ARTICLES
https://doi.org/10.1038/s41589-018-0168-3
nature chemical biology

Escherichia coli "Marionette" strains with 12 highly optimized small-molecule sensors

Adam J. Meyer, Thomas H. Segall-Shapiro, Emerson Glassey, Jing Zhang and Christopher A. Voigt*

Cellular processes are carried out by many genes, and their study and optimization requires multiple levers by which they can be independently controlled. The most common method is via a genetically encoded sensor that responds to a small molecule. However, these sensors are often suboptimal, exhibiting high background expression and low dynamic range. Further, using multiple sensors in one cell is limited by cross-talk and the taxing of cellular resources. Here, we have developed a directed evolution strategy to simultaneously select for lower background, high dynamic range, increased sensitivity, and low cross-talk. This is applied to generate a set of 12 high-performance sensors that exhibit >100-fold induction with low background and cross-reactivity. These are combined to build a single "sensor array" in the genomes of E. coli MG1655 (wild-type), DH10B (cloning), and BL21 (protein expression). These "Marionette" strains allow for the independent control of gene expression using 12 small-molecule inducers.

Advances in biology are often tied to new methods that use external stimuli to control the levels of gene expression[1–3]. Pioneered in the early 1980s, so-called inducible systems were developed that allow genes to be turned on by adding a small-molecule inducer to the growth media[4]. These consist of a protein transcription factor (for example, LacI), whose binding to a DNA operator in a promoter is controlled by the inducer (for example, IPTG). Initially co-opted from natural regulatory networks, many versions of these systems have been designed over the years to improve performance. In the 1990s, additional naturally occurring systems that responded to other inducers, notably arabinose[5] and aTc[6], were developed into common tools in the field. In 1997, Lutz and Bujard published a seminal paper that combined three inducers (IPTG, arabinose, and aTc) that could be easily interchanged on a two-plasmid system[7]. Its organizational simplicity, compatibility, and quantified response functions were revolutionary. Beyond providing a new tool for biologists to control multiple genes with independent 'strings', it facilitated researchers with quantitative backgrounds to enter biology[8,9]. Armed with the new ability to control two genes with precision, physicists and engineers built the first synthetic genetic circuits, performed single-molecule experiments inside cells, deconstructed the origins of noise in gene expression, determined how enzyme balancing impacts metabolic flux, elucidated rules underlying the assembly of molecular machines, and built synthetic symbiotic microbial communities, just to highlight a few applications[10–14].

In synthetic biology, a transcriptional genetic sensor is defined as a unit of DNA that confers the ability to respond to a stimulus (for example, a chemical inducer) by controlling the activity of a promoter[15]. Sensor performance is quantified by its response function; in other words, how the concentration of inducer changes the activity of the output promoter (Fig. 1a). Often, this promoter retains a residual activity in the absence of inducer ('leakiness'). This hampers the ability to explore low expression levels or keep a gene in the off state, particularly needed for toxic proteins[16]. Another important parameter is the dynamic range, defined as the ratio of promoter activity in the on and off states. When this ratio is large, both high and low expression can be explored, as well as many intermediate states. The sensitivity is the concentration of inducer that turns a sensor on (defined as 50% activation). A lower sensitivity reduces the amount of a chemical that must be added to the media. Further, when multiple sensors are combined into one cell, they can interfere with each other's response functions (Fig. 1b). Some small molecules bind noncognate regulators, and this can lead to off-target activation (cross-reactivity) or competitive inhibition with the cognate small molecule (antagonism)[17–20]. Finally, each sensor requires cellular resources (for example, ribosomes) to function, and the activation of one sensor can influence another indirectly due to resource competition[21]. Each sensor also requires ~1–2 kb of DNA, and this becomes increasingly difficult to carry on plasmids. These challenges limit the number of sensors that can be put in a single cell, and the maximum reported to date is four[22].

Sensors can be improved using biophysical models, rational engineering, and directed evolution[7,17,19,23–28]. Improving performance requires screens or selections to be performed in the presence and absence of an inducer[29]. Such dual selections have been accomplished by sorting cells based on low and high fluorescence, using proteins that can be both toxic and selective (for example, TetA or HSV-TK), or by deploying separate positive and negative selections[19,27,29–33]. This has been applied to improving the response functions and eliminating cross-reactivity between pairs of regulators[17,19,27].

Here, we have developed a selection methodology that allows us to intervene at multiple steps to simultaneously select for improved response functions and decreased cross-talk (Fig. 1c). Using this method, we have generated a set of 12 optimized sensors with high dynamic range and low cross-reactivity. These have been integrated into the genomes of *E. coli* MG1655 (wild-type), DH10B (cloning), and BL21 (protein expression), resulting in the Marionette-Wild, Marionette-Clo, and Marionette-Pro strains. These strains (Supplementary Note 1) allow the independent control of up to 12 genes by applying 2,4-diacetylphophloroglucinol (DAPG), cuminic acid (Cuma), 3-oxohexanoyl-homoserine lactone (OC6), vanillic acid (Van), isopropyl β-D-1-thiogalactopyranoside (IPTG), anhydrotetracycline (aTc), L-arabinose (Ara), choline chloride (Cho), naringenin (Nar), 3,4-dihydroxybenzoic acid (DHBA), sodium

Synthetic Biology Center, Department of Biological Engineering, Massachusetts Institute of Technology, Cambridge, MA, USA. *e-mail: cavoigt@gmail.com

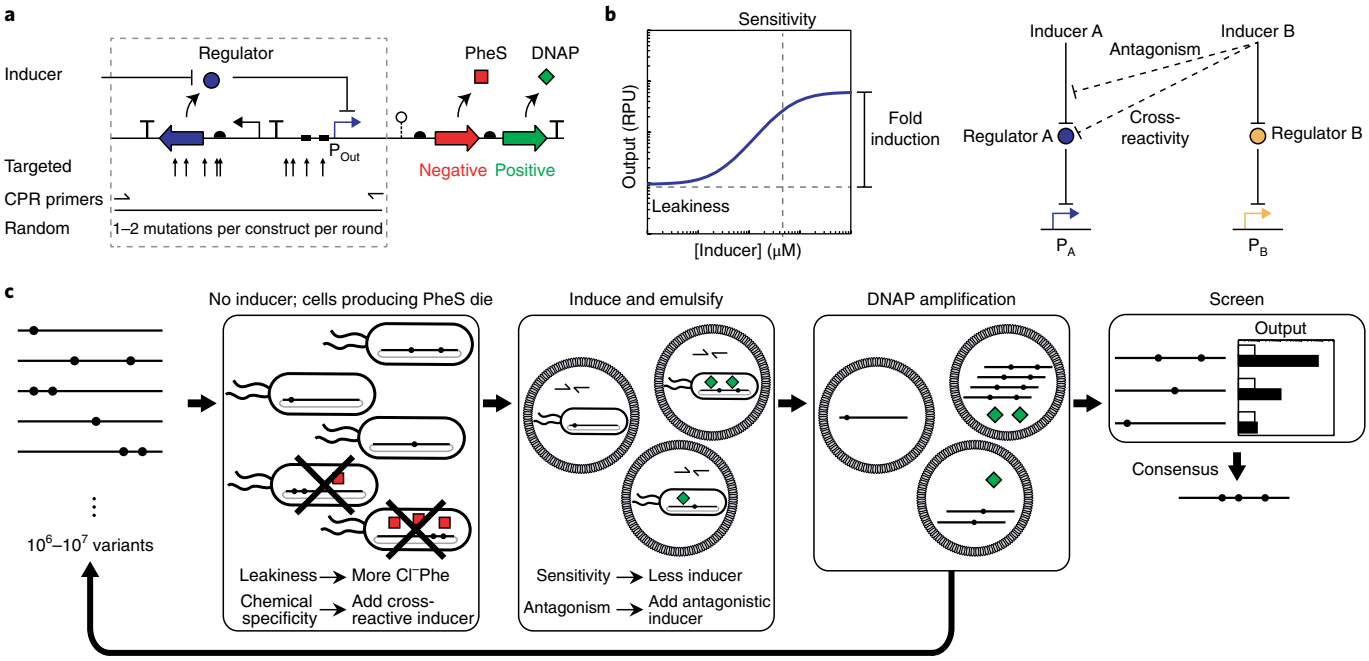

Fig. 1 | A dual selection for sensor optimization. a, A regulator (blue circle) is expressed from a weak constitutive promoter and controls expression from the output promoter (P_{Out}) while itself being controlled by an externally applied inducer molecule. Transcription from P_{Out} determines the expression level of an aminoacyl-tRNA synthetase (PheS; red square) and a DNA polymerase (DNAP: KOD or PK6; green diamond). Initial libraries may contain targeted degeneracy in the regulator RBS, regulator CDS, or P_{Out} (arrows). The library is amplified using the CPR primers shown, and random mutations are added throughout the entire library during each round of selection. **b**, A response function captures the activity of P_{Out} at various levels of inducer. Regulator A affects promoter A, while its own activity is affected by inducer A. Inducer B may affect regulator A (chemical cross-reactivity) or interfere with inducer A's action (antagonism). **c**, The dual-selection scheme is shown (see text for details). Dots denote mutations; half arrows represent PCR primers.

salicylate (Sal), and 3-hydroxytetradecanoyl-homoserine lactone (OHC14) to the growth media.

Results

Design of the Marionette sensor strains. First, we cloned 12 sensors that respond to different small molecules and made targeted changes to improve function. These sensors were then subjected to multiple rounds of negative selection to remove leaky sensors and positive selection to favor responsive sensors. This is accomplished by placing the transcription of a bicistronic mRNA under the control of the output promoter. Negative selection is facilitated by the first cistron, a promiscuous A294G mutant of the phenylalanine aminoacyl tRNA-synthetase (PheS). In the absence of inducer, leaky transcription of PheS leads to the charging of phenylalanyl-tRNA with the noncanonical amino acid 4-chloro-DL-phenylalanine (Cl-Phe). Adding more Cl-Phe increases the stringency of the selection by making PheS transcription more toxic[33,34]. After the negative selection, Cl-Phe is removed from the growth media, thus making PheS inert during the positive selection. Positive selection is facilitated by the second cistron, a thermostable DNA polymerase (DNAP), either KOD DNAP from the archaea *Thermococcus kodakarensis* (for stringent replication) or the engineered PK6 DNAP (to introduce mutations). In the presence of inducer, the level of DNAP can be quantified by PCR amplification of the sensor library after the cells are emulsified with primers that immediately flank the sensor library[35]. Encapsulated within its own water-in-oil compartment, the sensors responsible for the most expression of DNAP are most amplified by emulsion PCR. Multiple properties of the response function can be simultaneously improved during one cycle of negative and positive selection, and stringency can be altered by changing the concentrations of Cl-Phe and inducer. Chemical specificity can be achieved by adding a potentially cross-reactive inducer during the negative selection, thus selecting against cross-reactive mutants. Chemical antagonism can be selected against by adding a potentially antagonistic inducer to the positive selection, thus selecting for mutants that are less antagonized.

The dual selection is applied over multiple rounds to create a highly optimized set of sensors. From these, we identify 12 that can be used together in a single strain, and these are combined and integrated into the genomes of *E. coli* MG1655, DH10B, and BL21 to create the Marionette family of strains. Genome integration increases stability and reduces the cellular resources required to maintain regulator genes at low copy and without antibiotics[36]. It also simplifies the use of the inducible systems, in which only the output promoters need to be incorporated into a design (for example, for the expression of multiple proteins from a plasmid).

