Supplementary Information

A light tunable differentiation system for the creation and control of consortia in yeast

Chetan Aditya[1,2,3], François Bertaux[2,1], Gregory Batt[1,2] *, Jakob Ruess[1,2] *

1. Inria Paris, 2 rue Simone Iff, 75012 Paris, France

2. Institut Pasteur, 28 rue du Docteur Roux, 75015 Paris, France

3. Université de Paris, 85 Boulevard Saint-Germain, 75006 Paris, France

* These authors contributed equally to this work.

Correspondence: gregory.batt@inria.fr, jakob.ruess@inria.fr

Outline

1. Bioreactor platform and flow cytometry data analysis
2. Microscopy platform and microscopy data analysis
3. Construction, characterization and modelling of the differentiation system
4. Construction, characterization and modelling of Growth arrest upon differentiation system (GAuDi)
5. MPC control
6. Pattern formation setup and results for diverse patterns
7. Construction of the two differentiation programs and quantification of species' prevalence in multi-species consortia
8. Note on precautions undertook during handling and manipulation of the light sensitive strains

Supplementary Note 1

## Experimental platform, flow cytometry, and data analysis pipeline

We used our previously described turbidostat platform[1] to continuously culture cells and conduct time course experiments (Figure 2, 3, 4, 5). Briefly, the platform allows us to monitor 16 cultures in parallel with regular OD measurements and maintain them at a target cell density. The OD measurements and dilution data were used to estimate the growth rate. Further, culture vessels are equipped with LEDs and can be stimulated independently with light. Samples from the vessels are collected in a 96-well plate and, with the help of a pipetting robot, loaded into the cytometer. The cytometer acquisition is controlled with the help of click and point software. Supplementary Figure 1a gives a general overview of the experimental platform.
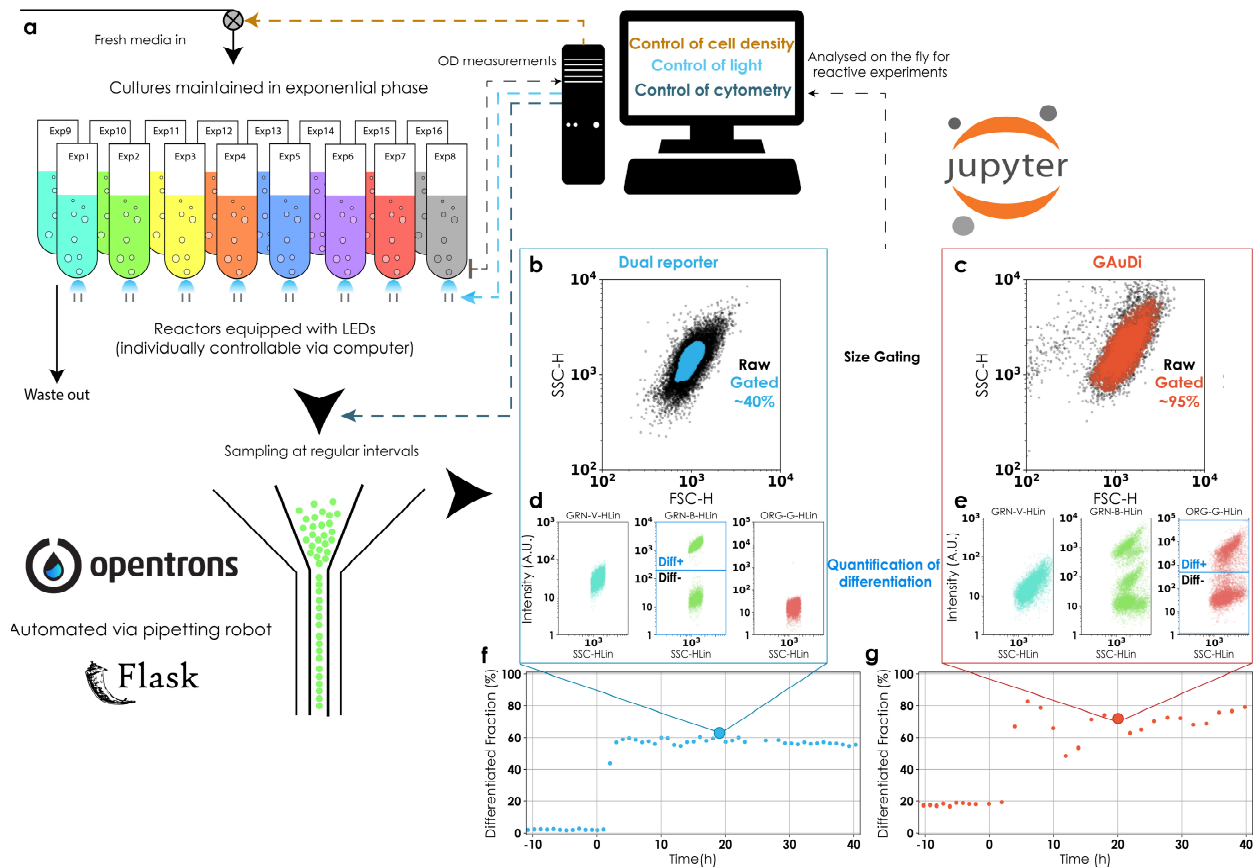
*Supplementary Figure 1. Experimental platform (**a**) and data overview. 16 experiments can be conducted in parallel in reactors with individual OD and LED induction control. The cytometry measurements are automated via a pipetting robot controlled with the Flask app. The raw data from the cytometer is parsed and stored as a csv file or can be analysed online to change experiment conditions such as cell density or induction profiles. All code required for the functioning of the reactors and analysis is developed in Python and implemented using Jupyter notebooks. Representative data from a single timepoint for one experiment each of the original differentiation system (blue box) and GAuDi strain (red box). The FSC vs SSC scatterplot is shown as an indicator of cell size (**b** and **c**). The GRN-B and ORG-G channels are used to detect differentiated cells for the original differentiation system (mNeonGreen fluorescence) and GAuDi (mScarlet-I fluorescence), respectively (blue boxes in **d** and **e**). At this representative timepoint (large circle in **f** and **g**), both exist in two subpopulations. A.U. stands for arbitrary units.*

All cytometry measurements were made with a Guava EasyCyte BGV 14HT benchtop flow cytometer. Settings and gains were kept constant for all the experiments.

5000 events were recorded for each sample unless specified differently. No compensation was used during acquisition. Dilutions were made with a pipetting robot so that the cell density was kept between 200 (to have 5000 events in the acquisition window) and 600 cells/µl (to ensure >90% singlets). Size gating and doublet removal were done using kernel density based methods. Singlets were selected based on deviation from linearity in Forward Scatter Height (FSC-H) vs. Forward Scatter Area (FSC-A). Cells were scored and a threshold was defined above which cells were classified as doublets and removed from analysis. For size gating, 2D kernel density estimates were obtained using SciPy gaussian kde package on Forward Scatter (FSC-H) vs. Side Scatter (SSC-H) and regions of density lower than a threshold were removed (Supplementary Figure 2). The two thresholds were kept constant for all measurements except those made with the GAuDi strain. In the latter case, thresholds were increased to include the entire population. This leniency was warranted because of considerable changes in the side scatter in growth arrested cells (compare Supplementary Figure 1b & Supplementary Figure 1c).
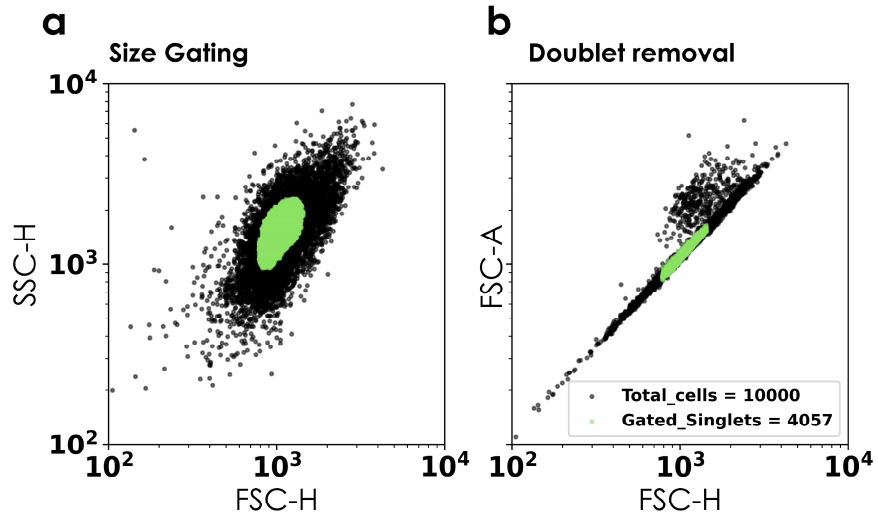
Supplementary Figure 2. Kernel density based gating of data on size (left) and doublet removal (right). Raw data is shown in black. Gated data is shown in green and represent ~40% of all cells.

In cases when multiple fluorescent proteins with overlapping spectra were present inside cells (two recombination cassette strains (Figure 5)), it was difficult to ascertain the differentiation status of the cells. To this effect, we implemented a deconvolution approach previously described in Bertaux et al. 2020[1]. Briefly, 4 single fluorescent protein control strains (mCerulean, mNeonGreen, mVenus, mScarlet-I) with the same promoter and terminator, and integrated in the same locus, were used to determine the spectral signature of each fluorescent protein across the 12 channels of the cytometer (Supplementary Figure 3).
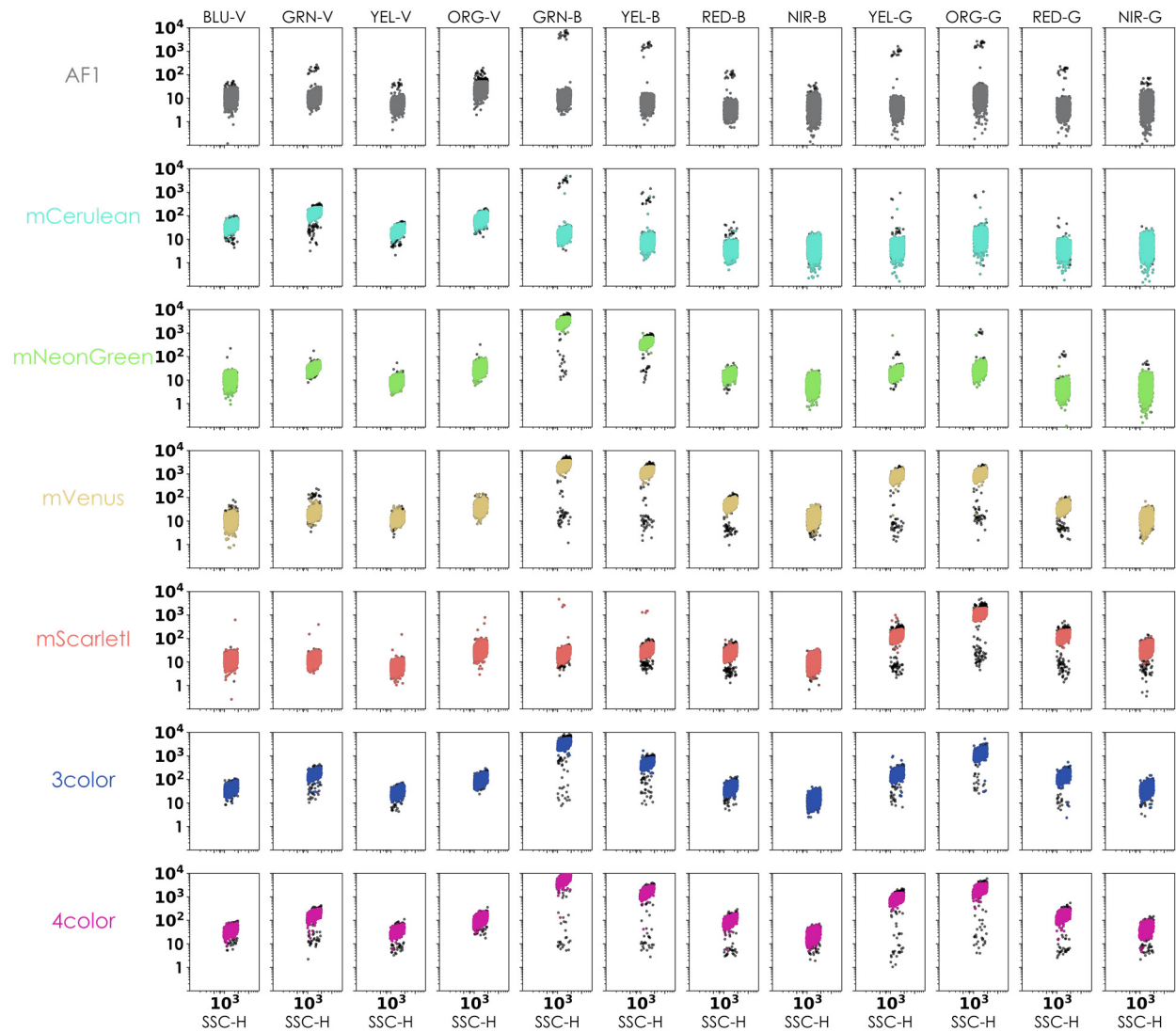
Supplementary Figure 3. Population distributions of measured intensity by different channels of the flow cytometer for an auto-fluorescence (grey) strain, four single fluorescent protein strains with identical promoters and terminators, mCerulean, mNeonGreen, mVenus and mScarletI, and a four colour strain that has all 4 fluorescent proteins listed above. Data was filtered for outliers using mean absolute deviation metric (4.8 medians) to remove sporadic cross-contamination from other cultures during sampling. We note that filtering was performed only for the experiment used to define spectral signatures.

