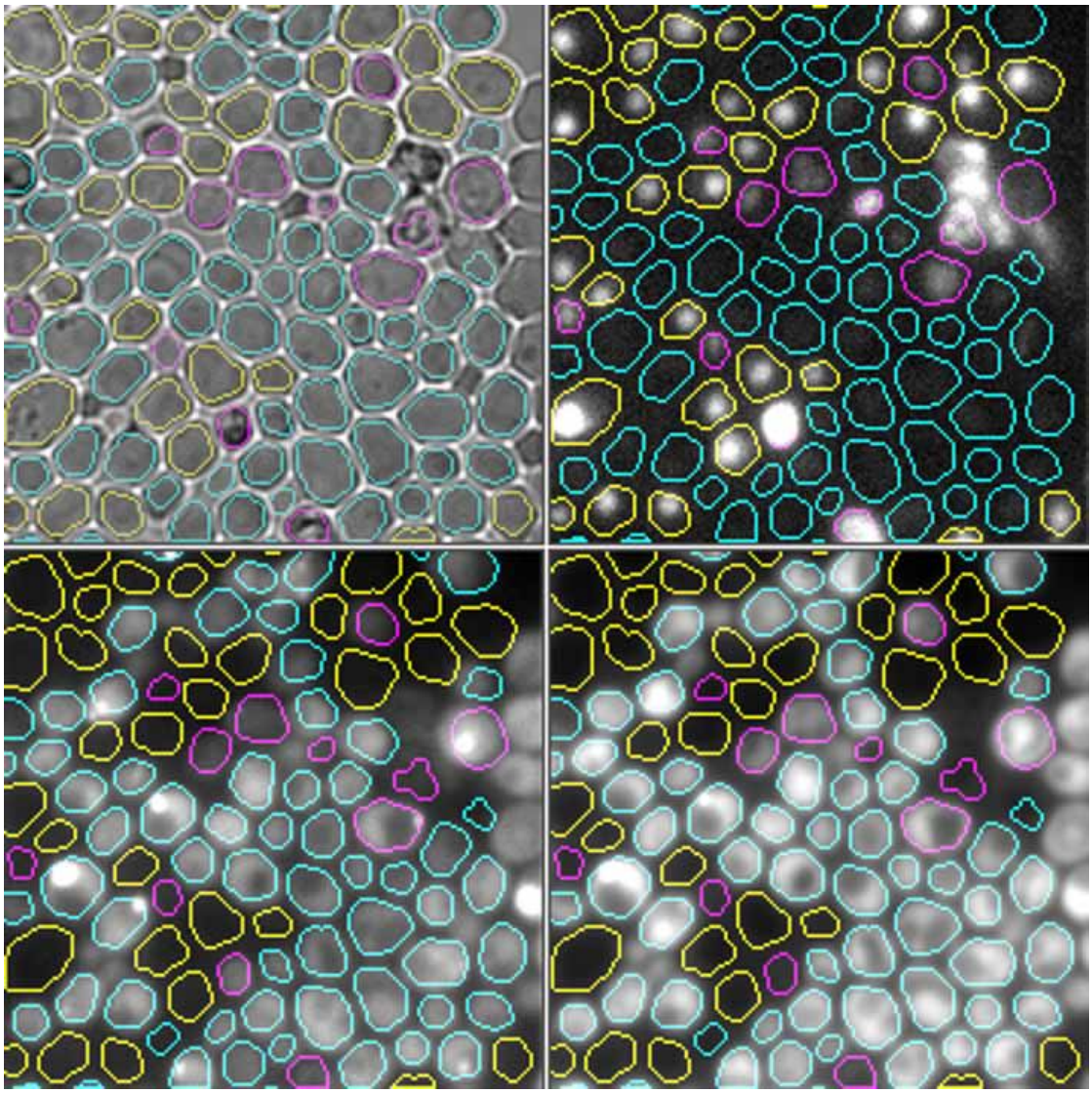

Figure S9

Fig. S9

Random forest classification of cells as experimental, control, or bad based on fluorescence. The images are (clockwise from the upper left): bright-field, mCherry, 2x-yECitrine, yECeruelan. Cyan outlines indicate experimental cells, yellow indicate control, and magenta indicate bad.

