Supporting Online Material

1 Materials and Methods

1.1 Strains and Plasmids

All plasmids used in this study were derived from a set of yeast single integration vectors which were gifts from N. Helman and W. Lim, and contain markers and targeting sequences for the TRP1, LEU2, HIS3 loci. The pGAL1 reporter plasmid was constructed in a HIS3 vector in two steps: the GAL1 promoter was amplified from yeast genomic DNA, and then cloned into PspOMI and xho1 restriction sites, followed by Venus YFP cloned between Xho1 and BamH1 with an ADH1 terminator. The PIF3 construct was assembled in a LEU2 vector. The construct contained an ADH1 promoter cloned between PspOMI and Xho1, a SV40-nuclear localization signal (NLS), and GAL4 activation domain (fused to PIF3) between Xho1 and Not1. The PHYB construct was assembled in the TRP1 vector with an ADH1 promoter. The PhyB N-terminal sequence was fused to the GAL4 DNA binding domain (residues 1-93, which contain an endogenous NLS sequence) and inserted between PspOMI and Not1. The PhyB and PIF components were amplified from PIFGAD and PhyBNT[7]. The strain used in all experiments was derived from W303 matA, transformed with ADE2 gene. This strain was then sequentially transformed with the above plasmids using standard genetic techniques. The final strain genotype was W303 MATa leu2::prADH1-GADPIF3 (LEU2), trp1::prADH1-PHYBGDBD (TRP1), his3:: prGAL1-YFP (HIS3), ade2::ADE2 can1 ura3.

1.2 Growth Conditions and Assay

For all experiments, cells were grown in SD complete media at 30C in the dark. Cell cultures were grown in exponential phase for a minimum of 8hrs before all experiments. At the beginning of an experiment, cells were diluted to an OD of 0.05 and 120μl placed in shallow 96 well plates (COSTAR) on a shaking plate incubator (Eppendorf). Chromophore (dissolved in DMSO, 3mM) was added to saturating concentration (1:300 dilution) and cells were incubated in the dark for 20 minutes. Light was then applied as described in main text using a custom-designed LED light source. Every 30 minutes, 15μl aliquots were removed, diluted 6 fold in media and measured on the LSRII cytometer (BD). YFP was excited at 488nm, and Fluorescence was collected through a HQ530/30 bandpass filter (Chroma). After each sample, 15μl of fresh media with chromophore was added to the cells to maintain constant volume, cell density increased slightly across the experiment as the doubling time was just over 2hrs in these conditions.

1.3 Data Analysis

Flow cytometry distributions were gated by forward and side scatter to remove aberrant cells from the acquired data set. The fluorescence measurements were then normalized with respect to forward scatter and the median value was taken as a measure of the average population fluorescence. All

Nature Biotechnology: doi:10.1038/nbt.2018

data processing (including model parameter identification and real-time estimation and control) was completed using custom Matlab scripts and are available upon request.

1.4 Light Pulse System (Experimental Setup)

For the application of all light treatments, a custom-built "light pulser" system was designed and built (Supplementary Fig. 1). Red (650nm) and far-red (730nm) wavelengths were provided by two triplets of high-power (5W) LEDs (LedEngin LZ1-10R305 and LZ1-10R205, `http://www.ledengin.com/led_products.htm`), which were controlled by two high-brightness LED driving boards (Maxim Max16816). Both the "light pulser" (Supplementary Fig. 1C, D) and the electronics box (Supplementary Fig. 1A, B) were designed using a 3D CAD program (Solidworks, `http://www.solidworks.com/`) and prototyped on an Acrylonitrile Butadiene Styrene (ABS) thermoplastic template using 3D-printing technology (uPrint Plus 3D printing system, `http://www.dimensionprinting.com/`). The complete bill of materials detailing all electronics components, as well as the Matlab code used for the experiments (including parameter identification, closed loop control and system simulation) are available upon request.

1.5 Light Application Protocol

The light fluence (dose)-response curve of the PhyB/PIF system saturates for large fluence levels. The maximal achievable induction then starts to quickly decrease as the fluence increases beyond saturating levels, an effect attributed to chromophore-induced photodynamic damage to the cells[7]. We experimentally confirmed this phenomenon and observed that R and FR pulses of 1 minute duration are sufficient to drive the system to saturation both in the OFF-ON and ON-OFF directions, without the undesirable effects of excessive light exposure, for the used culture size and geometry (data not shown). Only such saturating R and FR pulses were used for all characterization and control experiments, while the cells were kept in the dark between light treatments.