These signatures were then used in a linear algebra framework to calculate the individual fluorescence of each fluorophore in a strain harbouring all 4 of the fluorescent proteins. The values for single colour control and the 4-colour strain were in good agreement after deconvolution (Supplementary Figure 4).
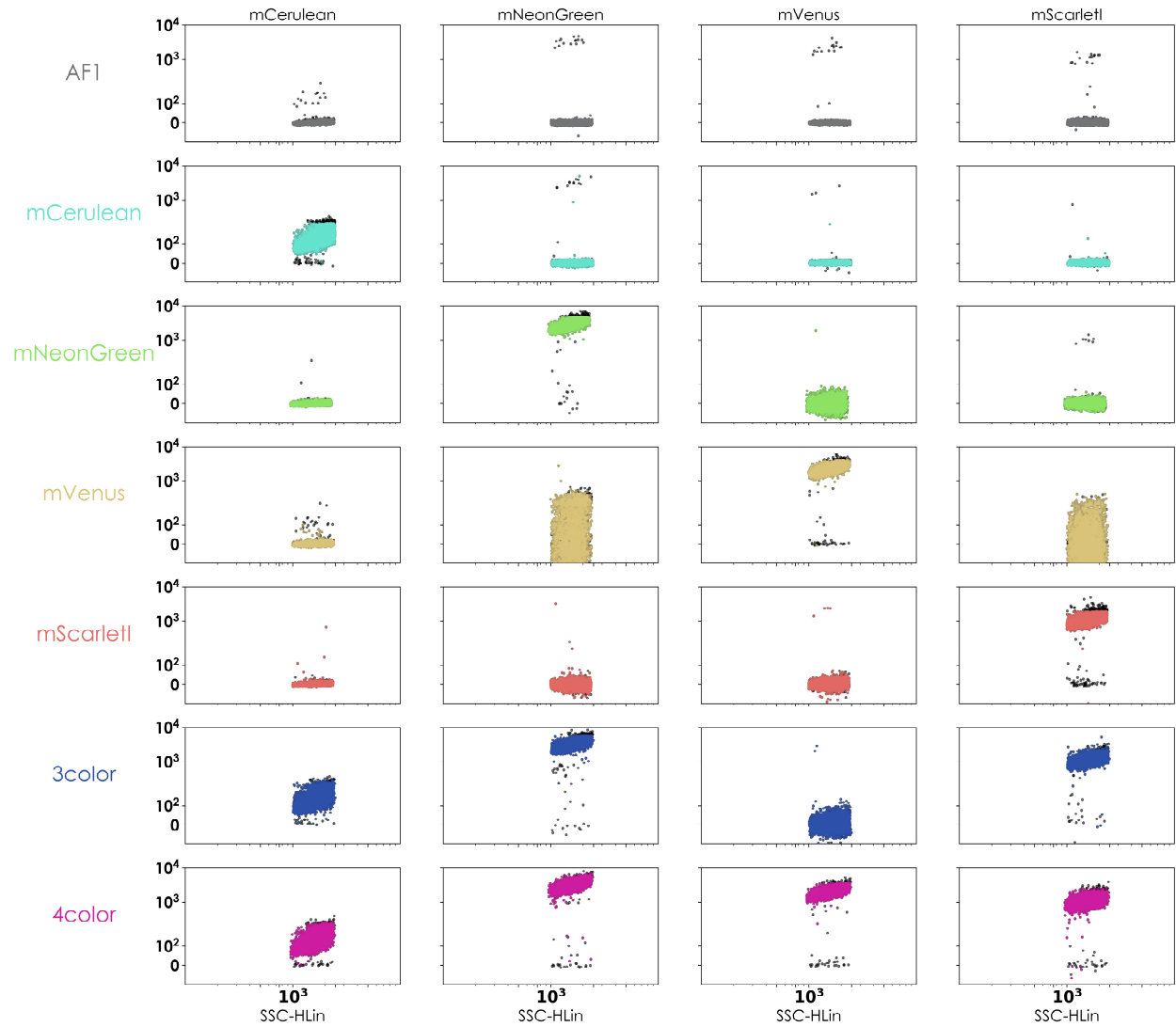
Supplementary Figure 4. Population distributions of deconvolved fluorescence intensity. Filtered channel data was used to compute spectral signatures of each fluorescent protein. Deconvolution also reduced size related heterogeneity in fluorescence.

Supplementary Note 2

Microscopy and image analysis pipeline

Microscopy was done on the inverted microscopy platform Leica DMi8 S. Illumination was effected via CoolLED PE-4000 illumination system. Cells were cultured in CellASIC ONIX plate for haploid yeast cells (Y04C) or µIbidi slides (Supplementary Figure 5).
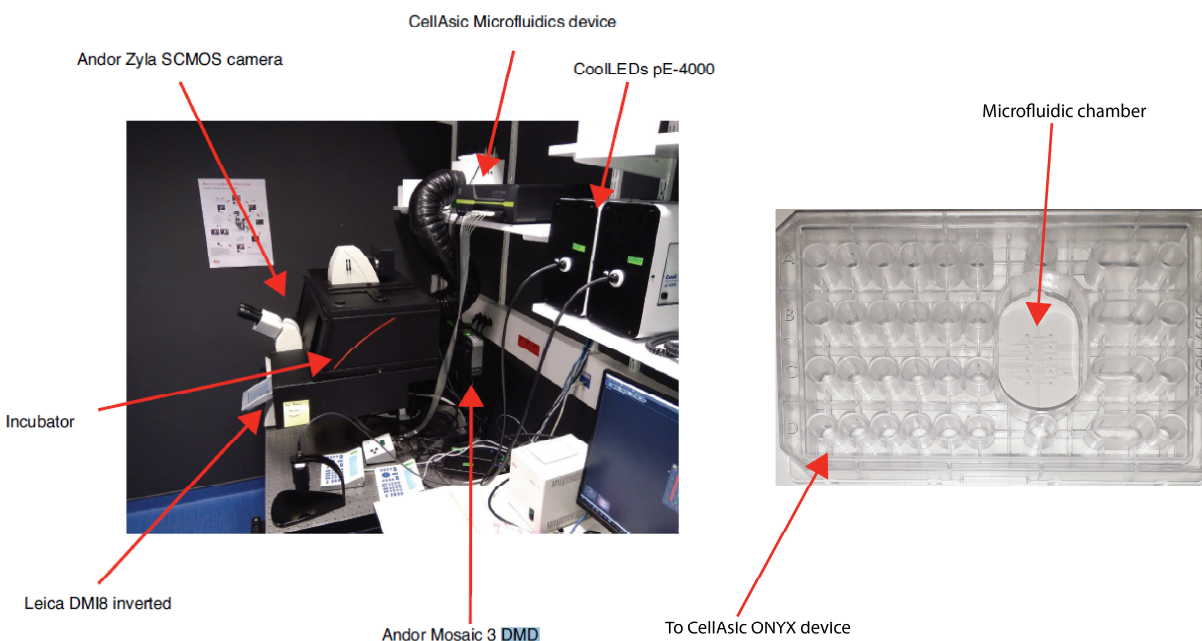
Supplementary Figure 5. Microscopy platform (left)[2] and microfluidic plate (right).
The spectra of the fluorescent proteins used in this study did not significantly overlap and therefore imaging could be done without the risk of crosstalk or bleed through.

Supplementary Table 1. Illumination and imaging conditions.

Protein	Filter Cube	Excitation (nm)	Emission (nm)	Dichroic (nm)	Light Source (nm) - Intensity	Exposure (ms)	EL222 Induction
mCerulean	CFP	436/20	480/40	455	435-6%	600	Yes
mNeonGreen	YFP	500/20	535/30	515	500-9%	600	Yes
mScarlet-I	Rhodamine	546/10	585/40	560	550-10%	600	No

During time-lapse live cell imaging the chamber temperature was maintained at 30°C. We used an in-house software, called MicroMator, for the automated acquisition and cell tracking[2]. Cell segmentation was achieved via SegMator, an in-house U-net based segmentation algorithm[2] (Supplementary Figure 6b). Once cells were segmented, median pixel fluorescence was computed for each cell (Supplementary Figure 6c). Cellular fluorescence was used for ascribing differentiation status to cells (Supplementary Figure 6d). Cells were considered differentiated once a threshold value for fluorescence was exceeded (Supplementary Figure 6e).