Initial sensor characterization and tuning. An initial set of 17 putative sensors were cloned and tested (Supplementary Tables 1–5). Each sensor consists of a weak constitutive promoter (P_{LacI}, Supplementary Table 1) driving the expression of the regulatory gene and an output promoter that is acted on by the regulator. The regulatory genes were either codon optimized and synthesized or cloned (Methods and Supplementary Table 4). The output promoters were either obtained from the literature or, in the case of P_{Van}, rationally designed by inserting cognate operator sequences into unregulated promoters (Supplementary Table 1)[37]. The activity of the output promoter was measured through the expression of yellow fluorescent protein (YFP) using flow cytometry and reported in relative promoter units (RPUs) (Supplementary Fig. 1 and Methods). Each complete sensor was cloned into the same sensor plasmid architecture (Supplementary Fig. 2). Some sensors require additional genes, which are encoded as an operon with the regulatory protein. For the Ara-inducible system, the transporter *araE*

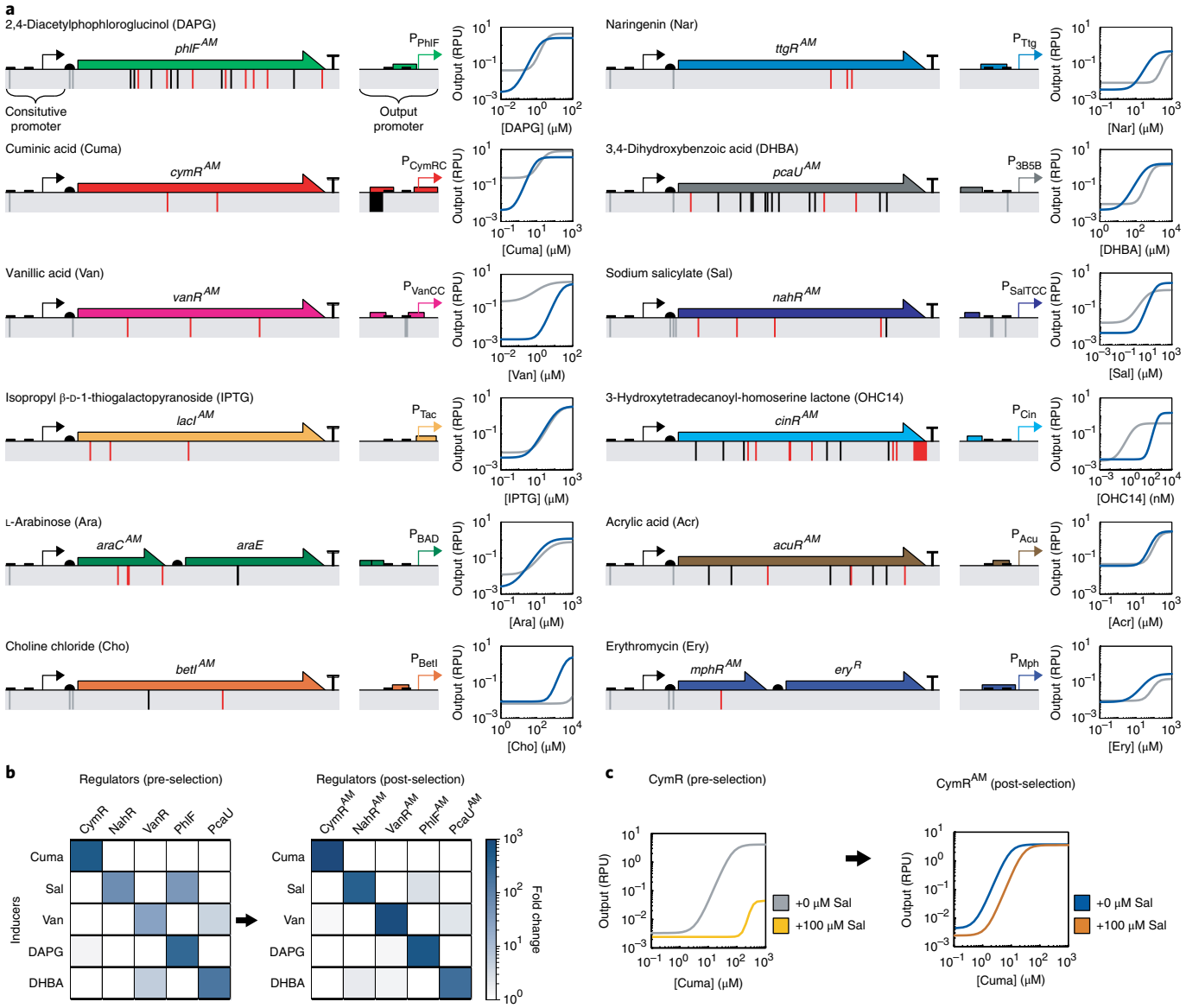

Fig. 2 | Improved sensor performance. a, The genetic design of each sensor is shown, with nonsynonymous (red) and noncoding or synonymous (gray/black) mutations noted. Mutations in gray were also applied to the parental sensor. The corresponding response functions comparing the evolved (blue) and parental (gray) sensors are shown. The fit of equation (1) to the mean of three replicates performed on different days is shown (see Supplementary Fig. 5 for data). The response function parameters for evolved sensors are provided in Table 1. **b**, The chemical cross-reactivity heat map of parental (left) and evolved (right) sensors are shown. Inducer concentrations were: 100 μM Cuma, 100 μM Sal, 100 μM Van, 10 μM DAPG, and 1 mM DHBA. The mean of three replicates performed on different days is shown (see Supplementary Fig. 8 for data). **c**, The response functions with Sal (light/dark orange) and without Sal (gray/blue) for parental CymR (left) and evolved CymR^{AM} (right) are shown. The fit of equation (1) to the mean of three replicates from different days is shown (see Supplementary Fig. 7 for data). The sequences of promoters and regulators are provided in Supplementary Tables 1 and 4. RPU, relative promoter units.

was included to produce a graded response[38]. For the Ery-inducible system, the ribosome methylase *ery*^{R} was included to confer resistance to Ery.

Rational mutations were made to improve some sensors before performing the directed evolution experiments. Multiple versions of each sensor were tested, each with a different promoter and RBS used to drive the expression of the regulator (Supplementary Fig. 3). The version with the largest dynamic range was chosen for further optimization. Then, a number of potential improvements that were gleaned from the literature or rationally designed were evaluated, the results of which are shown in Supplementary Fig. 4. After this step, we reduced the set of sensors to 14, removing copper, glucaric acid, and paraquat inducible systems because of inducer toxicity, incompatibility with rich media, interference with endogenous systems, and low response. The response functions of the initial sensors are shown in Fig. 2 (gray curves). The raw data points, including error bars, used for this fit are shown in Supplementary Fig. 5. Although they all show some response, the high leakiness, low dynamic range, and low sensitivity are apparent for many. Of the 14, the aTc and OC6 sensors produced a good enough response (high dynamic range and no cross-reactivity) to not require additional optimization (Table 1).

Sensor optimization by directed evolution. The remaining 12 were then subjected to directed evolution using the dual selection (Fig. 1c; Supplementary Fig. 6). For each sensor, a library was

Table 1 | Sensor response function parameters[a]

			Genome-based regulator (Marionette-Wild[c])				Plasmid-based regulator (*E. coli* DH10B)			
Inducer[d]		**Max inducer (μM)[b]**	y_{max} **(RPU)**	y_{min} **(RPU × 10^{-3})**	K **(μM)**	n	y_{max} **(RPU)**	y_{min} **(RPU × 10^{-3})**	**K (μM)**	n
DAPG	2,4-Diacetylphophloroglucinol	25	3.4	6.5	2.1	2.3	2.5	2.5	1.7	2.1
Cuma	Cuminic acid	100	7.6	3.6	22	2.3	3.7	4.3	8.9	2.4
OC6	3OC6-AHL	10	4.1	8.6	0.012	1.7	1.3	2.4	0.12	1.8
Van	Vanillic acid	100	4.5	4.0	14	2.1	3.0	2.4	26	2.3
IPTG	Isopropyl-β-D-thiogalactoside	1,000	2.6	8.3	190	1.7	3.3	4.8	140	1.8
aTc	Anhydrotetracycline HCl	0.2	3.2	3.6	0.012	4.4	2.4	4.9	0.013	3.8
Ara	L-Arabinose	4,000	1.8	3.6	43	1.7	1.2	2.4	37	1.5
Cho	Choline chloride	10,000	3.7	4.0	1900	2.0	2.6	8.5	4100	2.7
Nar	Naringenin	1,000	0.5	4.0	280	2.3	0.5	3.4	95	1.9
DHBA	3,4-Dihydroxybenzoic acid	1,000	0.8	8.0	240	1.5	1.6	4.5	370	1.8
Sal	Sodium salicylate	100	2.9	3.8	29	2.1	2.8	4.7	43	1.8
OHC14	3OHC14:1-AHL	10	1.2	3.6	0.25	3.0	1.5	3.0	0.43	2.3
Acr	Acrylic acid	1,000					3.1	37	130	2.5
Ery	Erythromycin	125					0.3	8.0	65	1.5

[a]Response data from 3 d were averaged and fit to equation (1) (Methods). Full response functions are provided in Supplementary Note 1. Definitions of parameters are provided in the main text. [b]Growth defects are observed above this concentration. [c]Based on *E. coli* MG1655. Data for Marionette-Clo (DH10B) and Marionette-Pro (BL21) are provided in Supplementary Note 1. [d]DAPG-Santa Cruz sc-206518; Cuma-Sigma 268402; OC6-Sigma K3007; Van-Sigma 94770; IPTG-Gold I2481C25; aTc-Sigma 37919; Ara-Sigma A3256; Cho-Sigma C7017; Nar-Sigma N5893; DHBA-Sigma 37580; Sal-Sigma S3007; OHC14-Sigma 51481; Acr-Sigma 147230; Ery-Sigma E5389.

constructed and cloned upstream of the operon containing PheS and DNAP (Fig. 1a) on a dual-selection plasmid. The initial library contained a mixture of random and rational mutations. For a subset of sensors, the output promoter was mutagenized: selected bases in the −10 box and −35 box were randomized. To control regulator expression, we randomized critical bases in the ribosome binding site (RBS). For LacI and AraC, we partially mutagenized key amino acid residues based on prior work[17,25,26]. Specifically, we partially PCR amplified the genes with primers of limited degeneracy (for example, VNA, WKK, or NDC; see Methods), thus allowing LacI Q18, F161, W220, Q291, and L296 and AraC L133, E165, E169, and C280 to sample a subset of possible amino acids. The specific mutations made to the initial library for each sensor are shown in Supplementary Note 2.

Multiple rounds of the dual selection were performed with the initial library. Between 4 and 23 rounds were performed, depending on how many issues needed to be corrected for each sensor. Different interventions were performed during each round to bias solutions to address problems identified for each sensor. The conditions for each round are presented in detail in Supplementary Note 2. Typically, the stringency of the negative selection was increased at a particular round by increasing the concentration of Cl-Phe from 2 mM to 4 mM. This biases against leakiness in the absence of inducer. During positive selection, inducer was added to the surviving cells, leading to the expression of DNAP. During early rounds, the maximum amount of inducer was added. For some sensors, we sought to increase the sensitivity of the response by reducing the amount of inducer after each round; for example, BetI was induced with 1 mM Cho in the final round of selection, down from 5 mM Cho in the first round. After induction, the cells were encapsulated, lysed, and PCR amplified using the DNAP expressed by the sensor (Methods). In early rounds, additional random mutations throughout the sensor were introduced by using the PK6 DNAP during positive selection (yielding an average of 1–2 mutations per kilobase). In later rounds, the stringent KOD DNAP was used to reduce the diversity in the population. After the amplification step of the positive selection, the constructs were recloned into the selection plasmid. Recloning allows the selection plasmid to be reset each round, thus preventing the accumulation of cheaters (for example, plasmids with a nonfunctional *pheS* gene) and offers the opportunity to change the DNAP as needed. In some cases, in an effort to combine multiple beneficial mutations into a single variant, the library was fragmented and reassembled (gene shuffling) between rounds (Methods).

During the selection rounds, additional interventions were included to reduce cross-talk between systems. For example, cross-talk between the IPTG and Ara sensors is well known, whereby IPTG reduces the output of the AraC/P_{BAD} system[17] (Supplementary Fig. 7). To identify mutants that reduce this cross-talk, 1,000 μM IPTG was included during all rounds of selection for the Ara sensor. IPTG was also included in the negative selection to prevent the evolution of an IPTG-induced AraC mutant. Cross-talk between Van and DHBA was also observed and was eliminated through negative selection (Fig. 2b; Supplementary Fig. 8). Salicylate was also found to antagonize the Cuma sensor (Fig. 2c; Supplementary Fig. 7) and cross-react with the DAPG sensor (Fig. 2b; Supplementary Fig. 8) and thus was added to both the positive and negative selections.

After all of the rounds of selection were complete, the library was assembled into a YFP screening plasmid (Supplementary Fig. 6). Several clones were picked and assayed for output expression in the presence and absence of inducer (Supplementary Note 2). In addition, when selecting against cross-talk, the clones were screened for induction by these molecules. The mutants showing the highest improvement (leakiness, sensitivity, dynamic range, and orthogonality) were identified and sequenced (Supplementary Note 2). Based on the mutations observed, one or more consensus sequences were constructed and re-screened to identify the best variant. The final sequences of the evolved sensors, including the mutations identified, are provided in Supplementary Table 4. Mutations were identified throughout the sensors, including the promoter and RBS controlling regulator transcription and translation, synonymous and nonsynonymous mutations throughout the regulator genes, and mutations/substitutions in the output promoters (Fig. 2a). On average, about eight cumulative mutations were made to each sensor as a result of the rounds of dual selection.