1.6 Stationary Conditions and Reproducibility

Cell cultures grown in exponential phase (in SD complete media at 30C in the dark) for a minimum of 8hrs achieved a unique stationary condition. To measure the basal fluorescence at stationary conditions, fluorescence distributions from 15 cell cultures were obtained experimentally (Supplementary Fig. 2A). To obtain an average population fluorescence value at stationarity, we took the arithmetic mean of the 15 population median fluorescence measurements and obtained a basal level of fluorescence of 0.018 a.u. The reproducibility of the basal average fluorescence level at stationary conditions enabled us to consider it as a reference value; data for all subsequent experiments is reported in terms of fold change over this value. Reproducibility of the system time-response curve for cell cultures initialized at stationary conditions and perturbed by identical light treatments was also demonstrated experimentally (Supplementary Fig. 2B).

1.7 Mathematical model

The light-induced genetic circuit is modeled as a switched continuous-time linear dynamic system with 4 states and 5 effective parameters. The dynamical system is reported in both absolute values and fold change above stationary conditions; both models are mathematically equivalent. The model states represent average values (although in the actual experiments single cell measurements in the form of distributions were acquired (Supplementary Fig. 3)). In addition to the dynamical model for the states of the system we define the input (control) model and the measurement model.

1.7.1 Absolute values model

The states of the mathematical model represent average values of the transcriptional input signal s, the mRNA for YFP m, the immature YFP p, and the fluorescent (mature) YFP f. The state equations are

$$\begin{aligned}\dot{s} &= -a_s s \\ \dot{m} &= -a_m m + b_m s + r_m \\ \dot{p} &= -a_p p + b_p m - kp \\ \dot{f} &= -a_p f + kp\end{aligned}$$

where the model parameters (descriptions and values) are reported in Supplementary Table 1. More concisely the system can be written as

$$\begin{bmatrix} \dot{s} \\ \dot{m} \\ \dot{p} \\ \dot{f} \end{bmatrix} = \begin{bmatrix} -a_s & 0 & 0 & 0 \\ b_m & -a_m & 0 & 0 \\ 0 & b_p & -a_p - k & 0 \\ 0 & 0 & k & -a_p \end{bmatrix} \begin{bmatrix} s \\ m \\ p \\ f \end{bmatrix} + \begin{bmatrix} 0 \\ r_m \\ 0 \\ 0 \end{bmatrix}. \tag{1}$$

with initial conditions $s(0)$, $m(0)$, $p(0)$, and $f(0)$. We denote the unique set of stationary (equilibrium) conditions s_{ss}, m_{ss}, p_{ss}, and f_{ss} where the following algebraic equations hold

$$\begin{aligned}s_{ss} &= 0 \\ m_{ss} &= \frac{r_m}{a_m} \\ p_{ss} &= \frac{b_p}{a_p + k} m_{ss} = \frac{r_m b_p}{a_m(a_p + k)} \\ f_{ss} &= \frac{k}{a_p} p_{ss} = \frac{k r_m b_p}{a_p a_m (a_p + k)}.\end{aligned}$$

We assume that under normal conditions, $s(0) = s_{ss}$, $m(0) = m_{ss}$, $p(0) = p_{ss}$, and $f(0) = f_{ss}$ is achieved when the growth protocol outlined in the Materials and Methods is followed.

1.7.2 Fold change model

The fold change dynamic model *is defined with respect to the stationary values of the nominal states* m, p, *and* f, i.e. m_{ss}, p_{ss}, and f_{ss}. The states for the fold change model consist of the signal s, and the average absolute fold changes of mRNA $\bar{m}$, dark protein $\bar{p}$, and fluorescent protein $\bar{f}$, where $m = \bar{m} m_{ss}$, $p = \bar{p} p_{ss}$ and $f = \bar{f} f_{ss}$. Taking into account these relations and the absolute values model, we obtain the following equivalent state equations for the fold change model:

$$\begin{aligned}\dot{s} &= -a_s s \\ \dot{\bar{m}} &= -a_m(\bar{m} - 1) + \bar{b}_m s \\ \dot{\bar{p}} &= (a_p + k)(\bar{m} - \bar{p}) \\ \dot{\bar{f}} &= a_p(\bar{p} - \bar{f})\end{aligned}$$

Parameter	Description	Units	Value	Notes
a_s	degradation rate of the transcriptional signal	min^{-1}	0.0155	half-life $\sim$ 45 minutes
a_m	degradation rate of the mRNA	min^{-1}	0.0300	half-life $\sim$ 23 minutes
b_m	production rate of the mRNA due to the transcriptional signal	min^{-1}	-	
r_m	basal production rate of the mRNA	$a.u.\ min^{-1}$	-	
a_p	degradation rate of YFP (both dark and fluorescent) which in this case is primarily due to dilution	min^{-1}	0.0066	half-life $\sim$ 105 minutes
b_p	production rate of the dark YFP due to the mRNA	min^{-1}	-	
k	maturation rate of dark YFP	min^{-1}	0.0419	half-life $\sim$ 16 minutes
$\bar{b}_m$	production rate of the mRNA fold change due to the transcriptional signal	min^{-1}	0.9587	