We observed that growing in the constrained environment of the microfluidic chamber did not have a significant growth defect (Supplementary Figure 6f & g).
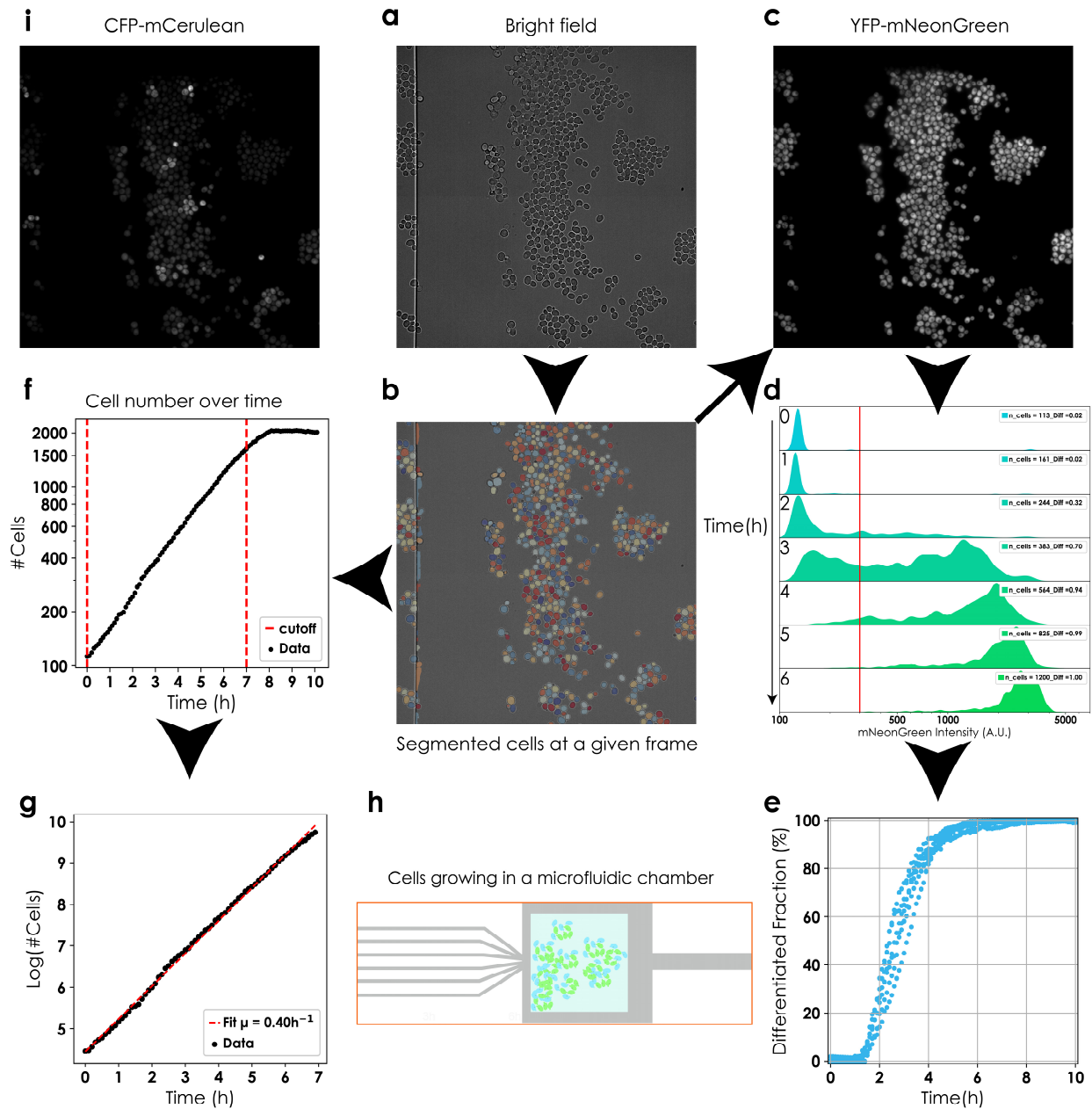
Supplementary Figure 6. Segmentation and growth in the microfluidic plate. *SegMator was used for segmentation and subsequent image analysis[2]. Numbers of cells, over time, in a field of view, were computed by counting the number of segments in each frame. Growth rates were computed by fitting a line to log (#cells) as a function of time. Mean growth rate ± s.d. across 8 fields of view over 2 experiments was found to be 0.394 h^-1 ± 0.029. Median cellular fluorescence was calculated from segmented images, and fluorescence in the YFP channel was used to distinguish between differentiated and non-differentiated cells. Cells possessing more than 300 units of median fluorescence (A.U.) were considered differentiated. This differentiation fraction was calculated at each frame to illuminate differentiation dynamics (Figure 1e). A.U. stands for arbitrary units.*
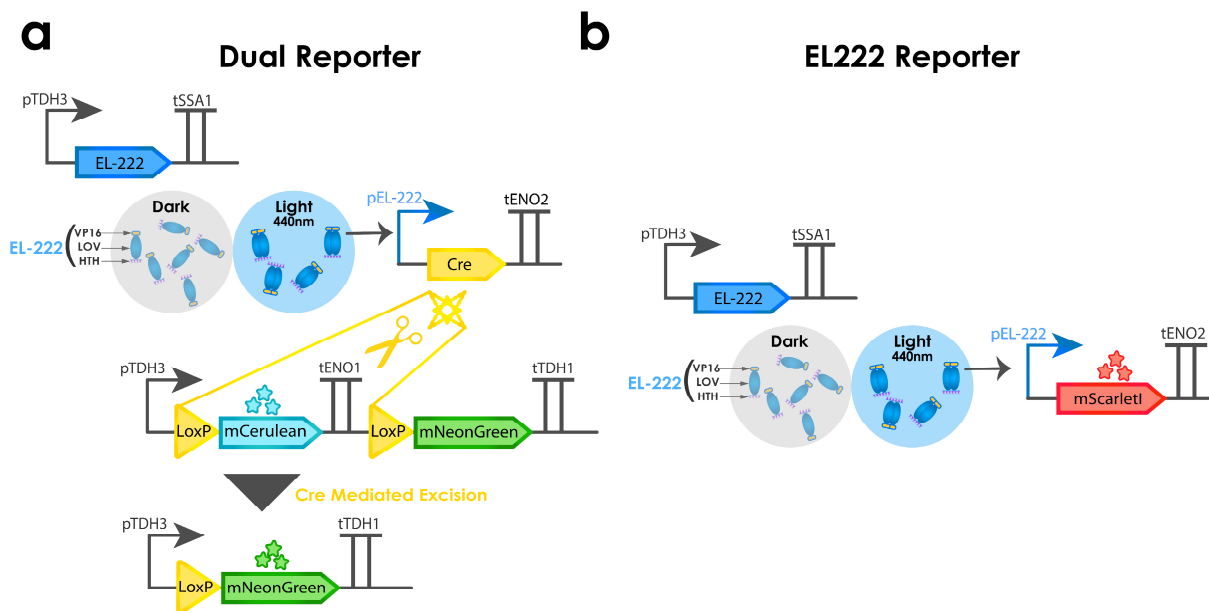
Supplementary Note 3

Construction, characterization and modelling of the differentiation system

While several solutions exist for optogenetic expression in yeast[3–6], we decided to use EL222 to drive Cre because it possessed several desirable features like,

- tighter control over background activity in the dark (range of promoters)

- control over the strength and/or the variability in expression

- it is a homodimer (therefore requires cloning of a single gene) and does not rely on the addition of expensive chromophores

We chose three EL222 promoters published in Benzinger and Khammash (2018)[3] to drive Cre recombinase, namely pEL222 3x binding sites (bs), pEL222 6x bs and pEL222 5x bs Gal1. Of these three, only EL222 5x bs Gal1 led to low leakage.

Supplementary Figure 7. a. Differentiation system (Dual reporter) and b. EL222 activity reporter. ELL222 promoter used in both reporters was pEL222 5X bs Gal1.

We observed no significant toxicity or growth defect due to expression of EL222, Cre or blue light

induction. We also did not notice a growth difference between differentiated and non-differentiated
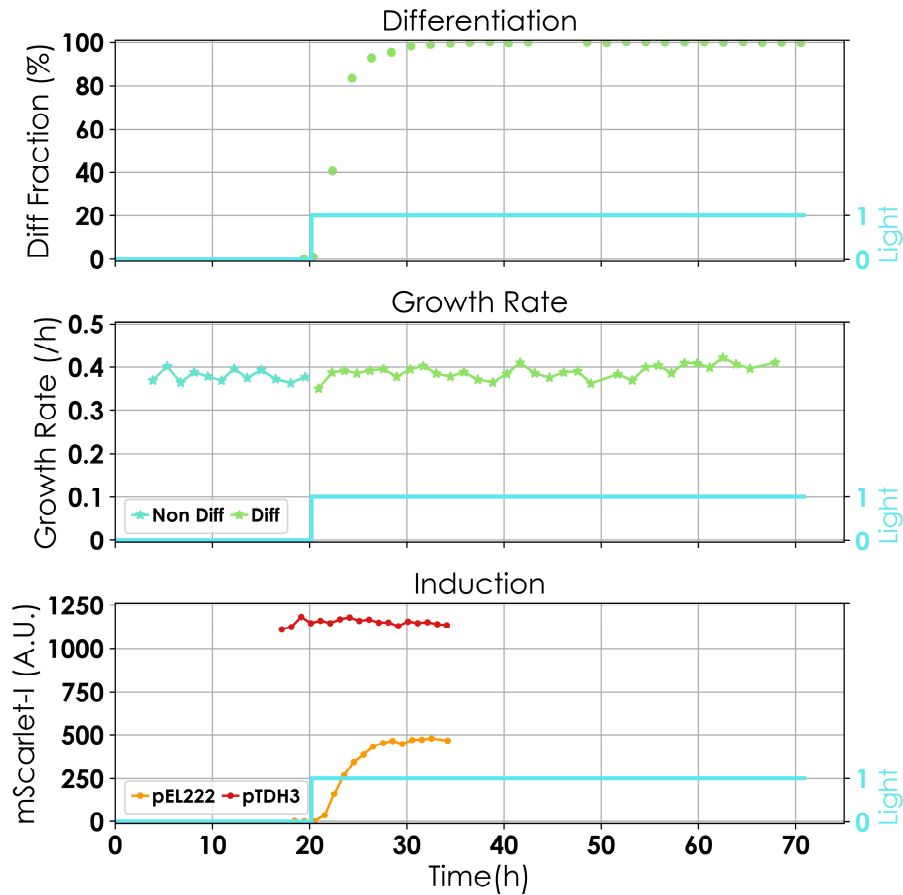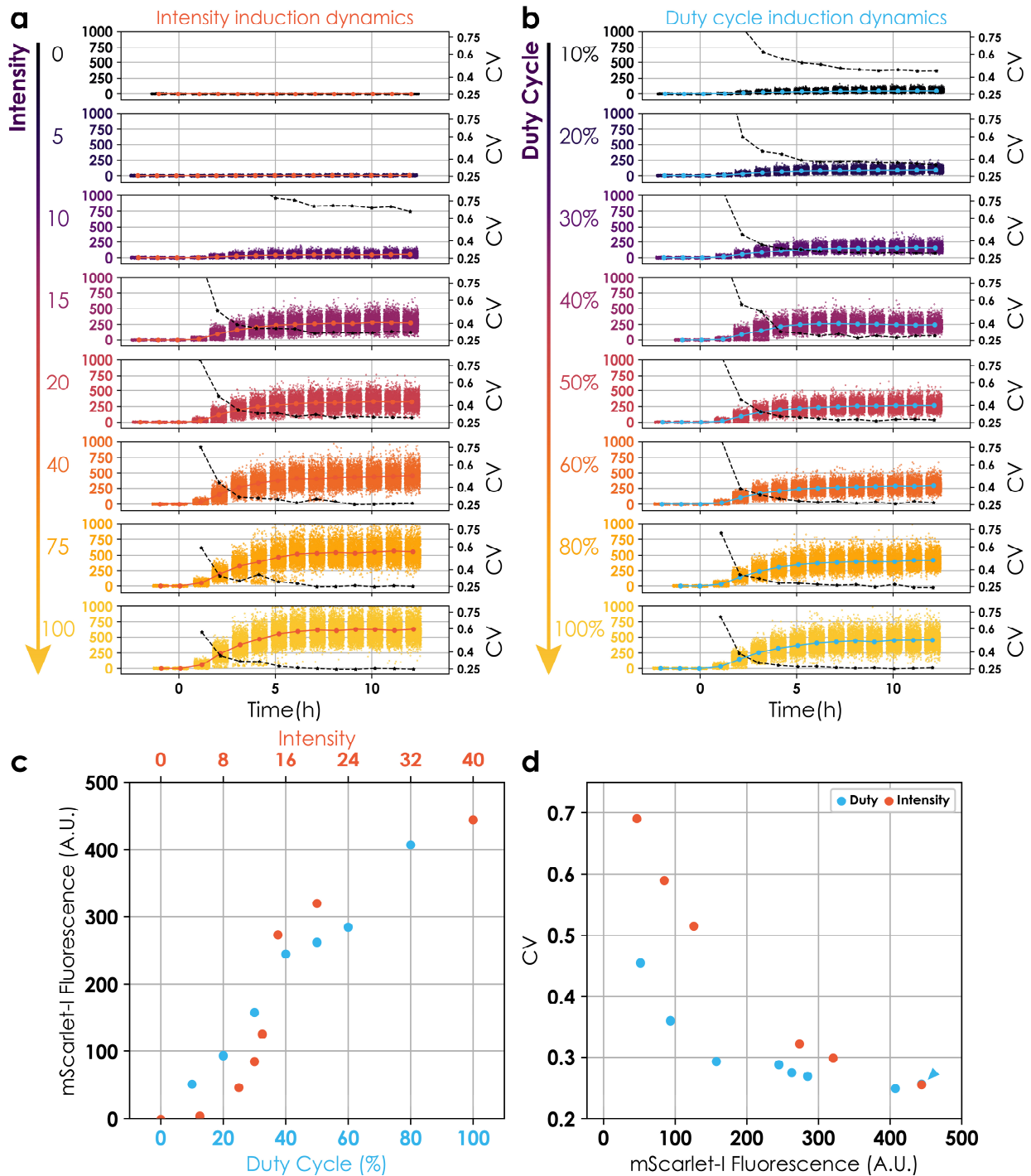
cells.

Supplementary Figure 8. No growth defect is found for differentiated cells, growing under blue light and expressing Cre (top and middle panel). The expression strength of the EL222 promoter is approximately 40% of the one of the TDH3 promoter (bottom) when induced with light of intensity 40. A.U. stands for arbitrary units.