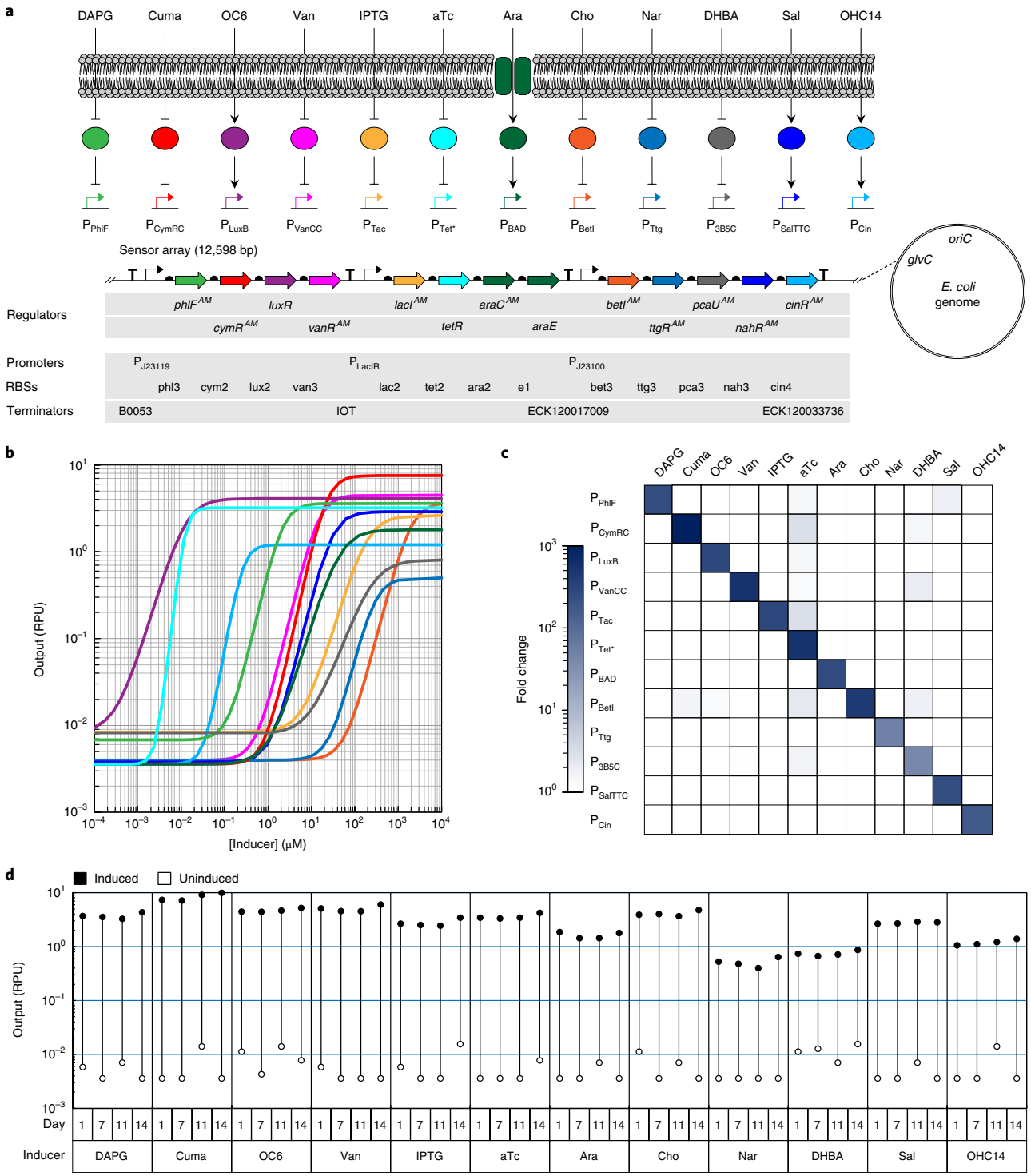

Fig. 3 | Marionette-Wild performance. a, The molecular and genetic schematic of the Marionette cluster and its location in the *E. coli* genome is shown. The cluster was inserted in the direction of leading-strand replication, between 3,860,010 and 3,861,627 in *E. coli* MG1655 (NCBI accession number NC_000913), between 3,957,594 and 3,959,211 in *E. coli* DH10B (NC_010473), and between 3,720,027 and 3,721,644 in *E. coli* BL21 (CP010816). Part sequences are provided in Supplementary Tables 1–5. The full cluster sequence is provided in Supplementary Table 6. **b**, The response functions of all 12 sensors are shown with colors corresponding to those in **a**. The lines represent fits of experimental data to equation (1), with the parameters provided in Table 1. Each individual line along with the corresponding data and error bars are provided in Supplementary Fig. 11. **c**, The chemical cross-reactivity heat-map of the 12 output promoters with each inducer is shown. Inducer concentrations were: 25 μM DAPG, 500 μM Cuma, 10 μM OC6, 100 μM Van, 1 mM IPTG, 200 nM aTc, 4 mM Ara, 10 mM Cho, 1 mM Nar, 2.5 mM DHBA, 250 μM Sal, and 10 μM OHC14. The mean of three replicates performed on different days is shown (see Supplementary Fig. 13 for data). **d**, The output of uninduced (open circles) and induced (closed circles) cultures on days 1, 7, 11, and 14 of passaging is shown. Inducer concentrations were: 25 μM DAPG, 500 μM Cuma, 10 μM OC6, 100 μM Van, 1 mM IPTG, 200 nM aTc, 4 mM Ara, 10 mM Cho, 1 mM Nar, 2.5 mM DHBA, 250 μM Sal, and 10 μM OHC14. A single evolutionary trajectory is shown. Passaging details and data from other days are provided in Supplementary Fig. 20.

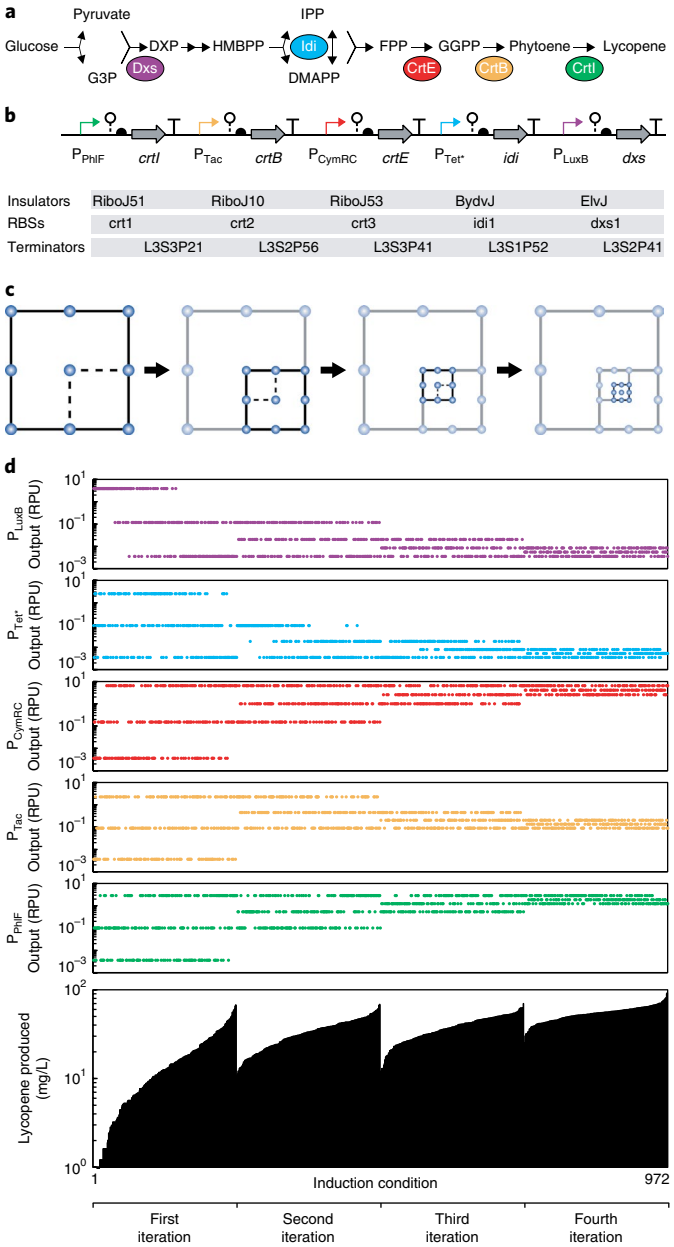

Fig. 4 | Marionette-Wild expression level optimization metabolic pathway balancing using Marionette-Wild. a, The metabolic pathway for the production of lycopene is shown. Enzymes known to be associated with increased production of lycopene are circled and colored. **b**, The genetic design for the sensor-controlled lycopene production pathway is shown. Expression of each of the five colored enzymes is controlled by a different sensor. Part sequences are provided in Supplementary Tables 1–5. **c**, The five-dimensional iterative grid search algorithm is visualized in two dimensions. Each circle represents a tested inducer combination. Dashed lines represent the bounds of a nested subvolume. The subvolume with the highest total titer is set as the search volume for the next iteration. **d**, The output of each promoter (colors) and the amount of lycopene production (black) is shown for each of 972 combinations of inducers, grouped by iteration (see Methods). Each vertical slice represents a single replicate. Cells grown with 3.3 nM aTc, 0 nM OC6, 500 μM Cuma, 26 μM IPTG, and 1.6 μM DAPG produced 90 mg/L lycopene.

The mutations occur in diverse locations, and it is difficult to deduce the mechanisms by which they impact performance. Still, some trends are apparent. Mutations to the output promoter occur either in the −35 or −10 box and only impact absolute expression. Mutations to the regulatory protein likely influence the ligand affinity and specificity, as well as its ability to bind DNA and interact with host transcriptional machinery. Some of these mutations come with trade-offs. For example, a mutation that restricts the ligand-binding pocket will reduce binding of cognate and noncognate ligands, thus decreasing sensitivity while increasing specificity. Similarly, mutations in the OHC14 sensor increased its dynamic range while decreasing its sensitivity (Fig. 2a).

The improvements in the response functions are shown in Fig. 2a (blue curves). Each function was obtained by fitting the experimental data to the equation

$$y = y_{\min} + (y_{\max} - y_{\min}) \frac{x^n}{K^n + x^n} \quad (1)$$

where y is the promoter activity in RPU (Methods), x is the concentration of the small molecule, $y_{\min}$ is the leakiness, K is the threshold (sensitivity), and n is the cooperativity. The fit parameters to equation (1) are provided in Table 1 and Supplementary Table 6 (raw data are provided in Supplementary Fig. 5). There is marked improvement in many of the response functions, sometimes showing orders of magnitude changes in the leakiness, dynamic range, and sensitivity. The Ery and Acr sensors showed slight improvements in their response functions, but were not chosen to be part of the final set because of their low dynamic ranges, which stem from a low $y_{\max}$ for the Ery and high $y_{\min}$ for the Acr sensor.

CymR, NahR, VanR, PhlF, and PcaU each responds to a substituted benzene. Therefore, we examined the activity of this set of sensors against all five inducers. The optimized sensors showed significant reduction in the cross-talk while maintaining high activity with their cognate inducer (Fig. 2b; Supplementary Fig. 8). Improvements in the antagonism between CymR and Sal were also tested (Fig. 2c; Supplementary Fig. 7). In the presence of 100 μM Sal, the ability for Cuma to induce its sensor drops by 1,200-fold. The sensor obtained by rounds of positive selection in the presence of Cuma reduces this by two orders of magnitude. There is also slight antagonism of AraC by IPTG, which also improved as a result of the selection (Supplementary Fig. 7). Collectively, these improvements allow all of these sensors to be used simultaneously in a single cell.

Creating the Marionette strains. The best 12 sensors were then combined to form a 'sensor array' that was inserted into the genome of *E. coli* MG1655 to create "Marionette-Wild" (Fig. 3; Supplementary Tables 7 and 8). Genomic insertion has several benefits: it stabilizes the cluster and simplifies the use of multiple systems without building large plasmids containing the regulators. The array consists of the 12 regulatory genes and an *araE* transporter organized into several operons (Fig. 3a; Supplementary Fig. 9). The genes were organized into three operons controlled by three medium-strength constitutive promoters. Each gene is encoded with its own RBS, which was rationally designed using the RBS calculator (Methods) to achieve an equivalent expression to that encoded on the plasmid (Methods and Supplementary Fig. 10). Strong terminators were included before and after the sensor array to insulate the array from context effects. Phage transduction was used to move the sensor array to create two additional cell lines: the RecA-deficient *E. coli* DH10B strain for cloning "Marionette-Clo" and the protease-deficient *E. coli* BL21 for protein expression "Marionette-Pro" (Methods and Supplementary Table 8).

The responses of the 12 genomically encoded sensors in Marionette-Wild are shown in Fig. 3b, the parameters derived from the fits to equation (1) are shown in Table 1 and Supplementary Table 9, and the raw data points are provided in Supplementary Fig. 11 and Supplementary Note 1. Each response was measured by transforming the strain with a p15A (Supplementary Fig. 2) plasmid containing the output promoter fused to YFP. Each response

function shows at least 100-fold induction with similar levels of on- and off-states. The technical information for the use of each sensor is organized as a series of datasheets in the Supplementary Information (Supplementary Note 1). Mass spectroscopy was performed to quantify potential differences between the intra- and extracellular concentrations of inducer, and only small differences were observed (Supplementary Table 10). The response of each sensor was also measured in Marionette-Clo and Marionette-Pro (Supplementary Note 1). The performances of the sensors closely match that of Marionette-Wild, with several exceptions; notably, the responses to IPTG (P_{Tac}) and choline (P_{BetI}) are leakier.

All of the sensors follow similar induction dynamics, with induction after 15 min and full induction by 2 h. Interestingly, those sensors based on activators were slower to turn on compared to those based on repressors (Supplementary Fig. 12). There is little cross-reactivity from the noncognate inducers (Fig. 3c and Supplementary Fig. 13). The response functions for each promoter were also measured in the presence of the maximum levels of all 11 other inducers, and there was little change in dynamic range, although the apparent sensitivity to Cuma, DHBA, and OHC14 is reduced when all 11 noncognate inducers are present (Supplementary Fig. 14). The sensor responses were measured during exponential growth. To evaluate performance in stationary phase, cells were grown overnight (~20 h) and response functions were measured (Methods). The responses closely matched those measured during exponential growth, with several exceptions exhibiting increased basal activity in the stationary phase (Supplementary Fig. 12). The performance of each reporter was also measured in different media compositions (LB, 2xYT, M9-glucose, and M9-glycerol). The induction is largely the same as that in LB media, although there is some variance in y_{min} among media compositions (Supplementary Fig. 15). Induction on plates in which the inducers are added to LB agar resulted in responses similar to those observed in early stationary phase (Supplementary Fig. 16).