Supplementary Table 1: Parameter description and values for mathematical model of light-induced genetic system

where the model parameters (descriptions and values) are reported in Supplementary Table 1. The new parameter $\bar{b}_m$, is defined $\bar{b}_m = b_m/m_{ss}$, where b_m is the production rate of the mRNA and m_{ss} is the absolute mRNA concentration at stationarity. More concisely the system can be written as

$$\begin{bmatrix} \dot{s} \\ \dot{\bar{m}} \\ \dot{\bar{p}} \\ \dot{\bar{f}} \end{bmatrix} = \begin{bmatrix} -a_s & 0 & 0 & 0 \\ \bar{b}_m & -a_m & 0 & 0 \\ 0 & a_p + k & -a_p - k & 0 \\ 0 & 0 & a_p & -a_p \end{bmatrix} \begin{bmatrix} s \\ \bar{m} \\ \bar{p} \\ \bar{f} \end{bmatrix} + \begin{bmatrix} 0 \\ a_m \\ 0 \\ 0 \end{bmatrix}. \tag{2}$$

with $s(0) = 0$, $\bar{m}(0) = m(0)/m_{ss}$, $\bar{p}(0) = p(0)/p_{ss}$, and $\bar{f}(0) = f(0)/f_{ss}$.

Note that in the transformed model the number of parameters is reduced from 7 to 5. Given that the basal value of the mature YFP f_{ss} has been measured, we henceforth adopt the fold change model with the acknowledgment that it is mathematically equivalent to the absolute values model.

1.7.3 Input Model

The input to the system (i.e., the light pulses) is represented by the variable u and can be applied to the system every 15 minutes. In each case, u can take 3 discrete values, resulting in a reset of the transcriptional signal $s \in [0, 1]$. In all cases, the control action at time t, $u(t) \in \mathcal{U} = \{0, 1, 2\}$ represents:

- $u(t) = 0$: Do nothing
- $u(t) = 1$: $R(t)$ (set $s(t) = 1$)
- $u(t) = 2$: $FR(t)$ (set $s(t) = 0$)

where $R(t)$ denotes apply R light at time t and $FR(t)$ denotes apply FR at time t.

1.7.4 Measurement Model

Population measurements are obtained from the flow cytometer at discrete points in time. Normalizing the (gated) fluorescence measurements with respect to forward scatter and taking the median, we obtain a measurement value $\bar{y}(t)$ at time t representing the average (normalized) fluorescence of the cell population. $\bar{y}(t)$ is the sum of YFP fluorescence (due to the basal gene activity of cells grown in darkness) and cellular background autofluorescence.

Given that cellular background autofluorescence is difficult to quantify precisely in general, in this work we make the assumption that $f_{ss} = \bar{y}(0)$ (i.e. ignore the effect of autofluorescence), with the knowledge that it does not affect the performance of our estimation and control algorithms, given that our control objective is to regulate the *observed* cell fluorescence to a given level above the basal. Given our assumption on autofluorescence, the measurement vector for the fold change model (to be used for real-time estimation and control) consists of $\overline{f}$. That is, for each time t and a known basal value $f_{ss} = 0.018$ a.u. (acquired experimentally, cf. Section 1.6), the measured state for the fold change model is

$$y(t) = \frac{f(t)}{f_{ss}} = \bar{f}(t).$$

In the sequel $y(t)$ is also the value targeted for regulation.

1.8 Parameter identification

We obtained parameter estimates for the fold change model using data from the identification experiments illustrated in the main text (Fig. 1 c, e, g). The purpose of these experiments was to provide a rich enough dataset to facilitate the parameter estimation procedure, as well as to confirm intuition about the behavior of the system.

- Application of R light every 60 minutes (Fig. 1 g) resulted in constant synthesis of YFP over the time-course, achieving a balance between the production (maturation) and degradation of the mRNA and protein under quasi-constant signaling conditions ($s \sim 1$) and approaching an upper stationary condition for the system. Thus, one (implicit) algebraic constraint was established for four parameters of the system (a_m, $\hat{b}_m$, a_p, k).
- Application of R pulse at time 0, followed by the application of a FR pulse at 30, 60, 90, and 120 minutes (Fig. 1 c), or no application of FR (Fig. 1 g) were performed in order to isolate

Nature Biotechnology: doi:10.1038/nbt.2018

the transcriptional signal half-life from the remaining parameters (e.g. using the comparison of the cell culture treated with FR at 120 minutes with the cell culture not treated with FR). Additionally, the application of FR (hence, setting the transcriptional signal to 0) effectively removed the signal half-life a_s and the mRNA production rate $\hat{b}_m$ from effecting the tail ends of the time-response. Hence, the initial transient responses implicitly define a relationship between all system parameters and were used to further identify the characteristics of the signal decay and mRNA production, while the tail responses establish a balance between the mRNA degradation a_m, protein degradation a_p, and protein maturation k.