Consequently, we decided to study the induction behaviour of this promoter. We put a red

fluorescent protein (mScarlet-I) under the control of this promoter (EL222 reporter) and

characterized response to various light intensities and duty cycles in cells carrying the EL222

reporter in continuous cultures. By duty cycle is meant a quantity between 0 and 1 that reflects the

percentage of light shown in a given period of 2h. For instance, a duty cycle of 0.5 for a period of

2h signifies 1h of light followed by one hour of darkness. Dynamics of population average fluorescence emerging from various different light intensities and duty cycles are displayed in Supplementary Figure 9a & b, respectively.

Overall, we obtained results that are consistent with Benzinger and Khammash (2018)[3]:

1. Steady state average fluorescence is linear in the duty cycle and shows a sigmoidal response in increasing intensity of light (Supplementary Figure 9c).

2. Cell-to-cell heterogeneity is larger with intensity modulation for the same mean level of fluorescence compared to duty cycle modulation (Supplementary Figure 9d).

Supplementary Figure 9. Characterizing EL222 induction behaviour. **a** and **b.** Population level induction dynamics for intensity modulation (**a**) and duty cycle modulation (**b**). Small circles represent fluorescence from individual cells. Big circles (red and blue) represent the mean fluorescent values at each time point (connected with red and blue lines, respectively). Black lines marked with stars denotes the CV (measured on the right y-axis). CV values prior to induction were not reliable due to low average population fluorescence. **c.** Red circles represent mean population fluorescence for duty cycle modulation and blue circles stand for mean population

*fluorescence for intensity modulation. Each condition has a single replicate. **d.** CV plotted against mean for intensity modulation (red) and duty cycle modulation (blue). A.U. stands for arbitrary units.*

Next, we tested the differentiation system for different intensities and duty cycles in exponentially growing cells in batch in duplicates. These experiments were done in batch because it allowed us to perform multiple induction experiments on the same day. Briefly, precultures could be diluted in a large volume and partially split to give 16 cultures in falcon tubes that could be induced with different profiles and removed from the batch inducer and placed in the dark while the same preculture could be used to conduct the next set of experiments. The resulting differentiation fractions (in duplicates) were reminiscent of EL222 induction profiles.
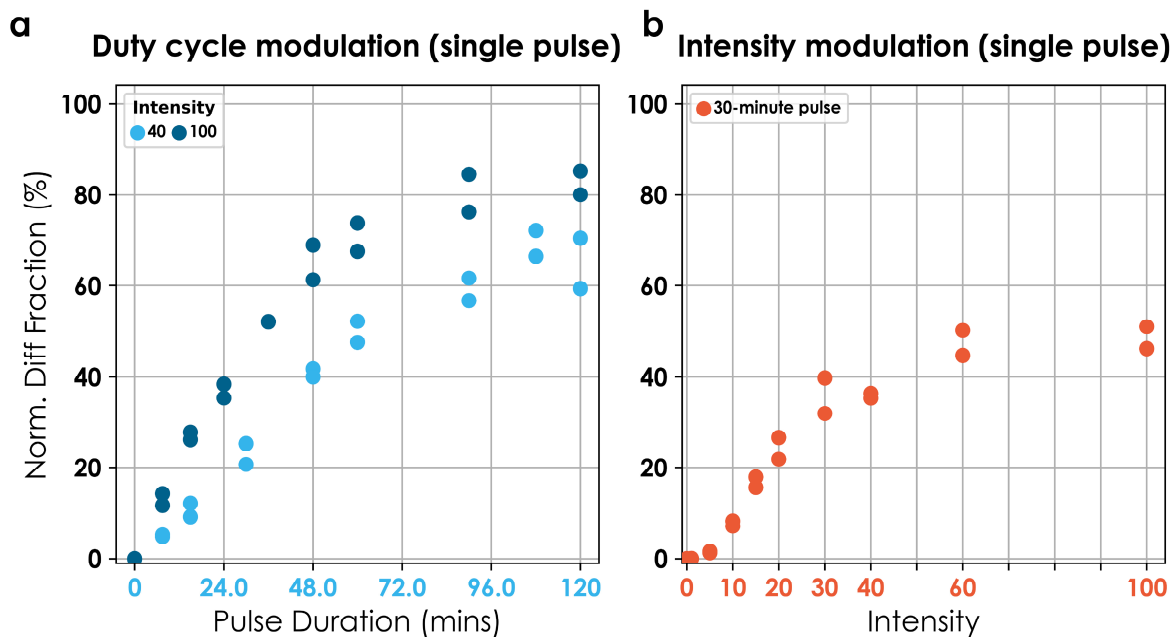
Supplementary Figure 10. Single pulse duration vs intensity modulation of differentiation behaviour. Solid circles are differentiation fractions observed. All conditions have two replicates. Differentiation w.r.t. pulse duration modulation over a period of 2h was assessed for two different intensities, 40 and 100. We noticed that the differentiation fraction increases linearly with pulse duration up to a certain point whence it plateaus. The slope of increase in differentiated fraction with respect to pulse duration was steeper for intensity 100. The linear response range was also smaller for intensity 100. Amplitude modulation led to a more switch like response of the differentiated fraction. Intensity 10 resulted in no significant recombination whereas intensity 30 resulted in 80% differentiation of the maximum intensity (100).

For all experiments in the main text, an intensity of 40 was used.

We compared the efficiency of our system with existing systems of optogenetic recombination in yeast. We found that our system was the most efficient light based recombination system to date. The table compares the efficiency in yeast. Since, the systems reported in Taslimi et al. 2016[7] and Kawano et al. 2016[8] consisted of results for mammalian cells (HEK293 and COS-7, respectively), data for these two is taken from Duplus-Bottin et al. 2021[9]. Leakage in dark is defined as percentage of cells identified as recombined after culture in the dark.

Supplementary Table 2. Characteristics of opto-inducible/photo-activable recombinases in yeast[4,7–9].

Publication	System	Growth conditions	Efficiency (mean ± s.d.) - Induction time	Leakage in dark (mean ± s.d.)
Ref 6 (Figure 3), adapted from Ref 4[†]	split Cre CIB1_CRY2	Stationary liquid culture	1.6% ± 0.8 at 90 mins	1.3% ± 0.5 over 24h of dark culture
Ref 6(Figure 3), adapted from Ref 5[††]	split Cre pMag-nMag	Stationary liquid culture	21.2% ± 5.8 at 90 mins	7.1% ± 1.1 over 24h of dark culture
Ref 7 (Figure 2) original study	split Cre PhyB-PIF3*	Stationary liquid culture	46.7% ± 5.3 at 24 h**	6.7% ± 2.6 over 24h of dark culture
Ref 6 (Figure 3 & 4) original study	destabilized Cre fused to asLOV2 (LiCre)	Stationary liquid culture Exponential liquid culture	41.2% ± 2.8 at 40 mins 66.7% ± 3.7 at 90 mins 66.8% ± 3.3 at 180 mins 7.6% ± 2.1 at 40 mins	0.7% ± 0.2 over 24h of dark culture
This study	EL222 inducible WT Cre	Exponential liquid culture	**43.1% ± 2.7** at 40 mins **76.8% ± 1.7** at 90 mins **94.4% ± 0.9** at 180 mins **99.7% ± 0.1** at 240 mins	**0.06% ± 0.05 over 72h of dark culture**

[†] Original study performed in HEK293 cells.
[††] Original study performed in COS-7 cells.
*Addition of PCB was necessary to achieve recombination with PhyB-PIF3 system. Results shown here correspond to a concentration of 25µM. Efficiency could be improved to 89.3% ± 4 by increasing the concentration of PCB to 100µM. This also resulted in an increased the background activity to 9.7% ± 2.4.
** Induction was carried out by delivering a single 5-min pulse of red light and followed by 10s pulses every 5 minutes for 24h.

Next, we asked the question whether the functioning of the system remains predictable in light of 15-fold variation in cell density. We found that the cell density does not significantly affect the

differentiation dynamics strongly suggesting that the system can be used to for dynamic control purposes at different ODs (Supplementary Figure 11a &b).
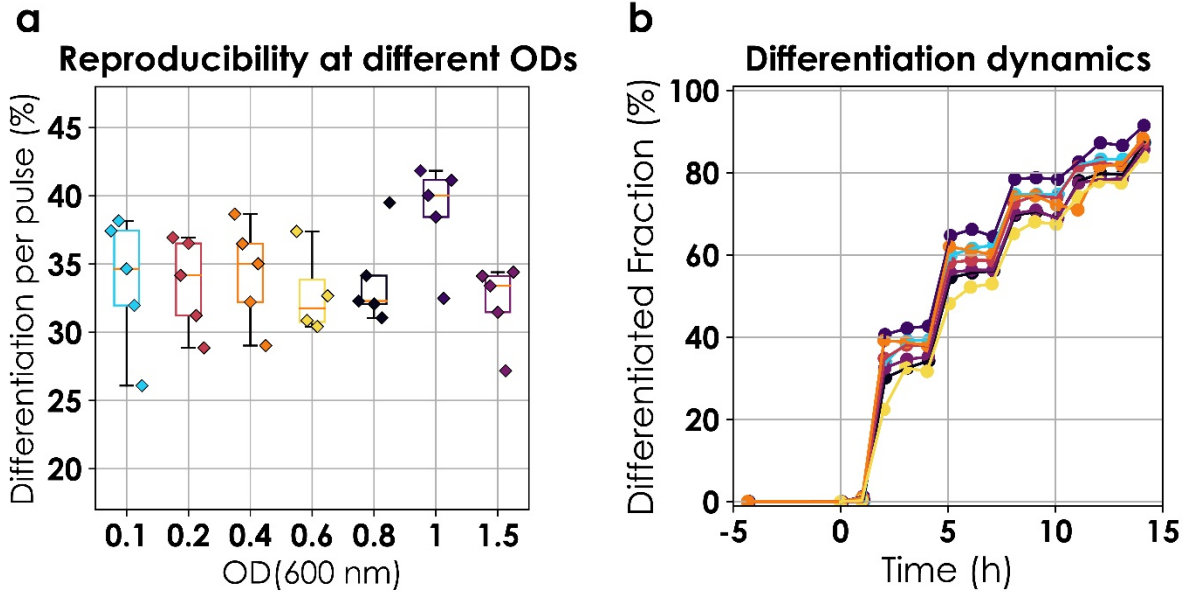
*Supplementary Figure 11. **a**. Comparing the variability in response to the same (30-min) pulse when cells are cultured continuously at different ODs. For each OD, a unique experiment was performed in which cells have been exposed to five 30min light pulses (n = 5) delivered 3h apart from each other and the differentiated fraction in response to each pulse was quantified (diamonds). Lines, boxes and whiskers denote median, quartiles and extreme values, respectively. **b**. Dynamics of differentiation at different ODs in response to repeated pulses of 30 mins that were used to quantify differentiation per pulse in a. Colors in b correspond to the ODs in a (see x-ticks).*

Next, we developed a model to assess whether our system is predictable i.e. whether we could predict the dynamics of the differentiated fraction emerging from given light inputs. Since the model was primarily developed for the purpose of deploying it in model predictive control of the population composition, we decided to use a simple ODE model that tracks population dynamics.

$$\dot{g}(t) = \mu g(t) - \boldsymbol{U}(t)r_{diff}g(t) - \lambda g(t)$$

$$\dot{p}(t) = \mu p(t) + \boldsymbol{U}(t)r_{diff}g(t) - \lambda p(t)$$

$$n = g(t) + p(t)$$

$$\lambda = \mu$$

g and p stand for specific cell density (in O.D. units) of non-differentiated and differentiated cells, respectively. μ is growth rate per hour of the culture. $U(t)$ is light signal as a function of time and can take values 0 or 1. r_{diff} is the differentiation rate under continuous light. n is the total cell density (in O.D. units) of the culture and is kept constant. λ is the dilution rate. At constant total cell density, dilution rate, λ equals growth rate, μ. No significant difference was observed in the growth rate of differentiated cells and non-differentiated cells (Supplementary Figure 8).

The ODE model was conceived with only one free parameter, namely the differentiation rate r_{diff}. Initial conditions and the growth rate were fixed from the data (Figure 2d). Experiments were performed in the bioreactor platform, which showed modest reactor-to-reactor variability.

Data fitting was implemented in Python using SciPy library. Bounds were imposed on the parameter search (10^{-10} -10) to preserve physiological relevance. Parameter searches were conducted locally but starting from different initial parameter values spanning four orders of magnitude using an approach similar to Branch, Coleman and Li[10]. In all cases, parameter searches converged to the same parameter value. Our results suggest that this simplistic model can be used to predict steady state differentiation levels given an input light signal. The model could well predict the differentiation dynamics if we assume a delay of 60 minutes to account for observation delays that were not explicitly included in the model. (Figure 2g).