The genome of Marionette-Wild was fully sequenced to identify changes that may have occurred during passage and cloning (Supplementary Table 8). The transcription and translation of the sensor array was also characterized using RNA-seq and ribosome profiling, thus providing the functions of all component parts and the relative expression levels of mRNA and protein (Supplementary Fig. 17; Supplementary Tables 11 and 12). These techniques can also be used to quantify the total impact of carrying the sensor array in the genome. Carrying the sensor array requires expressing 13 heterologous proteins that make up 1.3% of the proteome during exponential growth, as calculated from the ribosome profiling experiments (Methods and Supplementary Fig. 18). This is well below that required to impact growth[39]. Indeed, we find that the doubling times of Marionette-Wild, Marionette-Clo, and Marionette-Pro are 29.0 ± 1.9, 42.3 ± 3.8, and 31.1 ± 1.1 min, respectively, which are comparable to that of wild-type *E. coli* MG1655 (27.1 ± 1.2 min; Methods and Supplementary Fig. 19). The RNA-seq and ribosome profiling data also indicate that there is little impact on the expression of endogenous genes (notably, choline metabolism is repressed because of the expression of BetI) (Supplementary Fig. 18; Supplementary Tables 11 and 12).

Despite not having a measurable impact on growth, the sensor array's draw on cellular resources could confer a selective advantage to eliminating the array. Although genomic insertion improves evolutionary stability[16,40], it could still be disrupted over time, particularly for the *recA*-positive Marionette-Wild. To address this, we performed three independent experiments to assess the evolutionary stability of Marionette-Wild. First, we determined whether Marionette could reliably control a plasmid-based promoter, even after extended passage. The 12 reporter strains were passaged for 14 d in liquid culture without inducer, diluting the cells 10^6-fold each day (~280 cell doublings total). On each day, a subset of cells from each line was grown and assayed with and without inducer (Fig. 3d; Supplementary Fig. 20). Second, the Marionette-Wild strain was passaged for 9 d, streaking cultures on agar plates and inoculating single colonies into liquid culture each day. On the tenth day, the culture was transformed with each of the 12 reporter plasmids and assayed with and without the cognate inducer for each reporter. Third, we determined whether serial transfer would lead to the emergences of subpopulations[41]. We passaged the Marionette-Wild strain for 9 d in liquid culture, diluting the cells 10^6-fold each day (~180 cell doublings total). On the tenth day, the culture was transformed with each of the 12 reporter plasmids and assayed with and without the cognate inducer for each reporter. For all three evolution experiments, the sensors performed indistinguishably after growth and passaging (Supplementary Fig. 20). There was no decline in the fold-induction over time, and there was no emergence of 'broken' (constitutively on or constitutively off) subpopulations by flow cytometry (Supplementary Fig. 21).

Using Marionette-Wild to optimize a biosynthetic pathway. A metabolic engineering application was chosen to demonstrate the simultaneous use of multiple sensors for an optimization problem. A common observation was that the expression levels of enzymes need to be balanced to optimize flux through a pathway[42–45]. This can be accomplished by building large libraries in which the parts controlling expression (promoters, RBSs, etc.) are varied. There are two advantages to using Marionette: different expression levels can be tested without genetic perturbation, and continuous intermediate levels of expression can be probed.

A five-enzyme lycopene biosynthetic pathway (Fig. 4a) was selected, and each gene was placed under the control of different inducible promoters in Marionette-Wild (Fig. 4b). Strong RBSs (>60,000 arbitrary units) were designed to ensure that the optimal range of expression was explored (Methods). Importantly, it would be difficult to do this if the plasmid containing the pathway had to also carry the complete sensors; by placing the sensors on the genome, only the output promoters need to be used to control each gene.

A simple iterative grid search strategy was then implemented to identify the optimal levels of expression to maximize titer (Fig. 4c). Initially, we defined three levels of inducer for each sensor corresponding to low (zero), intermediate (half-maximum), and high (maximum) expression. This corresponds to 243 (3^5) combinations to fully characterize this space. All of these combinations were tested by varying inducer levels and the titer of lycopene evaluated (Fig. 4d). Using these data, the five-dimensional volume of the search space containing the optimum could be identified. Then, we defined the next search by selecting the midpoint along each dimension along with the two extrema. This creates a new set of 243 combinations, which we evaluated, and a new, smaller, hyper-dimensional volume is identified. This was repeated four times, after which the optimum was identified as producing 90 mg/L of lycopene. This is consistent with previous reports of optimization of lycopene biosynthesis[46]. However, our approach represents a significant advantage: only a single DNA construct had to be built to implement this search. A total of 972 combinations of inducer were tested, and if this were to be replicated with genetic parts, it would require 7 Mb of DNA construction. Although it is not cost-feasible to add five inducers to a fermentation, once the optimum is identified the inducible promoters can be replaced by constitutive ones, a process that we have demonstrated previously[45].

Discussion

Sensors have been used to control gene expression for decades, across all fields of biotechnology. However, the number of high-quality, well-characterized sensors available have been limited. Furthermore, each new use case has required placing the genes of interest into well-known expression plasmids (for example, pET,

pBAD, or Duet) or constructing and tuning large, intricate plasmids. As a result, the transcription of multigene systems is often controlled by a single sensor, and the relative expression level of each gene is controlled at the level of translation. This has limited the flexibility of the design of complex genetic systems, especially those based on functional RNAs.

This work represents a dramatic expansion in our ability to study and control genes in cells. The Marionette strains enable the modular control of up to 12 genes, simply by placing each one under the control of a small (50 to 289 base pair) inducible promoter. This means that a single construct can be built and then the expression levels perturbed in many ways through the combination of different small molecules. This could determine the role of proteins in a natural system, such as picking apart the stoichiometric requirements for a molecular machine[47,48]. This can also be part of rapid optimization of metabolic pathways, whereby the ideal stoichiometry can be identified without the need to build megabase-scale libraries[43,44]. Through the use of CRISPRi[49], sRNA[45], or other tools, endogenous genes can be inducibly downregulated as well as upregulated, thus enabling exquisite control of natural processes and metabolic flux. Further, dynamic systems can be probed by controlling the timing of the induction of each component to determine the role of ordered gene expression[13,50]. The ability to control gene expression has been a major limitation in genetic engineering; now, pulling the strings on Marionette enables unprecedented genetic control.

Online content

Any methods, additional references, Nature Research reporting summaries, source data, statements of data availability and associated accession codes are available at https://doi.org/10.1038/s41589-018-0168-3.

Received: 28 March 2018; Accepted: 5 October 2018;
Published online: 26 November 2018

References

1. Zuo, J. & Chua, N. H. Chemical-inducible systems for regulated expression of plant genes. *Curr. Opin. Biotechnol.* **11**, 146–151 (2000).
2. Keyes, W. M. & Mills, A. A. Inducible systems see the light. *Trends Biotechnol.* **21**, 53–55 (2003).
3. Mijakovic, I., Petranovic, D. & Jensen, P. R. Tunable promoters in systems biology. *Curr. Opin. Biotechnol.* **16**, 329–335 (2005).
4. de Boer, H. A., Comstock, L. J. & Vasser, M. The tac promoter: a functional hybrid derived from the trp and lac promoters. *Proc. Natl. Acad. Sci. USA* **80**, 21–25 (1983).
5. Guzman, L. M., Belin, D., Carson, M. J. & Beckwith, J. Tight regulation, modulation, and high-level expression by vectors containing the arabinose PBAD promoter. *J. Bacteriol.* **177**, 4121–4130 (1995).
6. Skerra, A. Use of the tetracycline promoter for the tightly regulated production of a murine antibody fragment in *Escherichia coli*. *Gene* **151**, 131–135 (1994).
7. Lutz, R. & Bujard, H. Independent and tight regulation of transcriptional units in *Escherichia coli* via the LacR/O, the TetR/O and AraC/I1-I2 regulatory elements. *Nucleic Acids Res.* **25**, 1203–1210 (1997).
8. Urban, J. H. & Vogel, J. Translational control and target recognition by *Escherichia coli* small RNAs in vivo. *Nucleic Acids Res.* **35**, 1018–1037 (2007).
9. Cookson, N. A. et al. Queueing up for enzymatic processing: correlated signaling through coupled degradation. *Mol. Syst. Biol.* **7**, 561 (2011).
10. Elowitz, M. B. & Leibler, S. A synthetic oscillatory network of transcriptional regulators. *Nature* **403**, 335–338 (2000).
11. Gardner, T. S., Cantor, C. R. & Collins, J. J. Construction of a genetic toggle switch in *Escherichia coli*. *Nature* **403**, 339–342 (2000).
12. Golding, I. & Cox, E. C. RNA dynamics in live *Escherichia coli* cells. *Proc. Natl. Acad. Sci. USA* **101**, 11310–11315 (2004).
13. Kalir, S. & Alon, U. Using a quantitative blueprint to reprogram the dynamics of the flagella gene network. *Cell* **117**, 713–720 (2004).
14. Golding, I., Paulsson, J., Zawilski, S. M. & Cox, E. C. Real-time kinetics of gene activity in individual bacteria. *Cell* **123**, 1025–1036 (2005).
15. Voigt, C. A. Genetic parts to program bacteria. *Curr. Opin. Biotechnol.* **17**, 548–557 (2006).
16. Caliando, B. J. & Voigt, C. A. Targeted DNA degradation using a CRISPR device stably carried in the host genome. *Nat. Commun.* **6**, 6989 (2015).
17. Lee, S. K. et al. Directed evolution of AraC for improved compatibility of arabinose- and lactose-inducible promoters. *Appl. Environ. Microbiol.* **73**, 5711–5715 (2007).
18. Scott, S. R. & Hasty, J. Quorum sensing communication modules for microbial consortia. *ACS Synth. Biol.* **5**, 969–977 (2016).
19. Tashiro, Y. et al. Directed evolution of the autoinducer selectivity of Vibrio fischeri LuxR. J. *Gen. Appl. Microbiol.* **62**, 240–247 (2016).
20. Halleran, A. D. & Murray, R. M. Cell-free and in vivo characterization of Lux, Las, and Rpa quorum activation systems in *E. coli*. *ACS Synth. Biol.* **7**, 752–755 (2018).
21. Gyorgy, A. et al. Isocost lines describe the cellular economy of genetic circuits. *Biophys. J.* **109**, 639–646 (2015).
22. Callura, J. M., Cantor, C. R. & Collins, J. J. Genetic switchboard for synthetic biology applications. *Proc. Natl. Acad. Sci. USA* **109**, 5850–5855 (2012).
23. Bintu, L. et al. Transcriptional regulation by the numbers: models. *Curr. Opin. Genet. Dev.* **15**, 116–124 (2005).
24. Salis, H., Tamsir, A. & Voigt, C. Engineering bacterial signals and sensors. *Contrib. Microbiol.* **16**, 194–225 (2009).
25. Daber, R., Sochor, M. A. & Lewis, M. Thermodynamic analysis of mutant lac repressors. *J. Mol. Biol.* **409**, 76–87 (2011).
26. Gatti-Lafranconi, P., Dijkman, W. P., Devenish, S. R. & Hollfelder, F. A single mutation in the core domain of the lac repressor reduces leakiness. *Microb. Cell. Fact.* **12**, 67 (2013).
27. Ike, K. et al. Evolutionary design of choline-inducible and -repressible T7-based induction systems. *ACS Synth. Biol.* **4**, 1352–1360 (2015).
28. Ellefson, J. W., Ledbetter, M. P. & Ellington, A. D. Directed evolution of a synthetic phylogeny of programmable Trp repressors. *Nat. Chem. Biol.* **14**, 361–367 (2018).
29. Yokobayashi, Y., Weiss, R. & Arnold, F. H. Directed evolution of a genetic circuit. *Proc. Natl. Acad. Sci. USA* **99**, 16587–16591 (2002).
30. Tang, S. Y., Fazelinia, H. & Cirino, P. C. AraC regulatory protein mutants with altered effector specificity. *J. Am. Chem. Soc.* **130**, 5267–5271 (2008).
31. Tashiro, Y., Fukutomi, H., Terakubo, K., Saito, K. & Umeno, D. A nucleoside kinase as a dual selector for genetic switches and circuits. *Nucleic Acids Res.* **39**, e12 (2011).
32. Taylor, N. D. et al. Engineering an allosteric transcription factor to respond to new ligands. *Nat. Methods* **13**, 177–183 (2016).
33. Maranhao, A. C. & Ellington, A. D. Evolving orthogonal suppressor tRNAs to incorporate modified amino acids. *ACS Synth. Biol.* **6**, 108–119 (2017).
34. Thyer, R., Filipovska, A. & Rackham, O. Engineered rRNA enhances the efficiency of selenocysteine incorporation during translation. *J. Am. Chem. Soc.* **135**, 2–5 (2013).
35. Ellefson, J. W. et al. Directed evolution of genetic parts and circuits by compartmentalized partnered replication. *Nat. Biotechnol.* **32**, 97–101 (2014).
36. Diaz Ricci, J. C. & Hernández, M. E. Plasmid effects on *Escherichia coli* metabolism. *Crit. Rev. Biotechnol.* **20**, 79–108 (2000).
37. Kunjapur, A. M. & Prather, K. L. J. Development of a vanillate biosensor for the vanillin biosynthesis pathway in *E. coli*. https://doi.org/10.1101/375287 (2018).
38. Khlebnikov, A., Datsenko, K. A., Skaug, T., Wanner, B. L. & Keasling, J. D. Homogeneous expression of the P(BAD) promoter in *Escherichia coli* by constitutive expression of the low-affinity high-capacity AraE transporter. *Microbiology* **147**, 3241–3247 (2001).
39. Scott, M., Gunderson, C. W., Mateescu, E. M., Zhang, Z. & Hwa, T. Interdependence of cell growth and gene expression: origins and consequences. *Science* **330**, 1099–1102 (2010).
40. Lee, J. W. et al. Creating single-copy genetic circuits. *Mol. Cell* **63**, 329–336 (2016).
41. Barrick, J. E. & Lenski, R. E. Genome dynamics during experimental evolution. *Nat. Rev. Genet.* **14**, 827–839 (2013).
42. Lee, M. E., Aswani, A., Han, A. S., Tomlin, C. J. & Dueber, J. E. Expression-level optimization of a multi-enzyme pathway in the absence of a high-throughput assay. *Nucleic Acids Res.* **41**, 10668–10678 (2013).
43. Zelcbuch, L. et al. Spanning high-dimensional expression space using ribosome-binding site combinatorics. *Nucleic Acids Res.* **41**, e98 (2013).
44. Smanski, M. J. et al. Functional optimization of gene clusters by combinatorial design and assembly. *Nat. Biotechnol.* **32**, 1241–1249 (2014).
45. Ghodasara, A. & Voigt, C. A. Balancing gene expression without library construction via a reusable sRNA pool. *Nucleic Acids Res.* **45**, 8116–8127 (2017).
46. Yoon, S. H. et al. Engineering the lycopene synthetic pathway in E. *coli* by comparison of the carotenoid genes of *Pantoea agglomerans* and P*antoea ananatis*. *Appl. Microbiol. Biotechnol.* **74**, 131–139 (2007).
47. Alon, U., Surette, M. G., Barkai, N. & Leibler, S. Robustness in bacterial chemotaxis. *Nature* **397**, 168–171 (1999).