- Repeated application of tightly spaced R and FR light pulse trains (Fig. 1 e) can capture the time-response of the system starting from non-stationary initial conditions and $s = 0$, thus adding more implicit constraints to the free parameters.

To estimate free model parameters using the data described above, we employed an Approximate Bayesian Computation method[14]. Although the maximum likelihood parameter estimates were eventually used (reported in Supplementary Table 1), the posterior distributions obtained helped us quantify the uncertainty regarding each parameter, as well as identify correlations among parameters.

1.9 *In silico* feedback

1.9.1 Discrete-time switched linear system

A continuous-time switched linear system was developed for the fold change model of the light switch system. In this model, the switch or event takes the form of a state reset at discrete points in time. Specifically, we considered a switched linear system of the form

$$\dot{x} = Ax + B$$
$$y = Cx$$

where x is the state vector and A, B, and C are (constant) state matrices. State resets on $x_1 = s$ occur whenever light pulses are applied to the system. Since light pulses are applied to the system at a maximum frequency of 15 minutes, while measurements are taken every 30 minutes. We discretized this continuous time model with sampling rate of $\Delta t = 5$.

An exact discrete version of the switched linear system is given by the equations

$$x_k = e^{A\Delta t}x_{k-1} + \left(\int_0^{\Delta t} e^{A(\Delta t-\tau)}d\tau\right)B = \overline{A}x_{k-1} + \overline{B} \tag{3}$$

$$y_k = Cx_k = \overline{C}x_k. \tag{4}$$

1.9.2 *In silico* state estimation

Since YFP measurements are only reporting on the output of the system, other internal variables (e.g. m and p) need to be estimated. Therefore an observer (state estimator) is necessary to reconstruct (estimate) the unmeasured variables.

It is easy to check that the linear model developed above is observable, hence it is possible to construct an observer to estimate all the system states from the measured outputs (average fluorescence measurements).

We assume that the system (3) we want to observe has the form:

$$x_k = \overline{A}x_{k-1} + \overline{B} + w_k, \tag{5}$$

where $w_k \sim \mathcal{N}(0, Q)$ is the process noise, which is assumed to be Gaussian with covariance matrix Q (i.e. $w_k \sim \mathcal{N}(0, Q)$). Note that the stochastic term in the dynamics is also used to represent uncertainty with respect to parameter values.

At times $\hat{k} = 6k$ (since we consider a 30 minute sampling rate) we obtain a measurement from the cytometer (average fluorescence fold change)

$$y_{\hat{k}} = \overline{C} x_{\hat{k}} + v_{\hat{k}}, \tag{6}$$

where $v_{\hat{k}} \sim \mathcal{N}(0, R)$ is measurement noise of the system.

Given the process and measurement model presented above, we can construct the minimum variance filter for this system[15]. Thus, let us denote by $\hat{x}_{k|k}$ the estimate of the state and by $P_{k|k}$ the estimation error covariance matrix at time k, given measurement information up to time k. Then, between observations (i.e. when $\mod(k, 6) \neq 0$) these quantities evolve according to the difference equations

$$\hat{x}_{k|k-1} = \overline{A}\hat{x}_{k-1|k-2} + \overline{B} \tag{7}$$

$$P_{k|k-1} = \overline{A} P_{k-1|k-2} \overline{A}^T + Q, \tag{8}$$

that is, the observer is working "open loop". Next, when an observation is made, i.e. when $\mod(k, 6) = 0$, we correct the open loop estimates:

$$\hat{x}_{k|k} = \hat{x}_{k|k-1} + K_k(y_k - C\hat{x}_{k|k-1}) \tag{9}$$

$$P_{k|k} = P_{k|k-1} - K_k C P_{k|k-1}, \tag{10}$$

where the Kalman gain K is given by

$$K_k = P_{k|k-1} C^T (C P_{k|k-1} C^T + R)^{-1}. \tag{11}$$

The design of the state estimator weight matrices R and Q was based on *in silico* experimentation (data not shown), an underlying knowledge of the biological system, and an understanding of measurement quality. Given that the measurement of the average YFP was attained through flow cytometry and that measurements were highly reproducible (data not shown), we attributed very little noise to the measurements. Consequently the weight matrix R was set to a very small value $R = 0.05$ (which effectively forced the Kalman filter to believe the measured YFP as the truth). The design of the weight matrix Q was driven more by the inability of the model to reproduce the data perfectly than by any conceived process noise driving the system dynamics. Through simulation analysis, a weight matrix Q was constructed that, in effect, considers the state estimates of the dark YFP and mRNA to be highly inaccurate (equivalently driven by a significant amount of process noise) compared to the fluorescent YFP and signal. Hence, the process noise covariance was defined as $Q = \mathrm{diag}([0.1,\ \ 10,\ \ 10,\ \ 0.1])$.