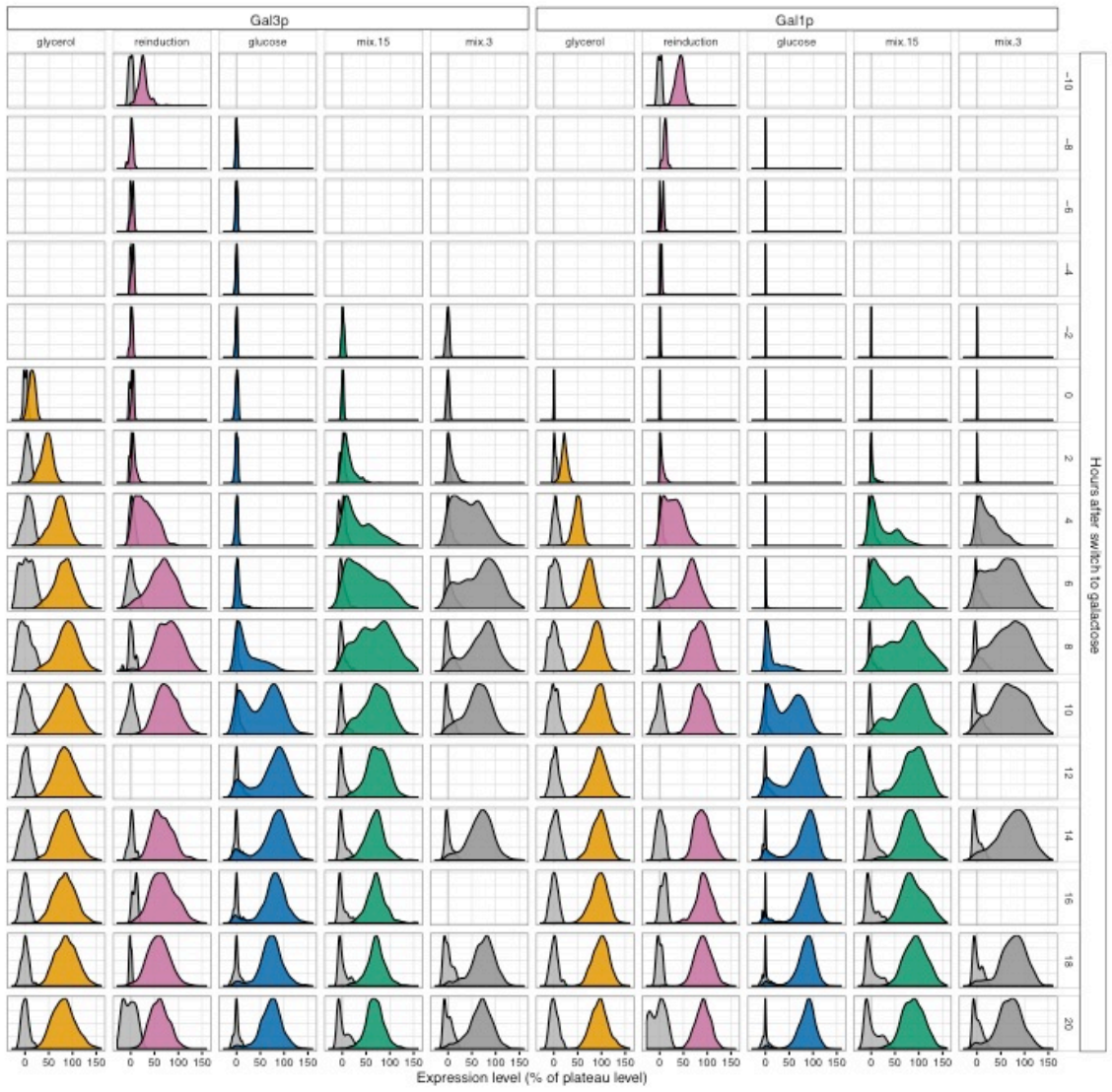

Fig. S10

Population level depiction of galactose network induction. Empirical population densities for Gal3p and Gal1p induction levels in glycerol-history, reinduction, LTGR, and the two glucose-galactose mix conditions. Grey densities represent control cells which have neither yECerulean or 2x-yECitrine. Empty blocks indicate frames in specific experiments where there were not enough cells to estimate a density reliably. In a few of the frames across the experiments the bright field images did not give a reliable segmentation.