Supplementary Note 4

Construction, characterization and modelling of GAuDi strain

This Supplementary Note 1s dedicated to the construction and characterization of the differentiation system coupled to a growth arrest module such that cells growth arrest upon differentiation.

In order to arrest the cells upon differentiation, we hijacked the mating factor pathway by overexpressing a variant of the downstream effector FAR1. Concretely, prior to recombination, cells express mNeonGreen. After recombination, cells start expressing ATAF1, an orthogonal transcription factor, that drives expression of FAR1M[11] (growth arrest) and mScarlet-I (classification) with 4x binding site promoter[12]. A positive feedback loop was introduced to increase the levels of ATAF1 and consequently, FAR1M. Concretely, we added a transcriptional unit with pATAF1 driving ATAF1 TF. The strain will be referred to as Gaudi in the following (Supplementary Figure 12).
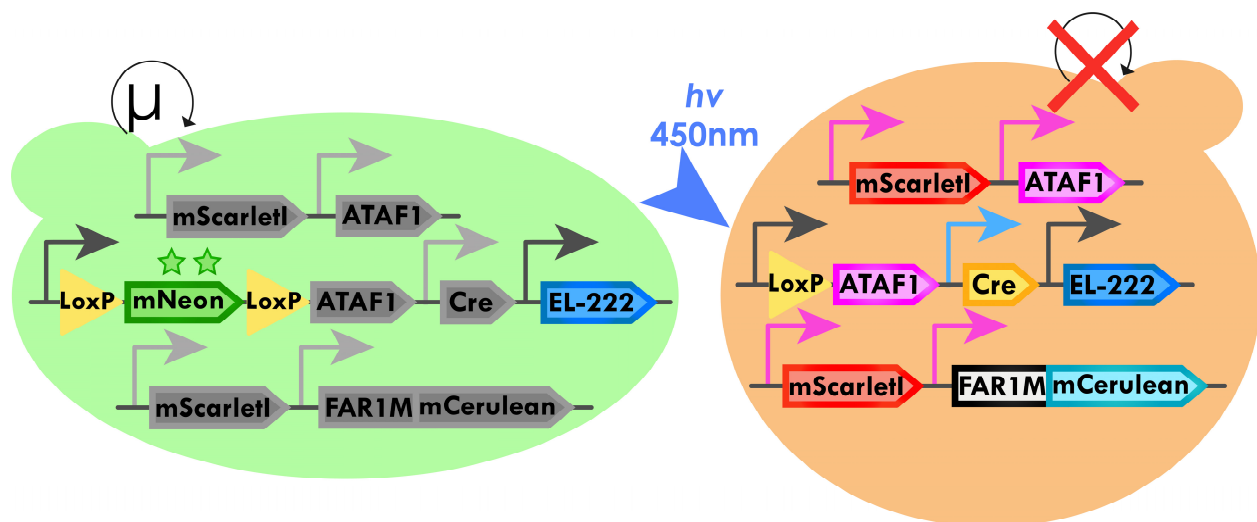

Supplementary Figure 12. Circuit for GauDi system. We added a positive feedback loop with ATAF1 TF driving its own expression to ameliorate the expression levels of FAR1M.

We conducted experiments in turbidostat and under the microscope in a microfluidic chamber and found that this topology enabled growth arrest upon differentiation. Under the microscope, cells did not undergo any cell divisions after light application (Supplementary Movie 2).

We grew cells in the turbidostat with continuous light and estimated the growth rate by measuring the OD (Supplementary Figure 12). We noticed that growth rate came to a halt roughly 4 hours and the arrest lasted for 15-20h post light induction and then the cultures escaped. Along with the OD measurements after light induction, we also analysed the population composition and discovered a variable, but consistently present, population of cells that contained neither green nor red. Presumably, these cells are dead cells.

We suspected that the escapers were mutants and extracted the genomes of 5 escapers and sent them for targeted genome sequencing. Results suggested that loss of the integrative cassette was the most common mutation (3/5) that led to escape. We hypothesized that if this were the case, mutants that lose the entire integrative cassette, including the auxotrophic marker, would be selectively disadvantaged. Therefore, we cultured cells under continuous stimulation but this time in selective media. We observed that the cells could be arrested for 50-100% longer, suggesting

that other mutations could also lead to escape. Curiously, the growth rate of differentiated cells was higher in selective media (Supplementary Figure 13).
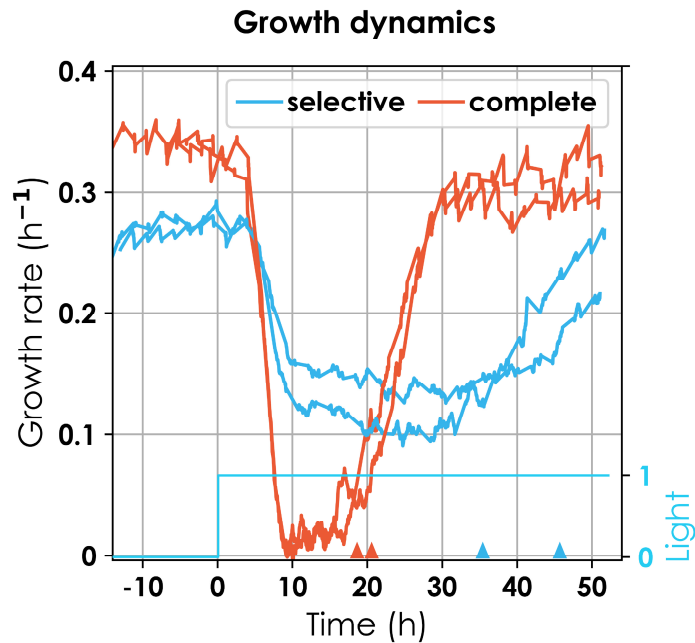
Supplementary Figure 13. Growth rate of GAuDi02 under continuous light in non-selective (red) and selective (blue) media in duplicates. The cyan line signifies light induction. Manually placed triangles indicate the time when the culture as a whole escaped the growth arrest.

Next, we characterized the differentiation dynamics of the strain in response to light. We applied different light pulses at different intervals to cultures. We observed a fast decrease in the differentiated fraction once light was removed suggesting that the growth arrest was strong and that this system could be used for dynamic control of microbial consortium in a self-contained configuration (Supplementary Figure 14).
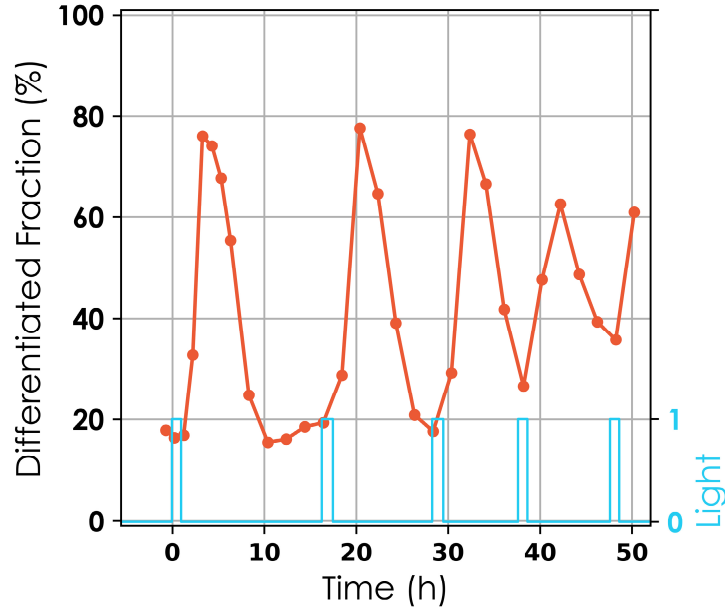
Supplementary Figure 14. Differentiation dynamics (red-dotted line) for periodic light pulses (cyan line). We observed that light pulses led to increases in differentiated fractions that subsided rapidly after removal of light (a, b).

Based on characterization data, we introduced two new state variables in the ODE model, escapers and dead cells. The rate of escape and the death rate could not be directly fixed from the data and are treated as unknown model parameters. Furthermore, due to the leakiness of the ATAF1 feedback loop, we observed a persistent fraction of cells that appeared to be initially differentiated given their levels of mScarletI fluorescence. Since empirically it was not possible to separate the differentiated cells from those with an overactive feedback loop, we introduced a basal differentiation rate present in the absence of light (and its presence). The growth rates of differentiated, non-differentiated and escapers were fixed from the data.

$$\dot{g}(t) = \mu_g g(t) - \left(r^{\circ}_{diff} + \mathbf{U}(t)r_{diff}\right)g(t) - \lambda g(t)$$

$$\dot{p}(t) = \mu_{diff} p(t) + \left(r^{\circ}_{diff} + \mathbf{U}(t)r_{diff}\right)g(t) - r_{esc}p(t) - r_{dead}p(t) - \lambda p(t)$$

$$\dot{e}(t) = \mu_{esc} e(t) + r_{esc}p(t) - \lambda e(t)$$

$$\dot{d}(t) = r_{dead}p(t) - \lambda d(t)$$

$$n = g(t) + p(t) + e(t) + d(t)$$

$$\lambda = \frac{g(t)}{n}\mu_g + \frac{p(t)}{n}\mu_{diff} + \frac{e(t)}{n}\mu_{esc}$$

$g, p, e,$ and d are specific cell density (in OD units) of non-differentiated, differentiated, escaper, and dead cells, respectively. μ_g, μ_{diff}, and μ_{esc} are growth rates per hour of non-differentiated, differentiated, and escaper cells, respectively. r_{diff}° is basal differentiation rate and r_{diff} is the differentiation rate per hour in presence of light. $U(t)$ is the light signal as a function of time and can take values 0 or 1. r_{esc} is the rate of escape per hour from growth arrest in differentiated cells per hour. r_{dead} is the rate of death per hour in growth arrested differentiated cells per hour. n is the total cell density (in O.D. units) of the culture. λ is the culture dilution rate and, at constant cell density, equals the weighted sum of growth rates of individual species.

The model was fitted to dynamical data (differentiated & dead fraction) with non-trivial light signals (Figure 4b). Concretely, the differentiation fraction and the dead fraction were computed from time-series flow cytometry data to fit the model. However, there were significant delays between differentiation events and the time by which corresponding cells could be classified as differentiated based on their fluorescence. Similarly, differentiated cells died only after being arrested for some time and the growth arrest required two cell generations to fully manifest. To account for these delays, model predictions were shifted in time (2h for differentiated and 6h for dead cells, and 5h for the growth rate). Bounds were imposed on the parameter values (10^{-10} -10) to preserve physiological relevance. The parameter search was conducted locally but starting from different initial parameter values spanning four orders of magnitude using an approach similar to Branch, Coleman and Li[10]. In all cases, the search converged to the same parameter values.

Supplementary Table 3: Parameter estimates (/h) for ODE model for GAuDi02.