48. Zhang, Y., Lara-Tejero, M., Bewersdorf, J. & Galán, J. E. Visualization and characterization of individual type III protein secretion machines in live bacteria. *Proc. Natl. Acad. Sci. USA* **114**, 6098–6103 (2017).
49. Larson, M. H. et al. CRISPR interference (CRISPRi) for sequence-specific control of gene expression. *Nat. Protoc.* **8**, 2180–2196 (2013).
50. Gupta, A., Reizman, I. M., Reisch, C. R. & Prather, K. L. Dynamic regulation of metabolic flux in engineered bacteria using a pathway-independent quorum-sensing circuit. *Nat. Biotechnol.* **35**, 273–279 (2017).

Acknowledgements

This work was supported by the US Office of Naval Research Multidisciplinary University Research Initiative grant #N00014-16-1-2388 (A.J.M., T.H.S.-S., E.G., J.Z., and C.A.V.). This work was supported in part by the Koch Institute Support (core) Grant P30-CA14051 from the National Cancer Institute. We would like to thank A.M. Kunjapur and K.L.J. Prather (Department of Chemical Engineering, Massachusetts Institute of Technology) for providing DNA templates for the amplification of P_{Van}, P_{3B5}, *vanR*, and *pcaU* and performing the initial characterization of VanR in *E. coli*. We would also like to thank S. Liu at the MIT-Broad Foundry for assisting in the RNA-seq and ribosome profiling sequencing run.

Author contributions

A.J.M. and C.A.V. conceived the study and designed the experiments; E.G. performed the mass spectrometry experiments; J.Z. performed the RNA sequencing and ribosome profiling experiments; A.J.M. performed all other experiments. A.J.M., T.H.S.-S., E.G., and J.Z., analyzed the data; A.J.M. and C.A.V. wrote the manuscript with input from all the authors.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information is available for this paper at https://doi.org/10.1038/s41589-018-0168-3.

Reprints and permissions information is available at www.nature.com/reprints.

Correspondence and requests for materials should be addressed to C.A.V.

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Methods

Strains, plasmids, and media. *Escherichia coli* DH10B (New England BioLabs, Ipswich, MA, USA) was used for all routine cloning and directed evolution. Plasmid-based regulator systems were characterized in *E. coli* DH10B. Marionette-Wild, -Clo, and -Pro were derived from *E. coli* MG1655 (ref. [51]), *E. coli* DH10B, and *E. coli* BL21 (New England BioLabs, Ipswich, MA, USA) cells, respectively. *E. coli* JTK164H was used to clone RK6 suicide vectors[52]. All other plasmids contain p15A origins of replication and kanamycin resistance (Supplementary Figs. 2 and 6). LB-Miller media (BD, Franklin Lakes, NJ, USA) was used for directed evolution, stability assays, and cytometry assays unless specifically noted. 2xYT liquid media (BD, Franklin Lakes, NJ, USA) and LB + 1.5% agar (BD, Franklin Lakes, NJ, USA) plates were used for routine cloning and strain maintenance and cytometry assays where noted. M9 media (1× M9 Salts (Millipore Sigma, St. Louis, MO,– USA), 2 mM $MgSO_4$, 100 µM $CaCl_2$, and 0.2% Casamino acids) supplemented with either 0.4% glucose or 0.4% glycerol was used for cytometry assays where noted.

Chemical inducers. Cells were induced with the following chemicals: 2,4-diacetylphophloroglucinol (DAPG) ≥97% purity from Santa Cruz Biotechnology (sc-206518); cuminic acid (Cuma) ≥98% purity from Millipore Sigma (268402); 3-oxohexanoyl-homoserine lactone (OC6) ≥98% purity from Millipore Sigma (K3007); vanillic acid (Van) ≥97% purity from Millipore Sigma (94770); isopropyl β-D-1-thiogalactopyranoside (IPTG) ≥99% purity from Gold Biotechnology (I2481C); anhydrotetracycline (aTc) ≥95% purity from Millipore Sigma (37919); L-arabinose (Ara) ≥99% purity from Millipore Sigma (A3256); choline chloride (Cho) ≥99% purity from Millipore Sigma (C7017); naringenin (Nar) ≥95% purity from Millipore Sigma (N5893); 3,4-dihydroxybenzoic acid (DHBA) ≥97% purity from Millipore Sigma (37580); sodium salicylate (Sal) ≥99.5% purity from Millipore Sigma (S3007); 3-hydroxytetradecanoyl-homoserine lactone (OHC14) ≥96% purity from Millipore Sigma (51481); acrylic acid (Acr) ≥99% purity from Millipore Sigma (147230); and erythromycin (Ery) ≥85% purity from Millipore Sigma (E5389).

Chemical transformation. For routine transformations, strains were made competent for chemical transformation. Overnight cultures (250 µl for *E. coli* DH10B derived cells, 100 µl for *E. coli* MG1655 and BL21 derived cells) were subcultured into 100 ml SOB media (BD, Franklin Lakes, NJ, USA) and grown at 37 °C, 250 r.p.m. for 3 h. Cultures were centrifuged (4,500 *g*, 4 °C, 10 min), and pellets were resuspended in 15 ml TFBI buffer[53] (30 mM KOAc, 50 mM $MnCl_2$, 100 mM RbCl, 10 mM $CaCl_2$, and 15% v/v glycerol, pH 5.0). After 1 h on ice, cells were centrifuged (4,500*g*, 4 °C, 10 min) and pellets were resuspended in 2 ml TFBII buffer (10 mM NaMOPS pH 7.0, 75 mM $CaCl_2$, 10 mM RbCl, and 15% v/v glycerol). Competent cells were stored at −80 °C until use.

Response function measurements (mid-log phase). All measurements shown were taken by cytometry of cells in mid-log growth except when noted. Glycerol stocks of strains containing the plasmids of interest were streaked on LB + 1.5% agar plates and grown overnight at 37 °C. Single colonies were inoculated into 1 ml LB + antibiotics in 2 ml 96-deep-well plates (USA Scientific, Orlando, FL, USA) sealed with an AeraSeal film (Excel Scientific, Victorville, CA, USA) and grown at 37 °C at 900 r.p.m. overnight in a Multitron Pro shaker incubator (INFORS HT, Bottmingen, Switzerland). The overnight growths were diluted 1:200 into 1 ml LB + antibiotics in 2-ml 96-deep-well plates + AeraSeal film and grown at 37 °C, 900 r.p.m. After 2 h the growths were diluted (*E. coli* DH10B/Marionette-Clo 1:500; *E. coli* BL21/Marionette-Pro 1:2,000; *E. coli* MG1655/Marionette-Wild 1:5,000) into prewarmed LB + antibiotics + inducer where necessary in 2 ml 96-deep-well plates + AeraSeal film and grown at 37 °C, 900 r.p.m. for 5 h. After growth, 20 µl of culture sample was diluted into 180 µl PBS + 200 µg/ml kanamycin.

Response function measurements (stationary phase). Measurements were taken from cells in the stationary phase to generate data shown in Supplementary Note 1; Supplementary Figs. 12 and 13. Glycerol stocks of strains containing the plasmids of interest were streaked on LB + 1.5% Agar plates and grown overnight at 37 °C. Single colonies were inoculated into 1 ml LB + antibiotics in 2 ml 96-deep-well plates (USA Scientific, Orlando, FL, USA) sealed with an AeraSeal film (Excel Scientific, Victorville, CA, USA) and grown at 37 °C, 900 r.p.m. overnight in a Multitron Pro shaker incubator (INFORS HT, Bottmingen, Switzerland). The overnight growths were diluted 1:200 into 1 ml LB + antibiotics in 2 ml 96-deep-well plates + AeraSeal film and grown at 37 °C, 900 r.p.m. After 2 h the growths were diluted (*E. coli* DH10B/Marionette-Clo 1:500; *E. coli* BL21/Marionette-Pro 1:2,000; *E. coli* MG1655/Marionette-Wild 1:5,000) into prewarmed LB + antibiotics + inducer where necessary in 2 ml 96-deep-well plates + AeraSeal film and grown at 37 °C, 900 r.p.m. for 20 h. After growth, 2 µl of culture sample was diluted into 198 µl PBS + 200 µg/ml kanamycin.

Time course (mid-log phase). For Supplementary Fig. 12: mid-log induction time course: glycerol stocks of strains containing the plasmids of interest were streaked on LB + 1.5% agar plates and grown overnight at 37 °C. Single colonies were inoculated into 1 ml LB + antibiotics in 2 ml 96-deep-well plates (USA Scientific, Orlando, FL, USA) sealed with an AeraSeal film (Excel Scientific, Victorville, CA, USA) and grown at 37 °C, 900 r.p.m. overnight in a Multitron Pro shaker incubator (INFORS HT, Bottmingen, Switzerland). The overnight growths were diluted 1:200 into 1 ml LB + antibiotics in 2-ml 96-deep-well plates + AeraSeal film and grown at 37 °C, 900 r.p.m. After 2 h the growths were diluted 1:500 into prewarmed LB + antibiotics. After 0, 1, 2, 3, 3.5, 4, 4.25, 4.5, or 4.75 h, cultures were further diluted 1:10 into prewarmed LB + antibiotics + inducer where necessary in 2 ml 96-deep-well plates + AeraSeal film and grown at 37 °C, 900 rpm for 5, 4, 3, 2, 1.5, 1, 0.75, 0.5, or 0.25 h (5 h total after the initial growth). After growth, 20 µl of culture sample was diluted into 180 µl PBS + 200 µg/ml kanamycin.

Time course (mid-log phase to stationary phase). For Supplementary Fig. 12: mid-log to stationary induction time course: glycerol stocks of strains containing the plasmids of interest were streaked on LB + 1.5% agar plates and grown overnight at 37 °C. Single colonies were inoculated into 1 ml LB + antibiotics in 2 ml 96-deep-well plates (USA Scientific, Orlando, FL, USA) sealed with an AeraSeal film (Excel Scientific, Victorville, CA, USA) and grown at 37 °C, 900 r.p.m. overnight in a Multitron Pro shaker incubator (INFORS HT, Bottmingen, Switzerland). The overnight growths were diluted 1:200 into 1 ml LB + antibiotics in 2 ml 96-deep-well plates + AeraSeal film and grown at 37 °C, 900 r.p.m. After 2 h the growths were diluted 1:5,000 into prewarmed LB + antibiotics + inducer where necessary in 2 ml 96-deep-well plates + AeraSeal film and grown at 37 °C, 900 r.p.m. for 5, 6, 7, 8, 9, 10, or 20 h. After growth, 2 to 20 µl of culture sample was diluted into 180–198 µl PBS + 200 µg/ml kanamycin.

Cytometry analysis. Fluorescence characterization with cytometry was performed on a BD LSR Fortessa flow cytometer with HTS attachment (BD, Franklin Lakes, NJ, USA). Cells diluted in PBS + kanamycin were run at a rate of 0.5 µl/s. Measurements were made using a green laser (488 nm) voltage of 400 V, a FSC voltage of 486 V, and a SSC voltage of 315 V. The events were gated by forward scatter height (mid-log: 1,000–10,000; stationary: 500–5,000) and side scatter area (mid-log: 1,000–10,000; stationary: 500–5,000) to reduce false events. After gating, 10^5–10^6 events were used for analysis. For each sample, the median YFP fluorescence was calculated. All output values are reported in terms of relative promoter units (RPU). For a given promoter measurement, the strain (*E. coli* DH10B, Marionette-Wild, etc.) is transformed with the plasmid. The strain is then assayed alongside a strain containing the RPU standard plasmid (Supplementary Fig. 1; Supplementary Note 4) as well as an autofluorescence control. The median autofluorescence value is subtracted from the all other median fluorescence values, including that of the RPU standard. The experimental sample value is then divided by the RPU standard value. The measured fluorescence corresponding to 1.0 RPU varies by cell type and growth conditions, but is typically in the range of 200–400 arbitrary units.