1.9.3 *In silico* model predictive control

The predictive control strategy we consider is a finite time horizon optimal control problem with quadratic cost, where the optimization objective is to minimize a weighted deviation from the set point Y for the mature YFP fold change $\overline{f}$. Specifically, the optimization problem is given by

$$\min_{u_{\bar{k}}, \ldots, u_{\bar{k}+\bar{N}-1}} \ \sum_{i=k}^{N+k} (\overline{f}_i - Y)^T W^{(i)} (\overline{f}_i - Y),$$

where $\bar{k} = 3k$ (since control moves are allowed every 15 minutes), subject to the following constraints:

- $x_k = \hat{x}_{k|k}$.
- $x_{i+1} = \overline{A}x_i + \overline{B},\ \ \forall i = k, k+1, ..., k+N-1$.
- $\forall i = \bar{k}, \bar{k}+1, ..., \bar{k}+\overline{N}-1,\ u_i \in \mathcal{U}$.
- during the time horizon, i.e. $\forall i = k, k+1, ..., k+N$, the state of the system obeys the reset laws defined for the system model.

In the above problem formulation N is defined as the horizon length, $\overline{N}$ is the decision horizon length (note the relation $N = 3\overline{N}$), and $W^{(i)}$ is the penalty attributed to the squared distance of the system state from the set point at the ith prediction step. For our experiments we used $\overline{N} = 6$, which implies a horizon length $N = 18$, or 90 minutes. Since the decision time horizon $\overline{N}$ is relatively short, enumeration of all possible light pulse trains is sufficiently fast for the solution of the optimization problem above.

The selection of the predictive control weights $W^{(i)}$ for all $i = k, k+1, ..., k+N$ was made using simulation analysis (data not shown). In order to minimize overshoot and to emphasize a gentle convergence to the set point, the weights were eventually designed such that the cost was penalized more at the end of the horizon than at the beginning. Specifically, for $i = k, k+1, ..., k+N$ we considered the cost weights

$$W^{(i)} = (i - k + 1).$$

Pseudocode for the feedback implementation follows:

Algorithm 1 *In silico* Feedback

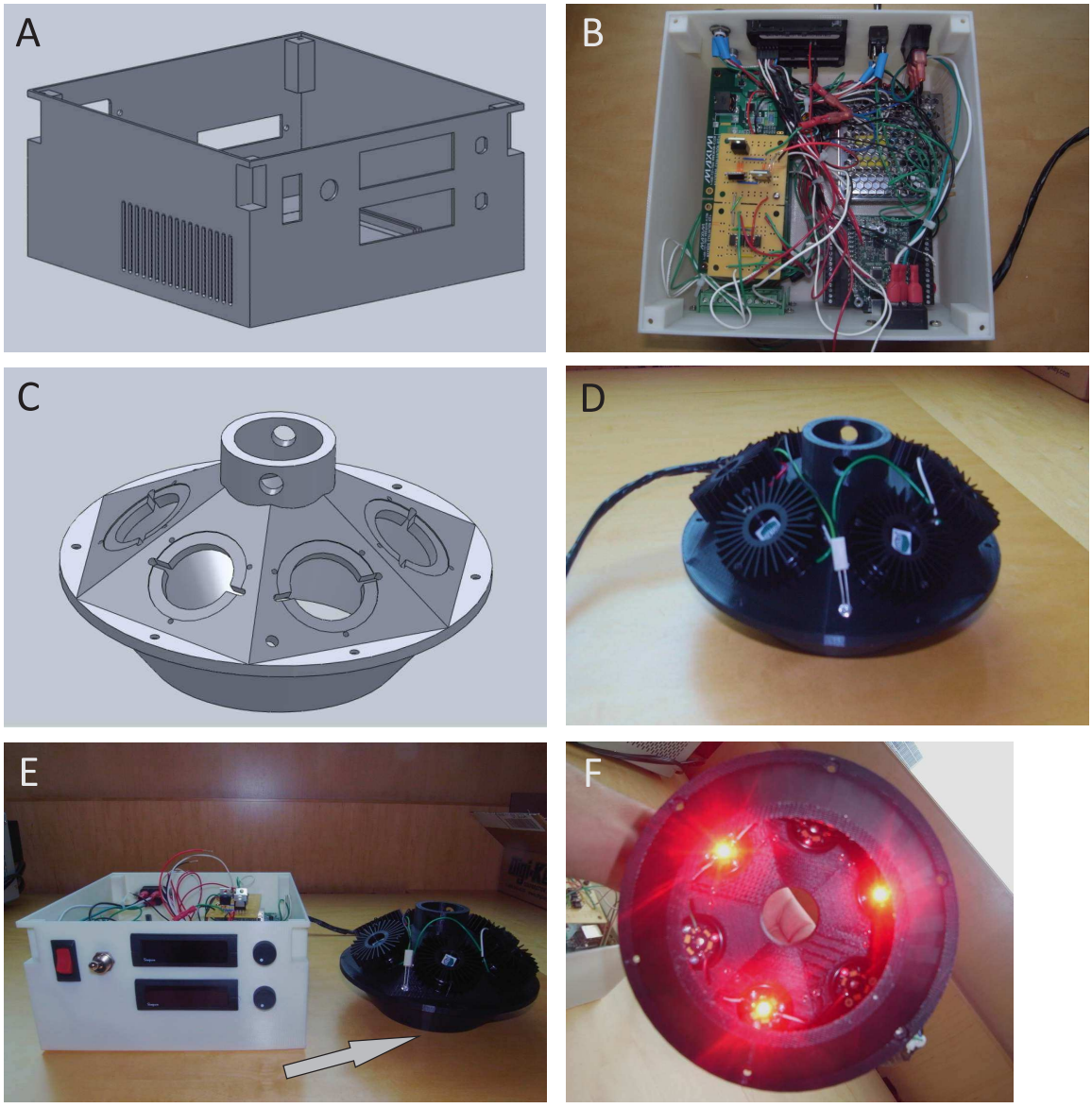
Initialize the state estimate of system $\hat{x}_{0|0}$ at time $k = 0$.
while $k < N$ **do**
  if mod $(k, 3) = 0$ **then**
    Solve the predictive control problem to obtain the optimal pulse train of light treatments $u_{\bar{k}}, ..., u_{\bar{k}+\overline{N}-1}$
    Apply the first light treatment from the optimal light pulse train $u_k = u_{\bar{k}}$ to the gene expression system
    if mod $(k, 6) = 0$ **then**
      Obtain a measurement of the YFP signal via flow cytometry
      Update the state estimate $\hat{x}_{k|k}$ using the Kalman Filter
    end if
  end if
  Evolve the Kalman filter gain $K_{k+1|k}$ and the state estimate $\hat{x}_{k+1|k}$ by one step
  $k = k + 1$
end while