Parameter	μ_g	μ_{diff}	μ_{esc}	r°_{diff}	r_{diff}	r_{esc}	r_{dead}
Estimated Value	3.9e-01	4.0e-02	3.9e-01	6.98e-02	1.16e+00	9.43e-08	7.12e-02

Parameters in green in Supplementary Table 3 were fixed directly from the data while those in yellow were estimated. These parameters led to good agreement between model outputs and observed data (Figure 4b). We further validated the model by predicting the outcome of four other experiments with dynamic non-trivial light inputs. Model predictions were in good agreement with observed data (Supplementary Figure 15). However, differentiation dynamics and growth arrest could not be predicted well under continuous light.
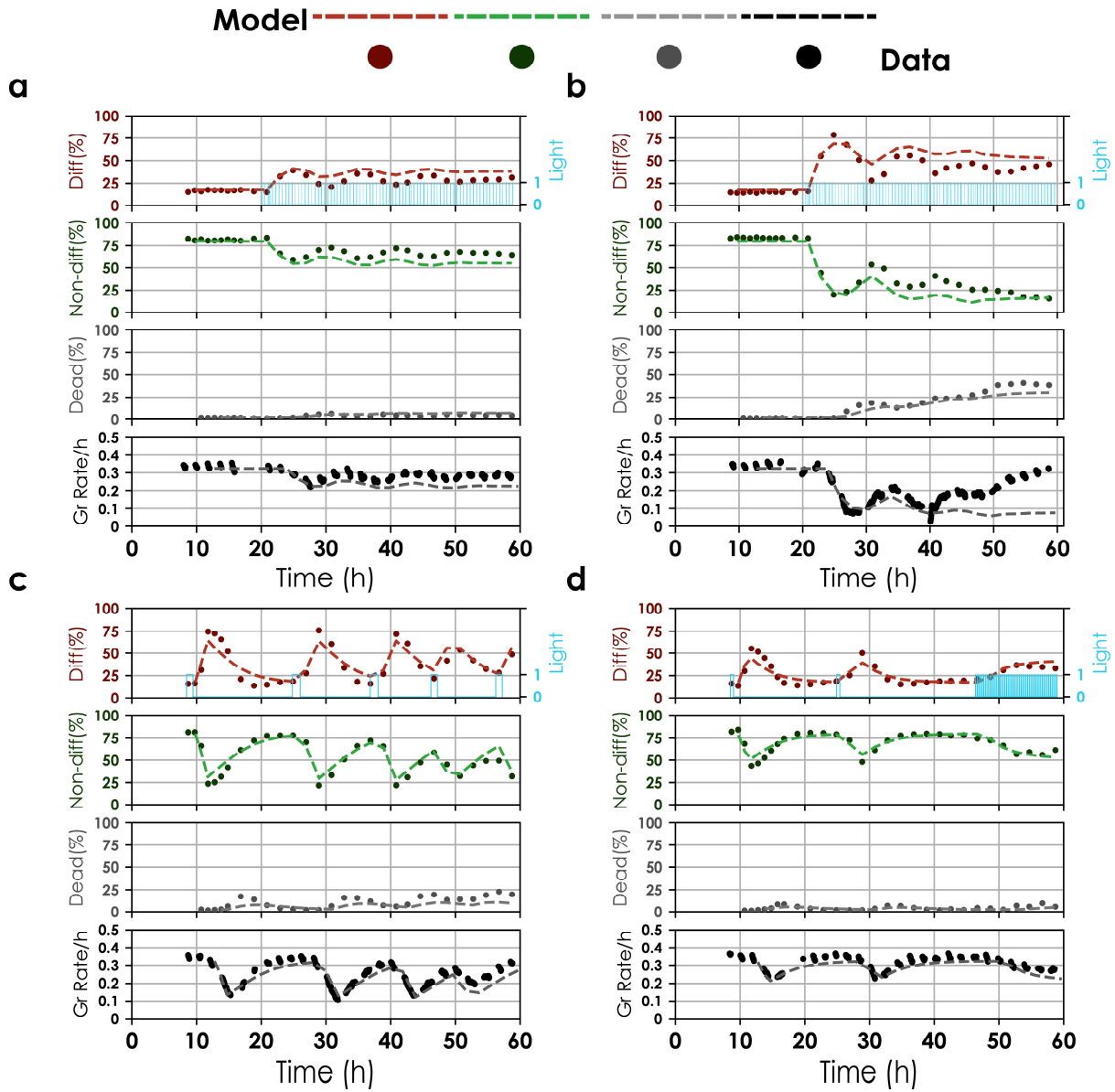
*Supplementary Figure 15. GAuDi model predictions and data for 4 light profiles. The fitted model was validated by using it to predict dynamic responses to light signals that were not used for the fit (blue lines). Circles represent differentiation fractions from a single experiment and dotted lines represent model predictions. The red line represents differentiated fraction, green is the non-differentiated fraction, grey is the dead fraction and black is the culture growth rate. Model predictions were shifted in time to account for observation delays. Initial conditions were fixed from the data. We note that in **c**, light pulses were applied with a decreasing interpulse period and that the response became dampened as the interpulse period decreased. Data suggests the cause to be accumulation of dead cells. Model predictions are in agreement with this hypothesis.*
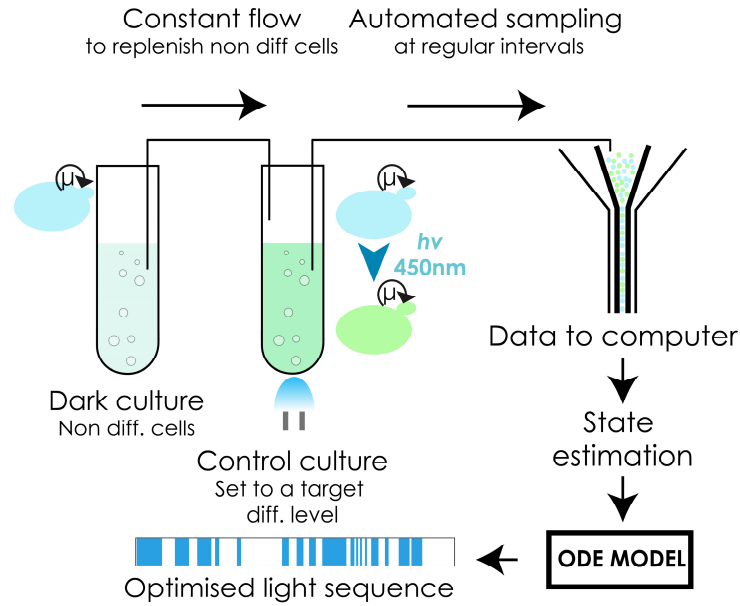
Supplementary Note 5

MPC Experiments

General strategy

1. Flow cytometry data is processed online to yield state estimates.
2. Some variables, like the escaper fraction, are non-measurable and therefore estimated from the model.
3. Conversion of this state estimate into a real-time state estimate.
   a. Because of delay not captured in the model, state estimates are taken to correspond to;

   T_0 = sampling time – (sampling delay + observation delay)
      i. Sampling delay was the time between sample was taken and the data acquisition finished.
      ii. Observation delay was the delay between change of genetic state of a cell and by the time enough fluorescence was produced to classify it as differentiated cell.
   b. Forward simulation of the model to go from T_0 to T_{now} (given the actual light sequence applied during that time window)
4. Optimization of the next sequence of 10 duty cycles (horizon = 5h) so as to minimize the predicted error between target and model prediction (gradient descent, scipy.optimize)

Dynamic control and ODE model adjustments

Supplementary Figure 16. Experimental setup.

Both reactors were maintained at constant ODs but different from each other. We wondered how to set the ODs of the two reactors so as to ensure that up to 90% of cells could be differentiated in the control reactor. Since there was a constant flux of non-differentiated cells, a flux term had to be added in the ODE model. We also observed that growth rates of cultures at different ODs were not equal. We assume, however, that cells that enter the control reactor from the reservoir reactor instantaneously change their growth rate.

$$\frac{dg}{dt} = \mu_{target}g - \boldsymbol{U}(t)\,r_{diff}g + Q - \lambda g$$

$$\frac{dp}{dt} = \mu_{target}p + \boldsymbol{U}(t)\,r_{diff}g - \lambda p$$

$$Q = \mu_{reservoir}n_{reservoir}$$

$$n_{target} = g + p$$

g and p stand for the specific cell density (in O.D. units) of non-differentiated and differentiated cells, respectively. $\mu_{control}$ and $\mu_{reservoir}$ are growth rates in the control and reservoir reactors, respectively. $U(t)$ is the light signal as a function of time and can take values 0 or 1. n_{target} and $n_{reservoir}$ are O.D. (total cell densities) at which control and reservoir reactors are maintained, respectively. r_{diff} is the differentiation rate under continuous light. Q is the flux of non-differentiated cells from reservoir to control reactor. λ is the dilution rate of the control reactor. Since we maintain the control reactor at constant OD, the number of cells in the control reactor stays constant at steady state.

$$\frac{dn_{target}}{dt} = 0$$

$$\frac{d(g + p)}{dt} = 0$$

$$\frac{dg}{dt} + \frac{dp}{dt} = \mu_{target}(g + p) + Q - \lambda(g + p)$$

$$0 = \mu_{target}(g + p) + Q - \lambda(g + p)$$

$$\lambda = \mu_{target} + \mu_{reservoir}\frac{n_{reservoir}}{g + p}$$

$$\lambda = \mu_{target} + \mu_{reservoir}\frac{n_{reservoir}}{n_{target}}$$

$$\lambda = \mu_{target} + \alpha\mu_{reservoir}$$

α is the ratio of reservoir OD to control reactor OD.

Let r be equal to the fraction of differentiated cells in the control reactor.

$$r = \frac{p}{g + p}$$

$$\frac{dr}{dt} = \frac{\frac{dp}{dt}}{g + p}$$

$$\frac{dr}{dt} = \frac{\mu_{target}p + U(t)r_{diff}g - (\mu_{target} + \alpha\mu_{reservoir})p}{g + p}$$

$$\frac{dr}{dt} = U(t)r_{diff}(1 - r) - \alpha\mu_{reservoir}r$$

$$\frac{dr}{dt} = -(U(t)r_{diff} + \alpha\mu_{reservoir})r + U(t)r_{diff}$$

This is a first order non-homogeneous linear differential equation in standard form,

$$\frac{dy}{dt} = -p(t)y + f(t)$$

For maximum differentiation fraction achievable ($U(t) = 1$), at steady state,

$$\frac{dr}{dt} = 0$$

$$r_{diff} = (r_{diff} + \alpha\mu_{reservoir})r_{max}$$

$$\alpha = \frac{r_{diff}}{\mu_{reservoir}}\left(\frac{1}{r_{max}} - 1\right)$$

Plugging in values of $r_{diff} = 0.86 \text{ h}^{-1}$, $r_{max} = 0.9$ and $\mu_{reservoir} = 0.41\text{h}^{-1}$, we get $\alpha = 0.23$. Therefore, to reach a target differentiation level of 90 % in the control reactor, the OD of the reservoir had to be ~5 times lower. Based on this calculation, the OD of the reservoir reactor was set to 0.1 and that of control reactor to 0.6. Growth rates were determined from the data. This model was then plugged into the MPC framework described previously in Bertaux et al. 2020[1]. Briefly, the MPC algorithm utilizes the fitted model to search for the optimal sequence of light signals that would lead to desired target level of differentiated cells. The differential equation obtained in (2) could be solved analytically for constant light signal and was solved piecewise in the model. Receding time horizon for model predictions for the experiments shown here was 5 hours (10 cycles of period 30 minutes each). Automated flow cytometry measurements were performed every hour. The light sequence was automatically re-optimized and updated at every

timepoint based on the newly estimated state of the system. There existed a significant delay (1.5-2h) between the measurement and the state of the system at the time of measurements. To bridge this gap, state estimation was used. We assumed that the measured data corresponds to a snapshot of the population composition 1.5h in the past and to estimate the current state we solved equation (2) with the light sequence that was applied over the past 1.5h. This was critical to avoid oscillatory behaviour of the system. Furthermore, we noticed that the differentiation rate parameter estimated from experiments led to a steady state error. Therefore, the value of this parameter was reduced to 0.4 /h in the MPC model to ensure robust control.
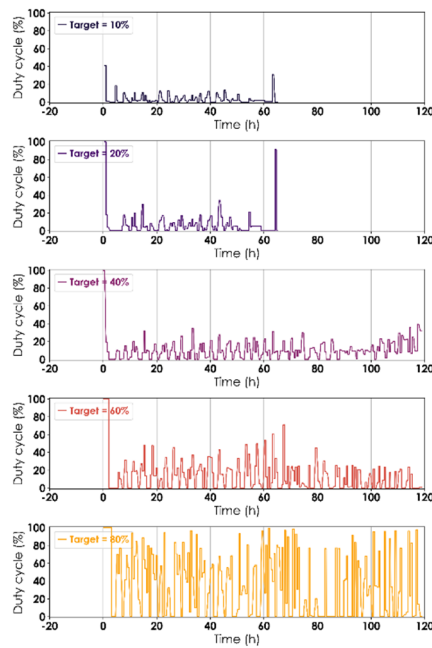
Supplementary Figure 17. Light profiles for experiments in Figure 3b. Duty cycles were estimated for a period of 30 minutes. Induction started at t=0.