Library generation. Portions of the initial libraries were created by PCR using degenerate oligonucleotides (Integrated DNA Technologies Coralville, IA - USA). These fragments were joined into a degenerate, full-length sensor module by overlap PCR. Sensor modules were assembled into selection vectors by Golden Gate assembly. Linear insert and plasmid selection vector were mixed at 1:1 molar ratio totaling ~1 µg DNA along with 5 µl 10× T4 ligase Buffer, 1 µl T4 DNA ligase (2,000,000 U/ml), and 2 µl BbsI (10,000 U/ml) (all from New England BioLabs, Ipswich, MA, USA) in 50 µl total. Reactions were cycled 45 times for 2 min at 37 °C and 5 min at 16 °C, and then incubated for 30 min at 50 °C, 30 min at 37 °C, and 10 min at 80 °C in a DNA Engine cycler (Bio-Rad, Hercules, CA, USA). An additional 1 µl BbsI was then added, and the assembly was incubated for 1 h at 37 °C. Assemblies were then purified using Zymo Spin I columns (Zymo Research, Irvine, CA, USA). Host cells were transformed with library plasmid by electroporation. Supplementary Note 2 contains a depiction of all degeneracy found in the "Initial library" for each selection. Degenerate oligonucleotides are coded as follows: *N* = A25:C25:T25:G25; W = A50:T50; S = C50:G50; M = A50:C50; K = G50:T50; D = 33 A:33 G:33 T; V = 33 A:33 C:33 G.

Negative selection. LB media + 8 mM Cl-Phe (4-chloro-DL-phenylalanine; Millipore Sigma, St. Louis, MO - USA) was mixed, autoclaved, and stored at room temperature (25 °C). Cl-Phe has a tendency to adhere to glassware, and care was taken to avoid disturbing the water + LB powder + Cl-Phe powder mixture before autoclaving. LB media + 8 mM Cl-Phe was mixed with plain LB media to achieve the desired concentration of Cl-Phe (See Interventions:[Cl-Phe] for each selection in Supplementary Note 2). Following transformation and outgrowth, cultures were diluted into 7 ml LB + Cl-Phe supplemented with antibiotics as well as any cross-reactive inducers (See Interventions:[Negative ligand] for each selection in Supplementary Note 2). Cultures were grown at 37 °C for 12–16 h.

Positive selection. Following negative selection, cultures were diluted 1:100 into 2 ml LB + antibiotics in culture tubes and grown at 37 °C, 250 r.p.m. for 2 h. The cultures were then diluted 1:100 into 2 ml prewarmed LB + antibiotics + inducer (See Interventions:[Positive ligand] for each selection in Supplementary Note 2) and grown at 37 °C, 250 r.p.m. for 4 h. Following induction, 1 ml of culture was centrifuged at 5,000 *g*, 25 °C, 10 min. Supernatant was removed and the cell pellet was resuspended in 50 µl 1× CPR buffer (50 mM Tris–HCl pH 8.8, 10 mM KCl,

2 mM $MgSO_4$, 10 mM $(NH_4)_2SO_4$). 5 µl resuspension was added to 95 µl of 1× CPR buffer plus 0.4 µM CPR primers and 200 µM dNTPs. This aqueous phase was added to a 2 ml centrifuge tube containing 438 µl Tegosoft DEC (Evonik, Essen, Germany), 42 µl AbilWE09 (Evonik, Essen, Germany) and 120 µl mineral oil (Millipore Sigma, St. Louis, MO, USA), with the rubber stopper from a 1 ml syringe. The mixture was vortexed at maximum setting for 2 min. The emulsion was split evenly into five 0.2 ml PCR tubes, and thermal cycled (95 °C for 5 min; 20 cycles of [95 °C for 30 s, 55 °C for 30 s, 72 °C for 2 min/kb]; and 72 °C for 5 min). Emulsions were then centrifuged (10,000 *g*, 25 °C, 10 min) and the upper (oil) phase was removed. Then, 100 µl H_2O and 500 µl chloroform were added, and the mixture was pipetted to disrupt the pellet. The resuspension was then transferred to a 1.5 ml heavy-gel phase-lock tube (5 Prime, San Francisco, CA, USA) and centrifuged (16,000 *g*, 25 °C, 2 min). The upper (aqueous) phase was collected, and DNA was purified using Zymo Spin I columns. The library was amplified in a recovery PCR using Accuprime *Pfx* and nested recovery primers, gel purified, and assembled as described above. A depiction of the CPR primer and recovery primer amplification are provided in Supplementary Fig. 6.

Library shuffling. Between some rounds of selection (See Interventions:Notes for each selection in Supplementary Note 2), libraries were shuffled upon themselves such that fragments of each mutant mutually serve as both primer and template for a primer-less PCR amplification[54]. 1 µg linear, library DNA (from the recovery PCR) was added to a mild DNAse reaction (500 mM Tris pH 7.4, 100 mM MnCl2, 0.5 U DNAse [New England BioLabs, Ipswich, MA, USA]) and lightly digested for 3 min at 15 °C. Fragmented DNA was purified using Zymo Spin I columns and reassembled in a primer-less PCR in 1× KAPA HiFi Master Mix (KAPA Biosystems, Wilmington, MA, USA) by thermal cycling (95 °C for 2 min; 35 cycles of [95 °C for 30 s, 65 °C for 90 s, 62 °C for 90 s, 59 °C for 90 s, 56 °C for 90 s, 53 °C for 90 s, 50 °C for 90 s, 47 °C for 90 s, 44 °C for 90 s, 41 °C for 90 s, 68 °C for 90 s]; and 72 °C for 4 min). The reassembly was purified using Zymo Spin I columns, re-amplified using Accuprime *Pfx* and CPR primers, gel purified, and assembled as described above.

On/off screen. At the end of each selection, libraries were assembled into the YFP screening plasmid (Supplementary Fig. 6). Host cells were transformed and plated on LB-agar. Between 35 and 92 individual clones were picked and assayed by cytometry as described. Cells were grown with no inducer, in the presence of cognate inducer, and when necessary in the presence of relevant noncognate inducers. Measurements were taken of cells in mid-log phase by cytometry as described. The most promising clones, as judged by dynamic range and orthogonality, were mini-prepped, and the sensor region was sequenced.

Genomic integration. In preparation for recombineering, cells were transformed with a plasmid containing Ara-inducible λ Red recombination machinery with a temperature-sensitive origin of replication[55]. 50 µl of overnight culture was subcultured in 50 ml LB medium and grown at 30 °C, 250 r.p.m. for 2 h. 2 mM Ara was added, and the culture continued to grow at 30 °C, 250 r.p.m. for 3 h. The culture was then centrifuged (4,500 *g*, 4 °C, 10 min) and washed with ice-cold 10% glycerol four times, with the fourth resuspension in 200 µl 10% glycerol. Recombineering-ready cells were stored at −80 °C until use. For the first insertion, six genes (*phlF*^{AM}, *cymR*^{AM}, *luxR*, *vanR*^{AM}, *lacI*^{AM}, and *tetR*) were Golden Gate assembled using BsaI into an RK6 suicide vector (which needs Pir protein in order to propagate)[52]. Pir-expressing *E. coli* JTK164H cells were transformed, and plasmids were purified, verified, and linearized with BpiI leaving homology to the *glvC* pseudogene. Recombineering-ready *E. coli* MG1655 cells were electroporated and transformed with gel-purified, linearized inserts. After an outgrowth of 1 h at 37 °C, transformations were plated on LB-agar plates + antibiotic (5 µg/ml chloramphenicol). Colonies were picked and grown at 37 °C in LB + antibiotic, and the presence of the insert was verified by colony PCR. For the second insertion, this strain was made recombineering-ready, and the process was repeated with the next set of genes (*araC*^{AM}, *araE*, *betI*^{AM}, and *ttgR*^{AM}) with the insert containing homology to *tetR* and *glvC* and conferring resistance to 20 µg/ml spectinomycin (replacing chloramphenicol resistance). For the third insertion, this strain was made recombineering-ready, and the process was repeated with the final set of genes (*pcaU*^{AM}, *nahR*^{AM}, and *cinR*^{AM}) with the insert containing homology to *ttgR*^{AM} and *glvC* and conferring resistance to 5 µg/ml chloramphenicol (replacing spectinomycin resistance). See Supplementary Fig. 9 for schematic details.

Phage transduction to transfer the Marionette cluster. 50 µl of overnight culture of Marionette-Wild was subcultured in 5 ml LB medium + 0.2% glucose and 5 mM $CaCl_2$ and grown at 37 °C, 250 r.p.m. for 30 min. 100 µl phage P1 lysate (10^9 pfu/ml) was added and the culture was grown at 37 °C, 250 r.p.m. until lysis (3 h)[56]. 50 µl chloroform was added, and the culture continued at 37 °C, 250 r.p.m. for 5 min. The culture was centrifuged (9,200 *g*, 25 °C, 10 min), and the supernatant was filtered (0.45 µM) and stored at 4 °C until use. 1.5 ml of overnight culture of *E. coli* MG1655, BL21, or DH10B (the *E. coli* DH10B cells contained a temperature-sensitive plasmid transiently expressing RecA, derived from the temperature sensitive, Ara-inducible, λ Red recombination plasmid[55] with *recA* replacing λ *gam*, *beta*, and *exo*) was centrifuged (10,000 *g*, 25 °C, 5 min). The cell pellet was resuspended in 750 µl P1 buffer (5 mM $MgSO_4$ and 10 mM $CalCl_2$), and up to 100 µl Marionette-P1 phage lysate was added. After 30 min at 25 °C, 1 ml LB + 200 µl 1 M sodium citrate was added and the culture was grown at 37 °C, 250 r.p.m. for 30 min. The culture was centrifuged (10,000 *g*, 25 °C, 2 min), and the pellets were resuspended in 100 µl LB and plated on LB-agar plates + 5 µg/ml chloramphenicol and 5 mM sodium citrate. Colonies were picked and grown at 37 °C in LB + 5 µg/ml chloramphenicol and 5 mM sodium citrate, and the presence of the insert was verified by colony PCR. The entire Marionette cluster was verified by sequencing PCR amplicons of the cluster from the genome of Marionette-Wild.

Genome sequencing. Genomic DNA from Marionette and parental strains were purified using the Wizard Genomic DNA Purification Kit (Promega, Madison, WI, USA) per the manufacturer's instructions. Samples were submitted for deep sequencing at MIT BioMicro Center (Cambridge, MA, USA). Samples were prepared using the Nextera XT DNA Library Preparation Kit (Illumina, San Diego, CA, USA) and sequenced using a MiSeq 300 cycle kit (Illumina, San Diego, CA, USA). FASTQ reads were assembled to the reference genome using Geneious v6.1.8 Software (Biomatters Ltd., Auckland, New Zealand). The consensus assembly was extracted and aligned to the reference sequence, and differences were noted.

Sensor induction on plates. Cells were transformed with reporter plasmids and, following outgrowth at 37 °C for 1 h, plated on LB-agar (LB-Miller powder + 1.5% agar (BD, Franklin Lakes, NJ, USA)) with and without the appropriate inducer. After an overnight incubation at 37 °C for 16 h, plates were incubated at 4 °C for 1 h and imaged using a ChemiDocMP Imaging system (Bio-Rad, Hercules, CA, USA) and Image Lab 4.0 software (Bio-Rad, Hercules, CA, USA) employing blue epi illumination, a 530/28 filter, and a 0.1 s exposure. Raw images were rotated and cropped using XnView (XnSoft, Reims, France). Individual colonies were scraped from the plate, resuspended in 200 µl PBS + 200 µg/ml kanamycin, and assayed by cytometry as described in 'Cytometry analysis'.

Ribosome profiling and RNA sequencing. Individual colonies were inoculated in LB broth in a 14 ml culture tube at 37 °C and 250 r.p.m. overnight. Cultures were diluted 200-fold into 1 ml LB media and grown under the same condition for 3 h. Cells were further diluted 500-fold into 250 ml LB media in 2,800 ml Erlenmeyer flasks. Culture flasks were incubated in an Innova shaker (Eppendorf, Hamburg, Germany) at 37 °C and 200 r.p.m. until the OD_{600} reached 0.3. Cells were rapidly filtered at room temperature using a nitrocellulose filter paper with a 0.22 µm pore size (GVS Life Sciences, Bologna, Italy), collected with a prewarmed spatula and flash frozen in liquid nitrogen. Cell pellets were saved in −80 °C until use. Ribosome footprinting and library preparation were done as previously described[57]. RNA sequencing was performed on the same cell lysate acquired from the above procedure. The ribosomal RNA fraction was subtracted with MICROExpress bacterial mRNA enrichment kit (Thermo Scientific Ambion, Waltham, MA, USA). The enriched mRNAs were fragmented in a heat block at 95 °C for 15 min. Fragments were converted to cDNA libraries using NEBNext Ultra RNA library prep kit for Illumina (New England BioLabs, Ipswich, MA, USA). Sequencing was performed on an Illumina HiSeq 2500 using a rapid run mode. For ribosome profiling libraries, 50-cycle single-end run was used. For RNA-seq cDNA libraries, 200-cycle pair-end run was implemented.