References

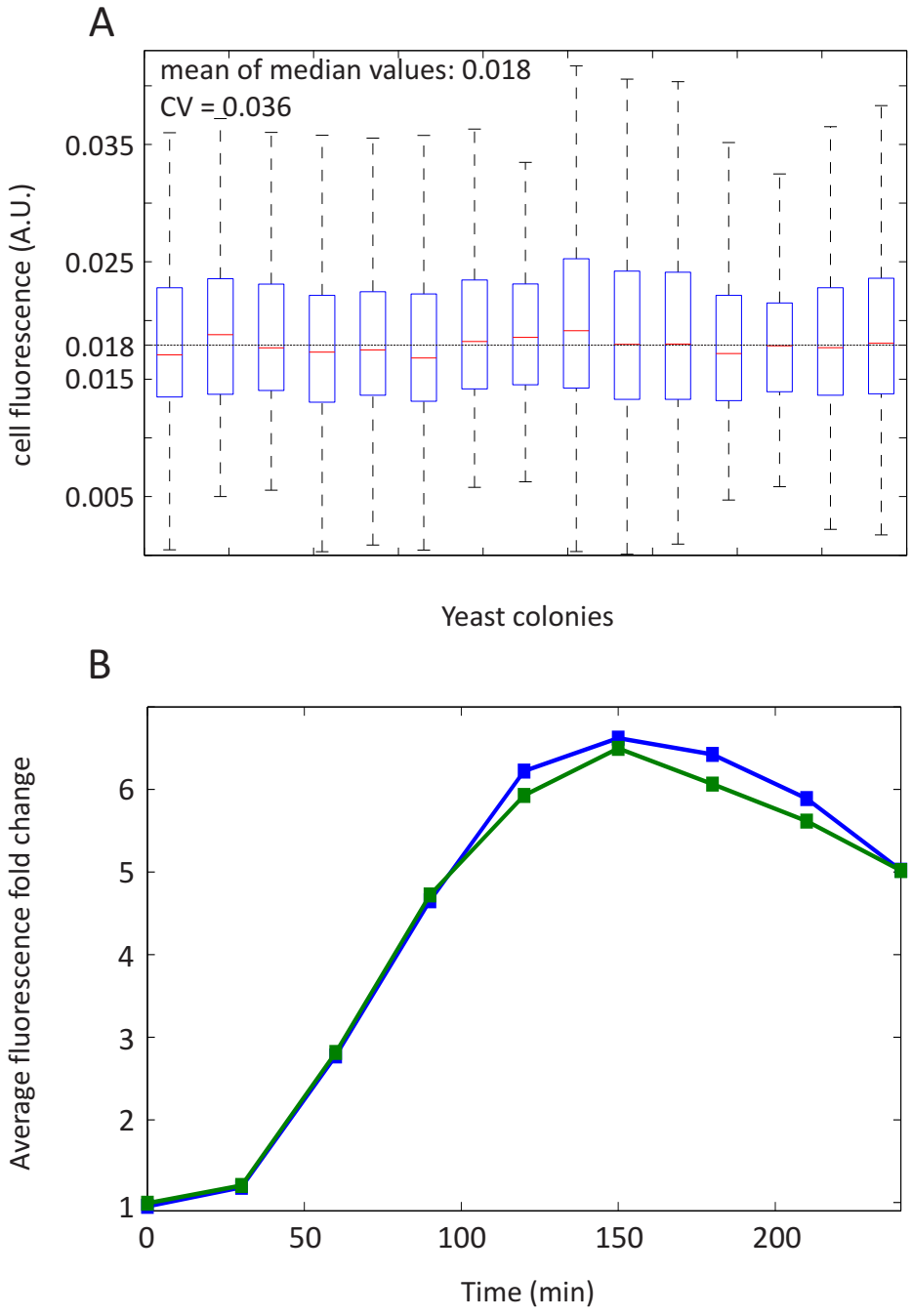
14. T. Toni, D. Welch, N. Strelkowa, A. Ipsen, M. P. H. Stumpf, Journal of The Royal Society Interface. 6, 187-202 (February 2009).

15. A. H. Jazwinski, Vol. 64 of Mathematics in Science and Engineering (Academic Press, Inc., London, 1970).

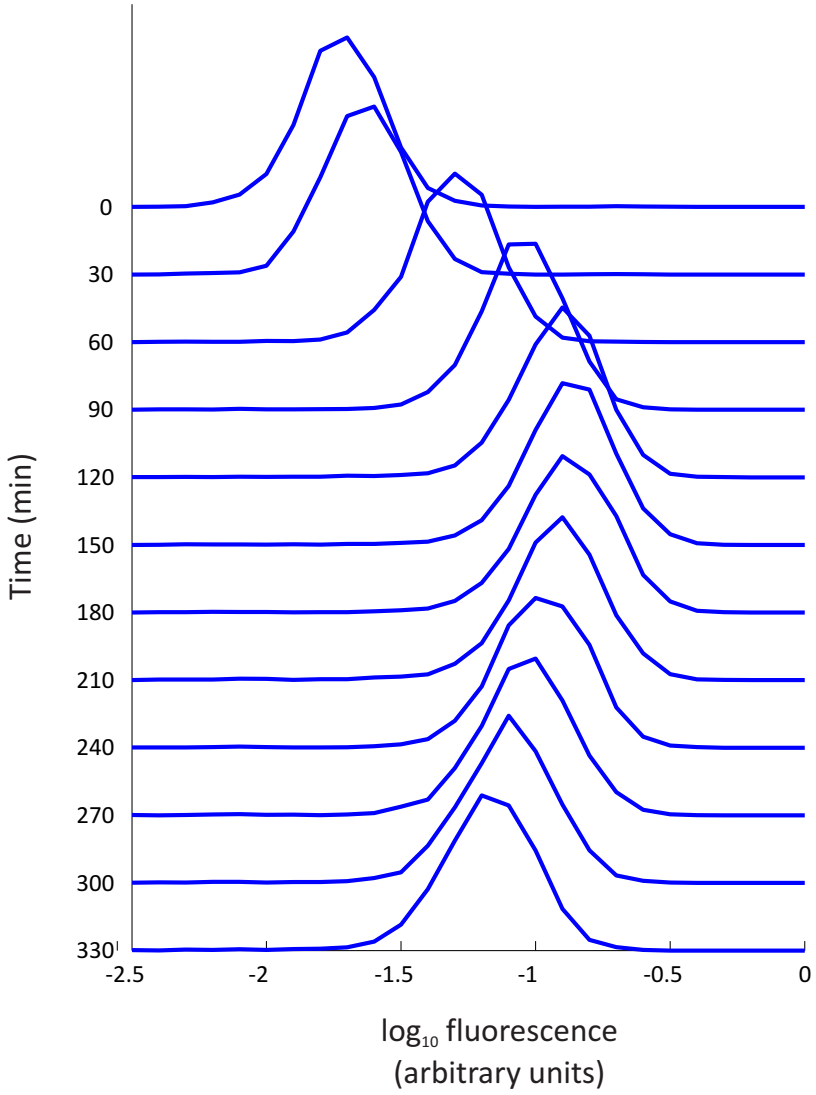
Nature Biotechnology: doi:10.1038/nbt.2018