Single vessel control
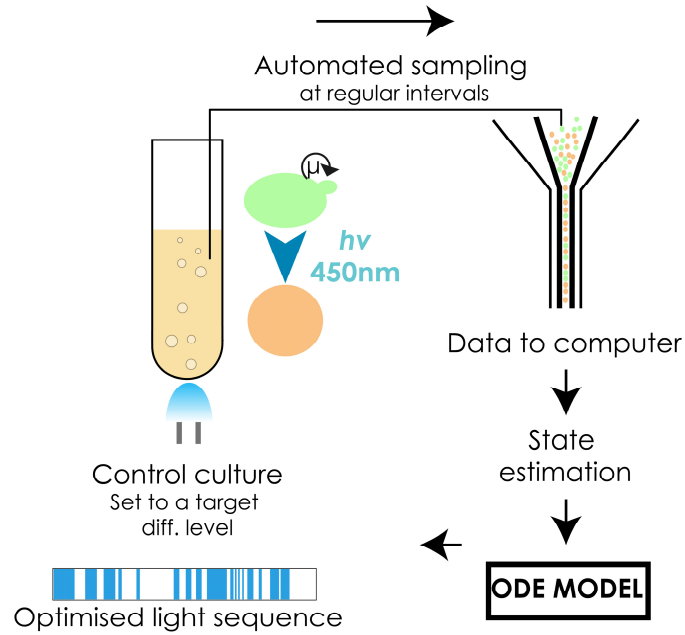
Supplementary Figure 18. Experimental setup.

The model for the GAuDi strain was used for the MPC. No adaptation to the model was necessary.

We could only sample the culture every two hours owing to volume limitations and consequently

the delay was increased to 2.5 hours.

Supplementary Figure 19. Light profiles for experiments in Figure 4e. Duty cycles were estimated
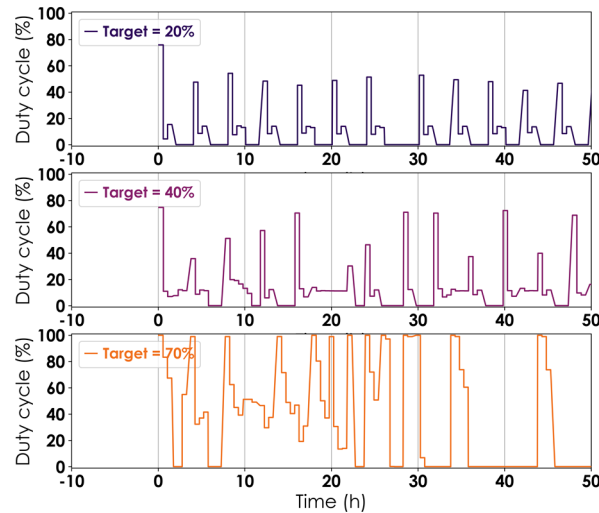*for a period of 30 minutes. Induction started at t=0.*

Supplementary Note 6

Spatial control

Cells harbouring the differentiation system were cultured in the dark to late exponential phase in 2% glucose synthetic complete LoFlo media at 30°C in 50mL falcon tubes. Cells were then loaded in a 6 channel μIbidi slide (Catalogue # 80606) such that they formed a monolayer. The μIbidi slide was induced under the microscope (10x magnification) within a user-defined pattern with the help of a digital mirror device (Andor Mosaic3). The pattern was a binary bitmap with black representing no light, and white representing induction.

A light of 460 nm was used for induction with the CFP band pass filter (Supplementary Table 2). Based on the results of the characterization experiment, we initially chose to induce the cells with 1s pulses every 6 mins at an intensity of 10%. The ibidi slide could not support a monolayer of yeast cells for more than 2.5-3h. Eventually bubbles would emerge disturbing the monolayer. It took about 30 minutes for all the floating cells to settle. We found that cells required at least 1h to generate enough mNeonGreen to be readily detectable (Supplementary Figure 6d). Due to these constraints, 1h was the maximum induction that led to reproducible patterns. Increasing beyond this induction period did not result in more detectable differentiation in the timeframe of the experiment. To improve the differentiation efficiency, we finally decided to double the frequency of pulses. We tried to ameliorate the growing conditions by culturing cells in a microfluidic plate but the media flow disallowed any meaningful quantification of efficiency of pattern formation at low cell density. This was because buds from inside the pattern would be transported with the flow, to regions outside the pattern. When we tried to do the experiment after cells had formed a monolayer, flow was obstructed leading to growth rate decrease and cell death due to lack of media. We note that the cells could not be imaged during the induction period as both mCerulean

and mNeonGreen required illumination that would activate the EL222 and lead to undesirable differentiation.
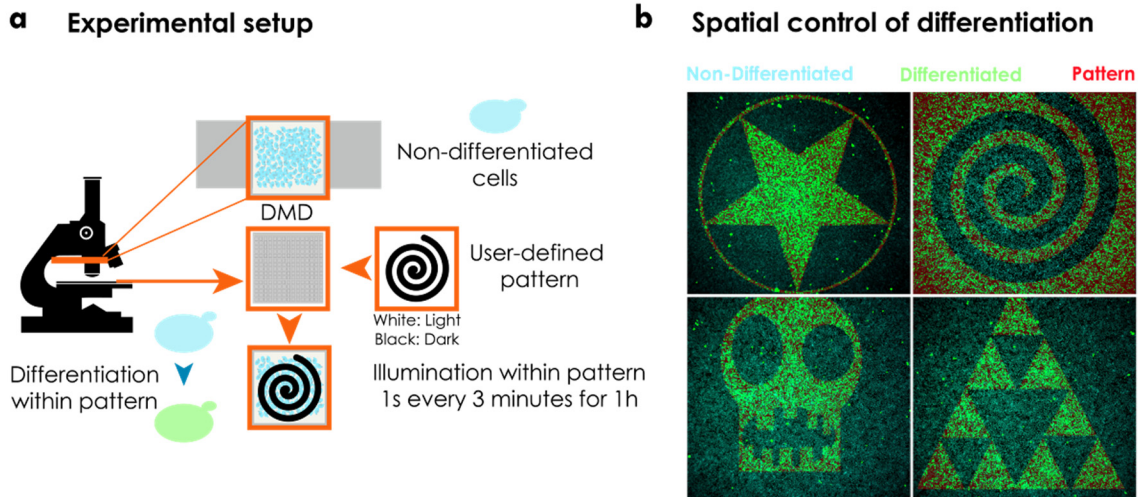

Supplementary Figure 20. Precise spatial control of heterogeneous microbial consortia originating from a single strain. a. Hardware setup for realizing pattern formation under the microscope. Cells carrying the differentiation system were cultured in the dark until late exponential phase (OD ~0.8) and loaded in a μIbidi slide. Cells were allowed to settle down and form a monolayer. Light was shone in a user-defined pattern over the monolayer using a digital mirror device (DMD). Light was shone as 1s pulses every 3 minutes for 1h (20 pulses). Following light delivery, cells were kept in darkness for 1 hour before imaging them. b. Images of imprinted patterns in the population. mCerulean fluorescence was ascribed a cyan colour, whereas mNeonGreen fluorescence was ascribed a green colour (false colour in microscopy images). The pattern, in red, is overlaid on top of the merge for cyan and green channels.

Supplementary Note 7

Expanding the differentiation system to give rise to multi-species consortia with differentiation programs

In order to explore the possibility of engineering differentiation programs, we cloned two recombination cassettes of unequal to-be-excised region but otherwise identical, in different strains. These strains were cultured continuously in the turbidostat and induced with full light. We found that changing the length of the to-be-excised region plays a crucial role in determining the rate of differentiation (Supplementary Figure 21). We observed that differentiation rate is inversely proportional to the length of the to-be-excised region.
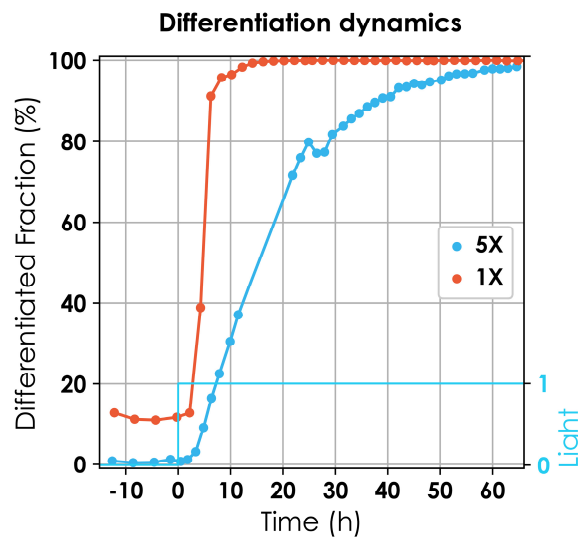
*Supplementary Figure 21. Differentiation fraction dynamics for recombination cassettes of different to-be-excised regions (**1X**: red and **5X**: blue) in continuous light. Red and blue dotted lines show differentiation fraction over time. Legend indicates the length of the to-be-excised region compared to the one used for Figure 1. For the experiment shown here, a significant part of the population for **1X** was recombined prior to light induction.*

Exploiting this information, we implemented differentiation programs by introducing two recombination cassettes of to-be-excised-regions of different lengths. These were named according to their differentiation behaviour with respect to the two recombination cassettes. If the recombination cassettes are of equal length (Supplementary Figure 22a), the recombination is

asynchronous meaning either recombination cassette could recombine independently of each other (Supplementary Figure 22a). We can achieve sequential differentiation if the to-be-excised regions are unequal (Supplementary Figure 22b) meaning that the recombination occurs first at the shorter site followed by recombination at the longer site. We did not observe recombination of the longer site exclusively (Supplementary Figure 22b).
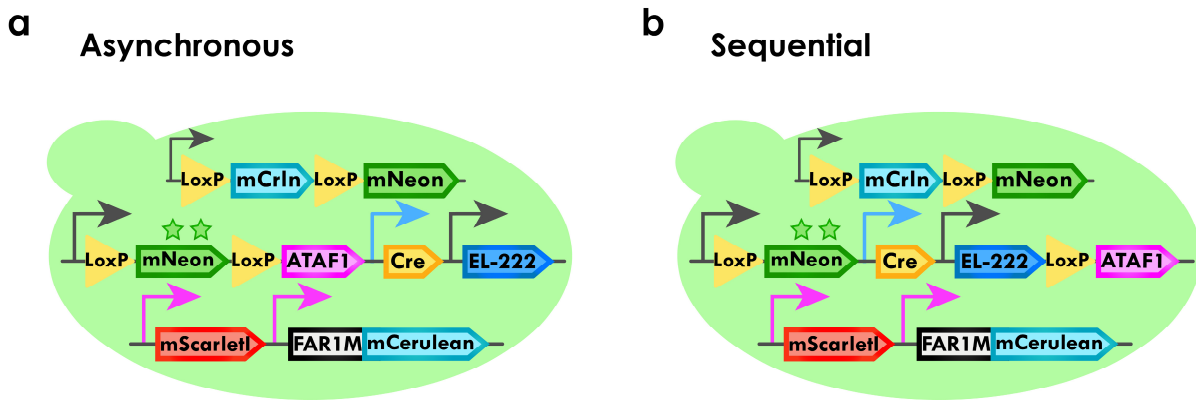

*Supplementary Figure 22. Circuits for (**a**) asynchronous differentiation program and (**b**) sequential differentiation program which result in 4 and 3 species microbial consortia, respectively.*

All four species could be observed for the asynchronous differentiation program while only 3 could be observed with the sequential differentiation program (Supplementary Figure 23). Note that EL222 and Cre are included within the flowed region and, therefore, following recombination, are removed from the cell. Our intention in removing Cre and EL222 upon excision of the longer site was to lock cells into the singly recombined state if they recombine the longer site before the short one. However, the fraction of cells that recombined the longer site first was too low for this feature of the design to matter much.