Processing of sequencing data. To process the ribosome profiling sequencing reads, adaptors were first trimmed from the raw reads by Cutadapt. Reference genomes 'sAJM.1506_Marion' and 'sAJM.240_NC_0009' modified from NC_000913.2 (NCBI) were used for Bowtie 1.1.2 sequence alignment, allowing for one mismatch and no more than two reported alignments. A center-weighting approach was taken to map the aligned footprint reads ranging from 22 to 42 nucleotides to the reference genome. For each footprint read, 11 nucleotides on either end were removed, and the nucleotides in the remaining region were given equal score, normalized by the length of the center region. Ribosome density was calculated by averaging reads mapped onto a given gene, excluding the first and last five codons, normalized by the total number of mapped reads. To adjust the elevation of densities for the first 50–100 codons, an exponential decay function was fitted to the average density at every 50 codons compared to the first 50 codons for genes with more than 128 mapped reads. The fitted function was subsequently used to adjust the ribosome density for each gene. To remove effects of ribosome pausing and other variations, the top and bottom 5% of the ribosome occupancy at each nucleotide of a gene is removed (90% winsorization) before averaging for the ribosome density calculation of each gene. The mass fraction in the proteome occupied by each protein, Φ_i, was estimated:

$$\Phi_i = \frac{RD_i \cdot MW_i}{\sum_P RD_P \cdot MW_P}$$

where the product of the ribosome density of protein *i*, RD_i, and its molecular weight, MW_i, was normalized by summation from all the proteins *P*. The fraction of the proteome corresponding to the Marionette cluster was obtained by summing the proteome fraction (reads per total mapped reads) of each of the component

genes. Raw reads from RNA sequencing were mapped to the reference genomes using BWA 0.7.4. The output SAM files were converted into a sorted and indexed BAM format using the "view" command of the SAMTools. The "view" command was also used to calculate the total mapped reads of each sample from its BAM file, with the options "-c –F 4". To obtain read counts of each gene, the "htseq-count" command of the HTSeq software 0.6.0 was applied given a GFF file and options "-s reverse –a 10 –m union". The fragment per kilobase per million mapped (FPKM) values were generated using the edgeR package 3.22.2. FPKM values between samples were normalized by applying a trimmed mean of *M*-values normalization method. With the same command, each BAM file was further split into two BAM files containing information from either the sense strand or the antisense strand. The separate BAM files were processed by Bedtools to obtain the base coverage. All scripts were written and executed in Python 3.4.

Plate reader assays to measure growth rates. Glycerol stocks of strains of interest were streaked on LB + 1.5% agar plates and grown overnight at 37 °C. Single colonies were inoculated into 1 ml LB in 2 ml 96-deep-well plates (USA Scientific, Orlando, FL, USA) sealed with an AeraSeal film (Excel Scientific, Victorville, CA, USA) and grown at 37 °C, 900 r.p.m. overnight in a Multitron Pro shaker incubator (INFORS HT, Bottmingen, Switzerland). The overnight growths were diluted 1:200 into 1 ml LB in 2 ml 96-deep-well plates + AeraSeal film and grown at 37 °C, 900 r.p.m. After 2 h the growths were diluted (*E. coli* DH10B/Marionette-Clo 1:500; *E. coli* BL21/Marionette-Pro 1:2,000; *E. coli* MG1655/Marionette-Wild 1:5,000) into prewarmed LB + inducer where necessary in 2 ml 96-deep-well plates. 100 µl of this culture was immediately transferred to 300 µl 96-well black-walled optical bottom plates (Thermo Scientific Nunc, Waltham, MA, USA) sealed with a BreathEasy film (Sigma-Aldrich, St. Louis, MO, USA), and grown in a Synergy H1 plate reader (BioTek, Winooski, VT, USA) at 37 °C, 1,000 r.p.m. OD_{600} was measured every 20 min over 12 h of growth. OD_{600} readings were also taken from wells containing media with no cells, and for each time point, readings from such wells were subtracted from the appropriate sample measurements to remove background. OD_{600} values were converted to equivalent 1 cm path-length measurements using a standard curve.

Calculation of growth rates. To calculate doubling times, the last measurement with $OD_{600} < 0.1$ and the first measurement with $OD_{600} > 0.4$ were identified, and the doubling time was calculated assuming exponential growth between those two points. Doubling time is calculated as elapsed time (in minutes) divided by the number of doublings that occurred in that time ($\log_2$[final OD_{600}/initial OD_{600}]).

Determination of pellet and supernatant inducer concentrations. Single colonies of Marionette-Wild were inoculated into 1 ml LB in 2 ml 96-deep-well plates (USA Scientific, Orlando, FL, USA) sealed with an AeraSeal film (Excel Scientific, Victorville, CA, USA) and grown at 37 °C, 900 r.p.m. overnight in a Multitron Pro shaker incubator (INFORS HT, Bottmingen, Switzerland). The overnight growths were diluted 1:200 into 1 ml LB in 2 ml 96-deep-well plates + AeraSeal film and grown at 37 °C, 900 r.p.m. After 2 h the growths were diluted 1:5,000 into prewarmed LB + inducers in 2-ml 96-deep-well plates + AeraSeal film and grown at 37 °C, 900 r.p.m. for 5 h. After growth, 1 ml of culture sample was pelleted at 5,000 *g*, 25 °C, 10 min. The supernatant was mixed 1:1 with methanol and saved for analysis. The pellet was resuspended in 50 µl methanol and then 50 µl LB, shaken for 5 min, and centrifuged at 21,000 *g*, 25 °C, 10 min, and the supernatant was collected for analysis. Samples were then analyzed via liquid chromatography coupled to mass spectrometry (LC–MS) using a 1260 Binary Pump (Agilent, Santa Clara, CA,USA) in low dwell volume mode, an Agilent column oven heated to 40 °C, and a 6420 Triple Quadrupole Mass Spectrometer with an ESI source (Agilent, Santa Clara, CA, USA). 5 µl of sample was injected onto a 2.6 µm particle size, 50 × 2.1 mm i.d., Kinetix Core-Shell F5 column (Phenomenex, Torrance, CA, USA) and analyzed using the following method: 0.0–0.5 min at 0% (vol/vol) acetonitrile, 0.5–5.0 min at 0 to 90% (vol/vol) acetonitrile gradient, and 5.0–5.5 min at 90% (vol/vol) acetonitrile, with 1.0 min of re-equilibration. Flow rate was maintained at 0.6 ml/min. Mobile phases were (1) acetonitrile with 0.1% (vol/vol) formic acid and 0.1% (vol/vol) H_2O and (2) H_2O with 0.1% (vol/vol) formic acid. The ESI source was run with a desolvation temperature of 350 °C, a drying gas flow rate of 12 L/minute, a nebulizer pressure of 15 psi, and a capillary voltage of 4,000 V. The MS was run in Dynamic MRM mode using Unit resolution for both precursor and product quadrupoles, 2 min delta retention time for dynamic acquisition, positive ionization mode, and cell accelerator voltage of 4. Compound specific parameters were optimized using the Agilent Optimizer software. Those optimized parameters are as follows: DAPG (precursor mass to charge in Daltons per electron volt) 211.1 *m/z* > (product mass to charge in Daltons per electron volt) 193.0 *m/z*, CE (collision energy in electron volts) 12 eV, FV (fragmentation voltage) 96 V, RT (retention time) 3.45 min; Cuma 165.1 *m/z* > 79.2 *m/z*, CE 16 eV, FV 77 V, RT 3.4 min; OC6 214.1 *m/z* > 102.1 *m/z*, CE 8 eV, FV 81 V, RT 2.10 min; Van 169.0 *m/z* > 93.1 *m/z*, CE 12 eV, FV 81 V, RT 2.15 min; IPTG 221.1 *m/z* > 89.1 *m/z*, CE 8 eV, FV 66 V, RT 0.75 min; aTc 427.2 *m/z* > 410.1 *m/z*, CE 12 eV, FV 50 V, RT 3.10 min; Ara 150.1 *m/z* > 73.2 *m/z*, CE 12 eV, FV 65 V, RT 0.50 min; Cho 104.1 *m/z* > 60.2 *m/z*, CE 16 eV, FV 108 V, RT 0.50 min; Nar 273.1 *m/z* > 153.0 *m/z*, CE 20 eV, FV 118 V, RT 3.30 min; DHBA 155.0 *m/z* > 93.1 *m/z*, CE 12 eV, FV 71 V, RT 0.95 min; Sal 139.0 *m/z* > 121.0 *m/z*, CE 12 eV, FV 56 V, RT 2.50 min; OHC14 328.2 *m/z* > 102.1 *m/z*, CE 8 eV, FV 81 V, RT 4.00 min. Post-analysis data processing was done using MassHunter Quantitiative Analysis software (Agilent, Santa Clara, CA –USA) to extract MRM chromatograms and integrate peaks. Area under the curve was converted to raw concentrations using a standard curve. For the pellet, raw concentrations were converted to the final concentrations by assuming a total cell volume of 3.4 µl per OD_{600} per ml (ref. [58]) and a 2 µl "dead volume" of supernatant associated with the pellet.

Evolutionary stability (passaging with reporters). Marionette-Wild was transformed with each of the 12 reporter plasmids and plated on LB-agar. Single colonies were inoculated into 1 ml LB + antibiotics in 2 ml 96-deep-well plates (USA Scientific, Orlando, FL - USA) sealed with an AeraSeal film (Excel Scientific, Victorville, CA - USA) and grown at 37 °C, 900 r.p.m. overnight in a Multitron Pro shaker incubator (INFORS HT, Bottmingen, Switzerland). The overnight growths were diluted 1:200 into 1 ml LB + antibiotics in 2 ml 96-deepwell plates + AeraSeal film and grown at 37 °C, 900 r.p.m. After 2 h, the growths were diluted 1:5,000 into prewarmed LB + antibiotics + inducer where necessary in 2 ml 96-deep-well plates + AeraSeal film and grown at 37 °C, 900 r.p.m. for 5 h. After growth, 20 µl of culture sample was diluted into 180 µl PBS + 200 µg/ml kanamycin and assayed by cytometry as described in 'Cytometry analysis'. Uninduced cultures were allowed to grow at 37 °C overnight (17 h), and the process (beginning with the 1:200 dilution of the overnight culture) was repeated each day for 14 d.

Evolutionary stability (passaging on plates). A single colony of Marionette-Wild was inoculated into 1 ml LB in 2 ml 96-deep-well plates (USA Scientific, Orlando, FL, USA) sealed with an AeraSeal film (Excel Scientific, Victorville, CA, USA) and grown at 37 °C, 900 r.p.m. for 12 h in a Multitron Pro shaker incubator (INFORS HT, Bottmingen, Switzerland). The overnight growth was streaked onto LB-agar plates and grown at 37 °C for 12 h. This process was repeated nine times. The final culture was made chemically competent as described in 'Chemical transformation', transformed with relevant reporter plasmids, and assayed with cytometry as described in 'Cytometry analysis'.

Evolutionary stability (passaging in liquid culture). A single colony of Marionette-Wild was inoculated into 1 ml LB in 2 ml 96-deep-well plates (USA Scientific, Orlando, FL, USA) sealed with an AeraSeal film (Excel Scientific, Victorville, CA, USA) and grown at 37 °C, 900 r.p.m. for 12 h in a Multitron Pro shaker incubator (INFORS HT, Bottmingen, Switzerland). The overnight growths were diluted 1:200 into 1 ml LB in 2 ml 96-deep-well plates + AeraSeal film and grown at 37 °C, 900 r.p.m. After 2 h, the growths were diluted 1:5,000 into prewarmed LB in 2 ml 96-deep-well plates + AeraSeal film and grown at 37 °C, 900 r.p.m. for 22 h. This process was repeated nine times. The final culture was made chemically competent as described in 'Chemical transformation', transformed with relevant reporter plasmids, and assayed with cytometry as described in 'Cytometry analysis'.