Supplementary Figure 1: Custom-built light pulse delivery system. Both the "light pulser" and the electronics box were designed using 3D CAD (panels A and C) and prototyped using 3D printing technology (panels B and D). The electronics were implemented to allow both manual and computer-controlled operation of the system (via a National Instruments NI USB-6009 multifunction DAQ). The finalized prototype is very compact, portable and inexpensive. The plates containing the yeast cultures were placed underneath the pulse delivery system, as indicated by the arrow on panel E. Tubes containing yeast cultures can also be inserted in the round opening at the top of the system.

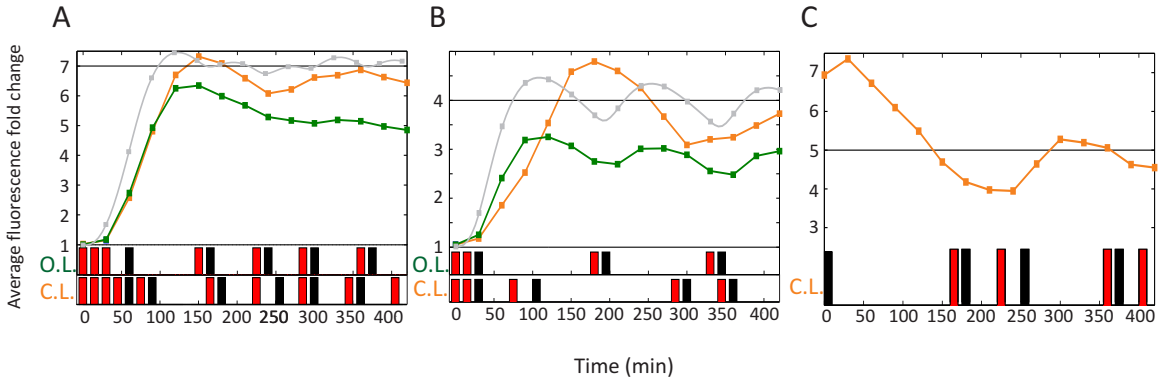


Supplementary Figure 2: Characterization of initial distribution and reproducibility of induction. A: Box plots of stationary ("steady-state") cell fluorescence distributions for 15 yeast colonies used for the characterization experiments. Yeast cells were grown according to the growth protocol described in the Materials and Methods and flow cytometry measurements were taken before the start of each experiment. The reproducibility of the basal fluorescence level enabled us to consider the average median fluorescence as a (fold change) reference value for all subsequent control experiments. Box plot legend: Central red mark: distribution median. Box edges: 25th and 75th percentiles. Whiskers: 99.3rd percentile. Outliers (lying beyond the whiskers) not shown. Horizontal black line: mean median value B: Dynamic response of the light-switchable system to a single R pulse given at $t = 0$. Results from two distinct yeast colonies are displayed, indicating the reproducibility of the system response.



Supplementary Figure 3: Dynamic response of the light-switchable system to a single R pulse applied at $t = 0$. Single cell measurements in the form of distributions are obtained by flow cytometry every 30 minutes for a period of 330 minutes. The distributions represent cell fluorescence normalized with respect to forward scatter and reported in log scale. While only the median values, reflecting the average population behavior, are used in the current study, the distributions contain additional information about the system that may be exploited in the future.

Nature Biotechnology: doi:10.1038/nbt.2018



Supplementary Figure 4: Biological replicates of the open- and closed-loop experiments reported in Fig. 3 of the paper. A: Regulation to 7-fold above the basal level. B: Regulation to 4-fold above the basal level. Open-loop trajectories are indistinguishable in the two replicates, while the closed-loop ones differ due to different light treatments starting at $t = 75$ minutes C: Regulation to 5-fold above basal, starting from unknown initial conditions. O.L. and C.L. denote open- and closed- loop control actions respectively.

Nature Biotechnology: doi:10.1038/nbt.2018