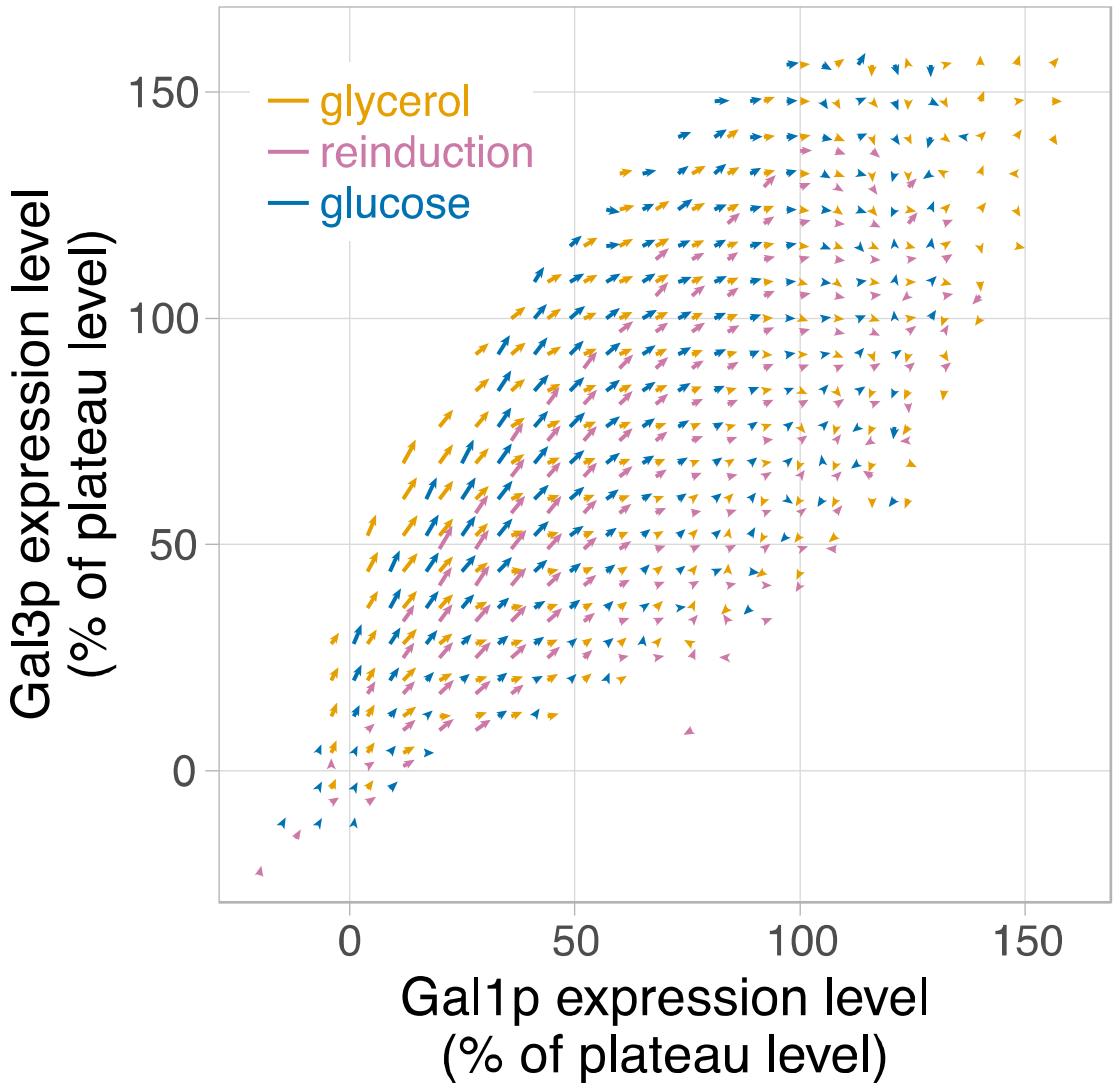

Fig. S11

The vector fields for the initial three experimental conditions plotted together. Bins are slightly offset to separate the vectors and vector lengths are 1/3 the mean unsigned displacement to reduce crowding. Once cells are induced, they move in a consistent direction to plateau levels in all three experiments.