Lycopene production optimization. Marionette Wild was transformed with a p15A plasmid containing the lycopene biosynthetic pathway under the control of five sensors. Single colonies were inoculated into 1 ml LB in 2-ml 96-deep-well plates (USA Scientific, Orlando, FL, USA) sealed with an AeraSeal film (Excel Scientific, Victorville, CA, USA) and grown at 37 °C, 900 r.p.m. overnight in a Multitron Pro shaker incubator (INFORS HT, Bottmingen, Switzerland). The overnight cultures were diluted 1:500 into 1 ml LB in 2 ml 96-deep-well plates + AeraSeal film and grown at 37 °C, 900 r.p.m. After 2 h the cultures were diluted 1:200 into 600 µl prewarmed LB + inducer where necessary in 2 ml 96-deep-well plates. After 20 h of growth, cultures were pelleted at 4,500 *g*, 4 °C, 10 min. Pellets were resuspended in 100 µl PBS and then 500 µl acetone. Resuspensions were shaken in an incubating microplate shaker (VWR, Radnor, PA, USA) at 55 °C, 400 r.p.m., 20 min. Extractions were pelleted at 4,500 *g*, 4 °C, 10 min, and 150 µl was transferred to a flat-bottomed polypropylene 96-well plate (Millipore Sigma, St. Louis, MO, USA). Samples were measured on a Synergy H1 plate reader (BioTek, Winooski, VT, USA) at an absorbance wavelength of 470 nm. Absorbance values were converted to lycopene production using a standard curve from purified lycopene (Millipore Sigma, St. Louis, MO, USA). The grid search algorithm was performed by subdividing the full (five-dimensional) volume into 32 smaller subvolumes, each defined by 32 vertices. A score was assigned to each subvolume by summing the lycopene titer for all 32 vertices. The subvolume with the highest score was then set as the full volume for the next iteration. The first iteration tested all combinations of: 0, 7.2, and 200 nM aTc; 0, 1.4, and 10,000 nM OC6; 0, 5.1, and 500 µM Cuma; 0, 26, and 1,000 µM IPTG; and 0, 0.4, and 25 µM DAPG. The second iteration tested all combinations of: 0, 5.0, and 7.2 nM aTc; 0, 0.4, and 1.4 nM OC6; 5.1, 12, and 500 µM Cuma; 26, 71, and 1,000 µM IPTG; and 0.4, 1.0, and 25 µM DAPG. The third iteration tested all combinations of: 0, 3.9, and 5.0 nM aTc; 0, 0.2, and 0.4 nM OC6; 12, 19, and 500 µM Cuma; 26, 42, and 71 µM IPTG; and 1.0, 1.6, and 25 µM DAPG. The fourth iteration tested all combinations of: 0, 3.3, and 3.9 nM aTc; 0, 0.1, and 0.2 nM OC6; 19, 28, and 500 µM Cuma; 26, 33, and 42 µM IPTG; and 1.6, 2.4, and 25 µM DAPG.

Computational methods. To parameterize the response function, error minimization was performed using the Solver function in Excel software (Microsoft, Redmond, WA - USA). Equation (1) was entered with y_{min}, y_{max}, K, and n as tunable parameters, x (inducer concentration) as the independent variable and y as the output. For each x, the error between the measured RPU value and the output of the function was determined. The total error was determined by summing the normalized square of the error ([y-measured RPU value]2/y) for each x. The Solver function minimized the total error by tuning y_{min}, y_{max}, K, and n.

Libraries for the selection of LacI and AraC used partial randomization of amino acids. Previous literature informed decisions regarding the desirability of including each possible residue in the library. The CASTER 2.0 tool[59] was used to design degenerate oligodeoxynucleotides that can sample the desired amino acids while limiting undesired amino acids and stop codons.

The initial RBS for each regulator in the Marionette cluster was designed using the RBS Calculator v1.1 (ref. [60]) using the *E. coli* DH10B 16S rRNA setting. The "pre-sequence" included the last 100 bp of the upstream CDS (where appropriate) as well as any scars used in Golden Gate assembly. The "Target Translation Initiation Rates" were chosen based on the RBS strength of the plasmid-based sensor (found using the "Reverse Engineer RBSs" tool) and compensating for the reduction in copy number and increase in promoter strength associated with the move to the genomic system. After initial assembly and testing, a decision was made based on the appearance of the response function as to whether the RBS was too low, too high, or good. Low RBSs were rationally mutated to more closely resemble the consensus Shine-Dalgarno sequence (TAAGGAGGT) while high RBSs were rationally mutated to less closely resemble the consensus Shine-Dalgarno sequence. All rationally designed RBS variants were checked using the "Reverse Engineer RBSs" to guard against making dramatic, unanticipated changes to the RBS strength. Rational mutations are noted in Supplementary Table 2. Translation Initiation Rate (in arbitrary units) for each RBS variant is provided in Supplementary Fig. 10.

The RBS for each gene in the lycopene pathway was designed using the RBS Calculator v1.1 (ref. [60]) using the *E. coli* DH10B 16S rRNA setting. The "Target Translation Initiation Rates" was set to maximize strength.

Reporting Summary. Further information on research design is available in the Nature Research Reporting Summary linked to this article.

Data availability

Additional data supporting this study are available from the corresponding author upon reasonable request. The sequences of the following plasmids and strains are provided in GenBank: pAJM.711 (P_{PhlF}-YFP) MH101715; pAJM.712 (P_{CymRC}-YFP) MH101716; pAJM.713 (P_{LuxB}-YFP) MH101717; pAJM.714 (P_{VanCC}-YFP) MH101718; pAJM.715 (P_{Tac}-YFP) MH101719; pAJM.717 (P_{Tet}-YFP) MH101720; pAJM.716 (P_{BAD}-YFP) MH101721; pAJM.718 (P_{BetI}-YFP) MH101722; pAJM.719 (P_{Ttg}-YFP) MH101723; pAJM.1459 (P_{3B5C}-YFP) MH101724; pAJM.721 (P_{SalTTC}-YFP) MH101725; pAJM.944 (P_{Cin}-YFP) MH101726; pAJM.847 ($PhlF^{AM}$ + P_{PhlF}-YFP) MH101727; pAJM.657 ($CymR^{AM}$ + P_{CymRC}-YFP) MH101728; pAJM.474 (LuxR + P_{LuxB}-YFP) MH101729; pAJM.773 ($VanR^{AM}$ + P_{VanCC}-YFP) MH101730; pAJM.336 ($LacI^{AM}$ + P_{Tac}-YFP) MH101731; pAJM.011 (TetR + P_{Tet}-YFP) MH101732; pAJM.677 ($AraC^{AM}$ + AraE + P_{BAD}-YFP) MH101733; pAJM.683 ($BetI^{AM}$ + P_{BetI}-YFP) MH101734; pAJM.661 ($TtgR^{AM}$ + P_{Ttg}-YFP) MH101735; pAJM.690 ($PcaU^{AM}$ + P_{3B5B}-YFP) MH101736; pAJM.771 ($NahR^{AM}$ + P_{SalTTC}-YFP) MH101737; pAJM.1642 ($CinR^{AM}$ + P_{Cin}-YFP) MH101738; pAJM.884 ($AcuR^{AM}$ + P_{Acu}-YFP) MH101739; pAJM.969 ($MphR^{AM}$ + EryR + P_{Mph}-YFP) MH101740. The following plasmids and strains can be acquired from Addgene: pAJM.711 (P_{PhlF}-YFP) 108512; pAJM.712 (P_{CymRC}-YFP) 108513; pAJM.713 (P_{LuxB}-YFP) 108514; pAJM.714 (P_{VanCC}-YFP) 108515; pAJM.715 (P_{Tac}-YFP) 108516; pAJM.717 (P_{Tet}-YFP) 108517; pAJM.716 (P_{BAD}-YFP) 108518; pAJM.718 (P_{BetI}-YFP) 108519; pAJM.719 (P_{Ttg}-YFP) 108520; pAJM.1459 (P_{3B5C}-YFP) 108521; pAJM.721 (P_{SalTTC}-YFP) 108522; pAJM.944 (P_{Cin}-YFP) 108523; pAJM.847 ($PhlF^{AM}$ + P_{PhlF}-YFP) 108524; pAJM.657 ($CymR^{AM}$ + P_{CymRC}-YFP) 108525; pAJM.474 (LuxR + P_{LuxB}-YFP) 108526; pAJM.773 ($VanR^{AM}$ + P_{VanCC}-YFP) 108527; pAJM.336 ($LacI^{AM}$ + P_{Tac}-YFP) 108528; pAJM.011 (TetR + P_{Tet}-YFP) 108529; pAJM.677 ($AraC^{AM}$ + AraE + P_{BAD}-YFP) 108530; pAJM.683 ($BetI^{AM}$ + P_{BetI}-YFP) 108531; pAJM.661 ($TtgR^{AM}$ + P_{Ttg}-YFP) 108532; pAJM.690 ($PcaU^{AM}$ + P_{3B5B}-YFP) 108533; pAJM.771 ($NahR^{AM}$ + P_{SalTTC}-YFP) 108534; pAJM.1642 ($CinR^{AM}$ + P_{Cin}-YFP) 108535; pAJM.884 ($AcuR^{AM}$ + P_{Acu}-YFP) 108536; pAJM.969 ($MphR^{AM}$ + EryR + P_{Mph}-YFP) 108537; sAJM.1504 (Marionette-Clo) 108251; sAJM.1505 (Marionette-Pro) 108253; sAJM.1506 (Marionette-Wild) 108254.

References

51. Blattner, F. R. et al. The complete genome sequence of *Escherichia coli* K-12. *Science* **277**, 1453–1462 (1997).
52. Kittleson, J. T., Cheung, S. & Anderson, J. C. Rapid optimization of gene dosage in *E. coli* using DIAL strains. *J. Biol. Eng.* **5**, 10 (2011).
53. Green, R. & Rogers, E. J. Transformation of chemically competent *E. coli*. *Methods Enzymol.* **529**, 329–336 (2013).
54. Meyer, A. J., Ellefson, J. W. & Ellington, A. D. Library generation by gene shuffling. *Curr. Protoc. Mol. Biol.* **105**, 15.12.1–15.12.7 (2014).
55. Datsenko, K. A. & Wanner, B. L. One-step inactivation of chromosomal genes in *Escherichia coli* K-12 using PCR products. *Proc. Natl. Acad. Sci. USA* **97**, 6640–6645 (2000).
56. Thomason, L. C., Costantino, N. & Court, D. L. *E. coli* genome manipulation by P1 transduction. *Curr. Protoc. Mol. Biol.* **79**, 1.17.1–1.17.8 (2007).
57. Li, G. W., Burkhardt, D., Gross, C. & Weissman, J. S. Quantifying absolute protein synthesis rates reveals principles underlying allocation of cellular resources. *Cell* **157**, 624–635 (2014).
58. Volkmer, B. & Heinemann, M. Condition-dependent cell volume and concentration of *Escherichia coli* to facilitate data conversion for systems biology modeling. *PLoS One* **6**, e23126 (2011).
59. Reetz, M. T. & Carballeira, J. D. Iterative saturation mutagenesis (ISM) for rapid directed evolution of functional enzymes. *Nat. Protoc.* **2**, 891–903 (2007).
60. Salis, H. M., Mirsky, E. A. & Voigt, C. A. Automated design of synthetic ribosome binding sites to control protein expression. *Nat. Biotechnol.* **27**, 946–950 (2009).

nature research

Corresponding author(s): Christopher Voigt

Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see Authors & Referees and the Editorial Policy Checklist.

Statistical parameters

When statistical analyses are reported, confirm that the following items are present in the relevant location (e.g. figure legend, table legend, main text, or Methods section).

n/a	Confirmed	
☐	☒	The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
☐	☒	An indication of whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☒	☐	The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☒	☐	A description of all covariates tested
☒	☐	A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☐	☒	A full description of the statistics including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☒	☐	For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted *Give P values as exact values whenever suitable.*
☒	☐	For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☒	☐	For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☒	☐	Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated
☐	☒	Clearly defined error bars *State explicitly what error bars represent (e.g. SD, SE, CI)*

Our web collection on statistics for biologists may be useful.

Software and code

Policy information about availability of computer code

Data collection	*Provide a description of all commercial, open source and custom code used to collect the data in this study, specifying the version used OR state that no software was used.*
Data analysis	CutAdapt, Bowtie, BWA, SAM Tools, HTSeq, edgeR, Bedtools, Agilent Optimizer, MassHunter Quantitiative Analysis, Excel, CASTER, RBS Calculator, Geneious.

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors/reviewers upon request. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

Key plasmids and strains are available from Addgene. Key sequences are provided on GenBank. All strains and plasmids are available upon request.

Field-specific reporting

Please select the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences ☐ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/authors/policies/ReportingSummary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	Most experiments are performed in triplicate on separate days. This is considered best practices in the field.
Data exclusions	No data were excluded from analysis unless there was a clear machine malfunction, in which case all data from the entire experiment were excluded.
Replication	All experimental claims have been tested three times on different days. The triplicate results closely matched the results from any given day.
Randomization	Randomization was not done as it was not relevant to the study.
Blinding	Blinding was not done as it was not relevant to the study.

Reporting for specific materials, systems and methods

Materials & experimental systems

n/a	Involved in the study
☒	☐ Unique biological materials
☒	☐ Antibodies
☒	☐ Eukaryotic cell lines
☒	☐ Palaeontology
☒	☐ Animals and other organisms
☒	☐ Human research participants

Methods

n/a	Involved in the study
☒	☐ ChIP-seq
☐	☒ Flow cytometry
☒	☐ MRI-based neuroimaging

Flow Cytometry

Plots

Confirm that:

☐ The axis labels state the marker and fluorochrome used (e.g. CD4-FITC).

☐ The axis scales are clearly visible. Include numbers along axes only for bottom left plot of group (a 'group' is an analysis of identical markers).

☐ All plots are contour plots with outliers or pseudocolor plots.

☐ A numerical value for number of cells or percentage (with statistics) is provided.

Methodology

Sample preparation	Bacterial cells were diluted into PBS containing Kanamycin to halt translation.
Instrument	BD LSR Fortessa
Software	FACS Diva software for collection; FlowJo for data analysis.
Cell population abundance	Typical samples contained 10,000 or more cells.

Gating strategy

Cells in mid-log phase were gated: 1,000-10,000 SSA and 1,000-10,000 FSH.
Cells in stationary phase were gated: 500-5,000 SSA and 500-5,000 FSH.

☐ Tick this box to confirm that a figure exemplifying the gating strategy is provided in the Supplementary Information.