We also note that the expression of FAR1M does have a negative effect on the growth rate of $\overline{CN}$ and $C\overline{N}$, however, the difference was not large enough to be observable at the timescale of the

induction profiles used in the experiments shown in the study (Figure 5b & d). The authors believe

that it will be possible to replenish the CN and $\bar{C}N$ species over longer timescales in the dark.
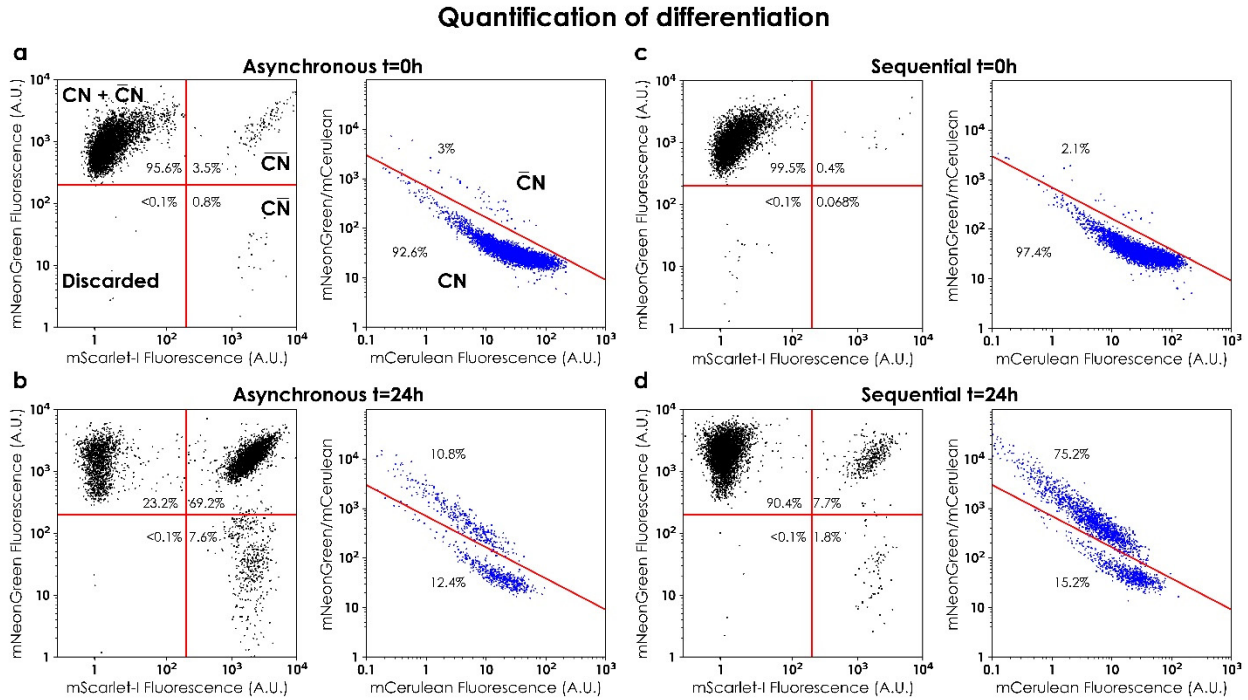
Supplementary Figure 23. Quantification of species prevalence. Cells were first separated by mScarlet-I fluorescence. Cells possessing red fluorescence were either $\overline{CN}$ or $C\bar{N}$ and those that did not were CN or $\bar{C}N$. To distinguish between $\overline{CN}$ and $C\bar{N}$ mNeonGreen fluorescence was enough. However, to distinguish between CN and $\bar{C}N$, both mNeonGreen and mCerulean fluorescence had to be analysed to obtain two well-separated populations (high mNeonGreen to mCerulean ratio corresponds to CN and the lower to $\bar{C}N$). Cells not possessing appreciable levels of either mNeonGreen or mScarletI were discarded from analysis (<1%). The figure shows the four subpopulations before induction, and after light stimulation, initiated at t= 4h and made of 3 light pulses of 30 minutes 6h apart, for both differentiation programs (Figure 5b & 5d). A.U. stands for arbitrary units.

Supplementary Note 8

Note on handling of light sensitive strains

To obtain pure non-differentiated fraction, all cell manipulation was done in red light including transformation and the creation of stocks. Stocks were covered with aluminium foil during storage. Revivals were performed in red light and yeast plates were encased in aluminium foil during growth in the incubator. Overnight cultures were grown in the dark using a falcon holder covered on all sides. Overnight cultures were measured using cytometry prior to the start of preculture to ensure maximum non-differentiated cells. Typically, we observed 0.03%-1% differentiation at this point (for GAuDi we always observed 10%-20% background differentiation but as noted in the Supplementary Note 4, we believe these cells were present due to leaky expression of ATAF1 TF and not background recombination). Some picked colonies were completely recombined and not used for experiments. We observed that streaked plates more than a week old led to higher levels of background differentiation in individual colonies. We note that the colonies were extremely sensitive to extraneous light and even opening them in red light led to significant increase in differentiated fraction over the next 96h (5%-10%). Such susceptibility to extraneous light was not observed for liquid cultures. Consequently, only freshly streaked plates were used for the experiments.

.

Supplementary Tables

Supplementary Table 4. List of primers used in the study.

Primer Name	Sequence	Target	Used for	Source
p95	gttccggctgtcttgcttag	Con E (YTK connector) R	Colony PCR single TU Level 1 plasmids	YTK
p128	ctgaactggccgataattgc	Con S (YTK connector) F	Colony PCR	YTK
p133	ggttcgtaacatctctgtaactgc	ConE (YTK connector) scar R	Colony PCR single TU Level 2 plasmids	YTK
p134	gagataggtgcctcactgattaag	AmpR cassette R	Colony PCR multiple TU Level 1 plasmids	YTK
CA6	GCATCGTCTCATCGGTCTCATTCTgccaccatggtgagcaagg	mCerulean NT F	New part Generation	This study
CA7	ATGCCGTCTCAGGTCTCAGGATcttgtacagctcgtccatgc	mCerulean CT R	New part Generation	This study
CA21	GCATCGTCTCATCGGTCTCATATGatgtctaatttgttgactg	Cre-Recombinase NT F	New part Generation	This study
CA22	ATGCCGTCTCAGGTCTCAGGATtctagatgcatgctcg	Cre-Recombinase CT R	New part Generation	This study
CA54	GCATCGTCTCATCGGTCTCATGGCatggtgagcaagggcgagg	mNeon NT F	New part Generation	This study
CA55	ATGCCGTCTCAGGTCTCACAGCcatccagccgttcaggg	tTDH1 R	New part Generation	This study
CA92	GCATCGTCTCATCGGTCTCATATGaagacaccaacaagagtttcg	Far1 NT F	New part Generation	This study
CA93	ATGCCGTCTCAGGTCTCAAGAAcctccaccgaggttgggaacttccagg	Far1 CT R	New part Generation	This study
CA131	GCATCGTCTCATCGGTCTCATGGCTATGGATGgataaagcggaattaattcc	ATAF1 NT F	New part Generation	This study
CA132	ATGCCGTCTCAGGTCTCACAGCagccgttcagggtaatatatttaacc	tTDH1 R	New part Generation	This study
	YTK stands for yeast tool kit created by Lee et al. 2015			
	Legend			
	hybridizing sequence			
	PART SPECIFIC OVERHANG			
	linker			

Supplementary Table 5. List of strains generated in the study with their genotype.

Title	Name	Description Complete	Genotype	Source	Appearance in the text
IB20019	WT/AF	BY4741	MATa his3Δ1 leu2Δ0 met15Δ0 ura3Δ0	EuroScarf	Parent strain
IB20199	EL222 Reporter	pIB0120 integrated in URA locus and pIB0515 integrated in the LEU locus	MATa his3Δ1 leu2Δ::pEL222_5x_Gal-mscarletI-tENO1-LEU2 met15Δ0 ura3Δ:: pTDH3-SV40 NLS - VP16AD - EL222-tSSA1-URA3	This study	Figure S3.2, Supplementary Note 3
IB20297	Dual Reporter	pIB0120 integrated in URA locus ; pIB0178 integrated in LEU Locus and pIB0684 integrated in HO locus	MATa his3Δ:: pEL222_5x_Gal-Cre-tENO2-HIS3 leu2Δ::pTDH3-LoxP-mCerulean-tENO1-LoxP-mNeonGreen-tTDH1-LEU2 met15Δ0 ura3Δ:: pTDH3-SV40 NLS - VP16AD - EL222-tSSA1-URA3	This study	Figure1,2,3, Supplementary Notes 1, 2, 3, 6
IB20268	GAuDi	pIB0593 integrated in LEU locus ; pIB0602 integrated in URA locus and pIB0610 integrated in HO locus	MATa his3Δ:: pATAF1_4x-mScarlet-tDIT1-pATAF1_4x-ATAF1-tRPL41b-HIS3 leu2Δ::pATAF1_4x-mScarlet-tDIT1-pATAF1_4x-Far1M_mCerulean-tDIT1-LEU2 met15Δ0 ura3Δ:: pTDH3_LoxP-mNeonGreen-tENO1_LoxP-ATAF1-tTDH1-pEL222_5x_Gal-Cre-tENO2-pTDH3-EL222-tSSA1-URA3	This study	Figure 4, Supplementary Notes 1, 4
IB20339	Asynchronous	pIB0593 integrated in LEU locus ; pIB0179 integrated in URA locus and pIB0603 integrated in HO locus	MATa his3Δ:: pTDH3_LoxP-mNeonGreen-tENO1_LoxP-ATAF1-tTDH1-pEL222_5x_Gal-Cre-tENO2-pTDH3-EL222-tSSA1-HIS3 leu2Δ::pATAF1_4x-mScarlet-tDIT1-pATAF1_4x-Far1M_mCerulean-tDIT1-LEU2 met15Δ0 ura3Δ:: pTDH3-LoxP-mCerulean-tENO1-LoxP-mNeonGreen-tTDH1-URA3	This study	Figure 5, Supplementary Note 7
IB20340	Sequential	pIB0593 integrated in LEU locus ; pIB0601 integrated in HO locus and pIB0179 integrated in URA locus	MATa his3Δ:: pTDH3_LoxP-mNeonGreen-tTDH1-pEL222_5x_Gal-Cre-tENO2-pTDH3-EL222-tENO1_LoxP-ATAF1-tTDH1-HIS3 leu2Δ::pATAF1_4x-mScarlet-tDIT1-pATAF1_4x-Far1M_mCerulean-tDIT1-LEU2 met15Δ0 ura3Δ:: pTDH3-LoxP-mCerulean-tENO1-LoxP-mNeonGreen-tTDH1-URA3	This study	Figure 5, Supplementary Note 7

References

1.   Bertaux, F. *et al.* Enhancing multi-bioreactor platforms for automated measurements and reactive experiment control. *bioRxiv* 2020.12.27.424467 (2020).

2.   Fox, Z. R. *et al.* MicroMator: Open and Flexible Software for Reactive Microscopy. *bioRxiv* 2021.03.12.435206 (2021).

3.   Benzinger, D. & Khammash, M. Pulsatile inputs achieve tunable attenuation of gene expression variability and graded multi-gene regulation. *Nat. Commun.* **9,** 3521 (2018).

4.   Hochrein, L., Mitchell, L. A., Schulz, K., Messerschmidt, K. & Mueller-Roeber, B. L-SCRaMbLE as a tool for light-controlled Cre-mediated recombination in yeast. *Nat. Commun.* **9,** 1931 (2018).

5.   An-adirekkun, J. *et al.* A yeast optogenetic toolkit (yOTK) for gene expression control in Saccharomyces cerevisiae. *Biotechnol. Bioeng.* **117,** 886–893 (2020).

6.   Xu, X. *et al.* A single-component optogenetic system allows stringent switch of gene expression in yeast cells. *ACS Synth. Biol.* **7,** 2045–2053 (2018).

7.   Taslimi, A. *et al.* Optimized second-generation CRY2-CIB dimerizers and photoactivatable Cre recombinase. *Nat. Chem. Biol.* **12,** 425–430 (2016).

8.   Kawano, F., Okazaki, R., Yazawa, M. & Sato, M. A photoactivatable Cre-loxP recombination system for optogenetic genome engineering. *Nat. Chem. Biol.* **12,** 1059–1064 (2016).

9.   Duplus-Bottin, H. *et al.* A single-chain and fast-responding light-inducible Cre recombinase as a novel optogenetic switch. *Elife* **10,** e61268 (2021).

10.   Branch, M. A., Coleman, T. F. & Li, Y. A subspace, interior, and conjugate gradient method for large-scale bound-constrained minimization problems. *SIAM J. Sci. Comput.* **21,** 1–23 (1999).

11.   Henchoz, S. *et al.* Phosphorylation-and ubiquitin-dependent degradation of the cyclin-dependent kinase inhibitor Far1p in budding yeast. *Genes Dev.* **11,** 3046–3060 (1997).

12.   Naseri, G. *et al.* Plant-derived transcription factors for orthologous regulation of gene expression in the yeast Saccharomyces cerevisiae. *ACS Synth. Biol.* **6,** 1742–1756 (2017).