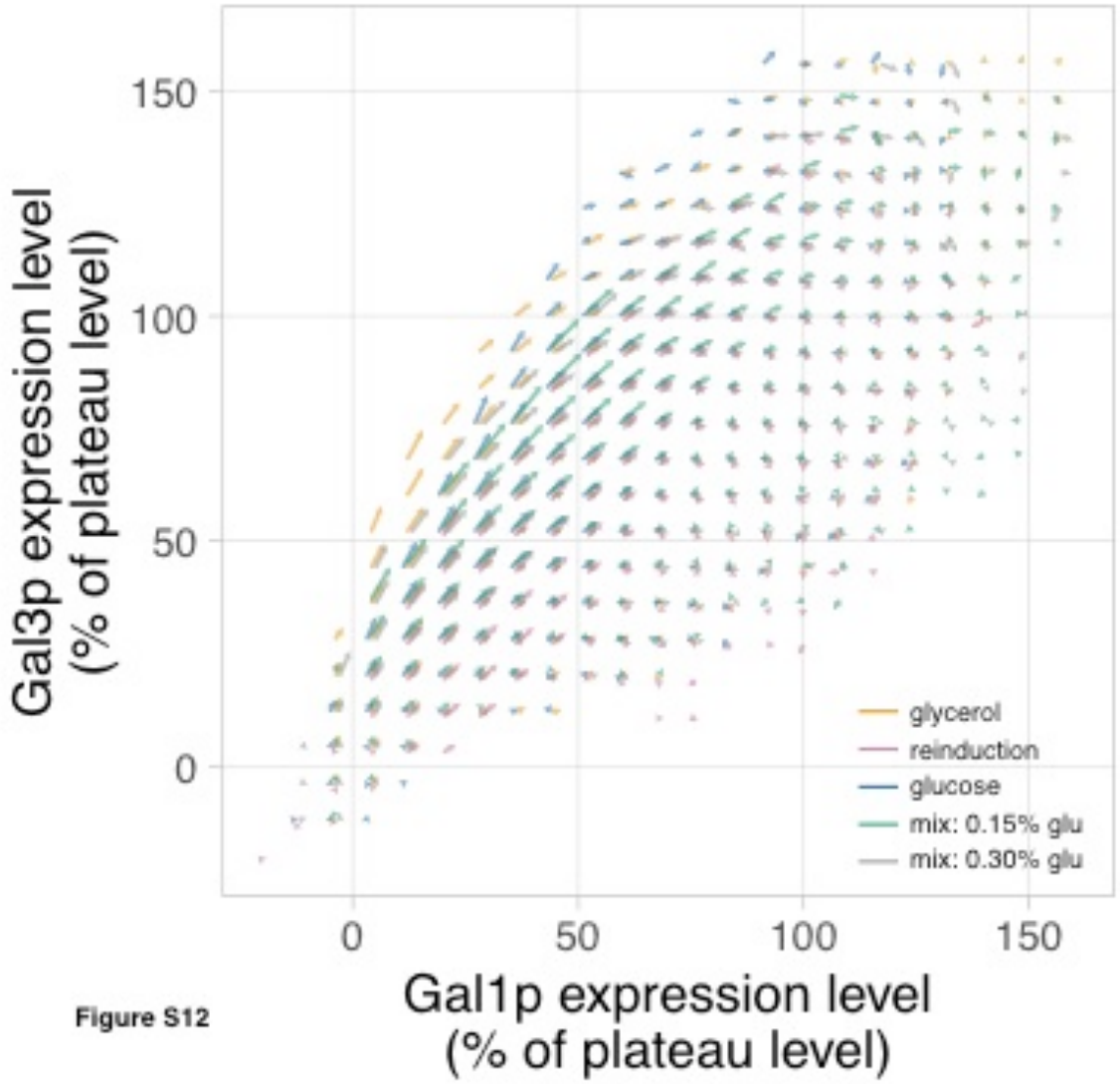

Fig. S12
Vector fields for all five experiments overlaid.

Movie S1

Movies of Gal1p induction in the five experimental conditions. The experiments have been aligned in time with respect to galactose exposure. Experimental cells are lightly outlined in cyan and control cells in magenta. The intensity of color in a lane of the clock corresponds to the fraction of inducing cells that reached 10% plateau level by that timepoint (Fig. 1b).

Movie S2

Movies of bright-field images in the five experimental conditions. These are the same fields of view as in Movie S1. Time runs from 1 hour before galactose exposure to 13 1/3 hours after. The intensity of the colors on the clock reflects the average speed of movement of cells (Fig. 3b), here scaled between 0 and 0.25 microns per minute with the few speeds over 0.25 um/min set to the maximum. Galactose addition is indicated by a yellow flash in all the fields of view.