



# Precision design of stable genetic circuits carried in highly-insulated *E. coli* genomic landing pads

Yongjin Park<sup>1</sup>, Amin Espah Borujeni<sup>1</sup>, Thomas E Gorochoowski<sup>1,2</sup> , Jonghyeon Shin<sup>1</sup> & Christopher A Voigt<sup>1,2,\*</sup> 

## Abstract

Genetic circuits have many applications, from guiding living therapeutics to ordering process in a bioreactor, but to be useful they have to be genetically stable and not hinder the host. Encoding circuits in the genome reduces burden, but this decreases performance and can interfere with native transcription. We have designed genomic landing pads in *Escherichia coli* at high-expression sites, flanked by ultrastrong double terminators. DNA payloads >8 kb are targeted to the landing pads using phage integrases. One landing pad is dedicated to carrying a sensor array, and two are used to carry genetic circuits. NOT/NOR gates based on repressors are optimized for the genome and characterized in the landing pads. These data are used, in conjunction with design automation software (Cello 2.0), to design circuits that perform quantitatively as predicted. These circuits require fourfold less RNA polymerase than when carried on a plasmid and are stable for weeks in a *recA*<sup>+</sup> strain without selection. This approach enables the design of synthetic regulatory networks to guide cells in environments or for applications where plasmid use is infeasible.

**Keywords** gene regulatory network; genetic circuit design automation; genome editing; synthetic biology; systems biology

**Subject Categories** Biotechnology & Synthetic Biology; Methods & Resources

**DOI** 10.15252/msb.20209584 | Received 18 March 2020 | Revised 7 July 2020 | Accepted 22 July 2020

**Mol Syst Biol.** (2020) **16**: e9584

## Introduction

Cells use regulatory networks, encoded in their genomes, to determine which genes need to be expressed based on cellular needs or to adapt to the environment (McAdams & Shapiro, 1995). Engineers reconstruct such networks as “genetic circuits” by connecting regulatory proteins to produce a desired computational operation (Elowitz & Leibler, 2000; Atkinson *et al.*, 2003; Basu *et al.*, 2004; Stricker *et al.*, 2008; Wang *et al.*, 2011; Moser *et al.*, 2012; Din *et al.*, 2016; Fernandez-Rodriguez *et al.*, 2017). The majority of these circuits

have been characterized using plasmids. One motivation is that circuit optimization requires genetic “tinkering” and making mutations to a plasmid is simpler (Gardner *et al.*, 2000; Tabor *et al.*, 2009; Fernandez-Rodriguez *et al.*, 2017). Another is that plasmids amplify regulator expression, making them easier to connect (Brophy & Voigt, 2014). Finally, they increase the expression of fluorescent reporters used to measure circuit function (Gardner *et al.*, 2000; Gorochoowski *et al.*, 2017). However, carrying a circuit on a plasmid also has disadvantages. They can lead to instability, cell-to-cell heterogeneity, and metabolic burden (Summers & Sherratt, 1984; Chiang & Bremer, 1988; Summers, 1991; Stoebel *et al.*, 2008; Kittleson *et al.*, 2011; Gyorgy *et al.*, 2015; Borkowski *et al.*, 2016; Wang *et al.*, 2016b). This can lead to evolutionary forces breaking a circuit through plasmid loss or mutagenesis to the plasmid or genome (to reduce the copy number; Mayo *et al.*, 2006; Stoebel *et al.*, 2008; Klumpp *et al.*, 2009; Sleight *et al.*, 2010; Chen *et al.*, 2013; Sleight & Sauro, 2013; Fernandez-Rodriguez *et al.*, 2015; Gyorgy *et al.*, 2015; Ceroni *et al.*, 2018; Liu *et al.*, 2018; Moser *et al.*, 2018, 2012). Because of these issues, it is standard practice in industrial biotechnology to introduce recombinant DNA in the genome, particularly if selective pressure from antibiotics is impossible or cost-prohibitive (Singh *et al.*, 2011; Isabella *et al.*, 2018).

Circuits have been designed for the genome for industrial and biomedical applications. For bio-production, circuits that switch on metabolic pathways after biomass accumulation have been encoded in the genome (Gupta *et al.*, 2017). For agricultural applications, sensors have been encoded in the genome to improve stability in soil (Brophy *et al.*, 2018). Bacteria engineered to be therapeutics have to be able to function in the human body without antibiotics. To this end, circuits have been integrated into the genome to identify environmental niches, respond to inflammation, induce expression in response to a consumed pharmaceutical, or kill the bacterium (Kong *et al.*, 2008; Kotula *et al.*, 2014; Danino *et al.*, 2015; Mimee *et al.*, 2015; Lee *et al.*, 2016; Riglar *et al.*, 2017; Stirling *et al.*, 2017; Isabella *et al.*, 2018; Chowdhury *et al.*, 2019; Naydich *et al.*, 2019). A genome-encoded memory switch was shown to be functional for 6 months in bacteria colonizing the gut of a mouse (Riglar *et al.*, 2017). Note that plasmids can also be stabilized for therapeutic using addiction systems (Fedorec *et al.*, 2019). For eukaryotes, where plasmids often are not available or unreliable (Davidsohn

<sup>1</sup> Synthetic Biology Center, Department of Biological Engineering, Massachusetts Institute of Technology, Cambridge, MA, USA

<sup>2</sup> Broad Institute of MIT and Harvard, Cambridge, MA, USA

\*Corresponding author. Tel: +1 617 324 4851; E-mail: cavoigt@gmail.com

*et al.*, 2015), it is standard to integrate circuits into the chromosome (Antunes *et al.*, 2011; Duportet *et al.*, 2014; Roybal *et al.*, 2016; Gander *et al.*, 2017; Gaidukov *et al.*, 2018; Chang *et al.*, 2019; Jusiak *et al.*, 2019).

These circuits were designed by hand to suit the needs of the application. Design automation software seeks to computationally map a desired circuit function to a DNA sequence (Cai *et al.*, 2007; Brophy & Voigt, 2014; Vaidyanathan *et al.*, 2015; Nielsen *et al.*, 2016; Guiziou *et al.*, 2018). Cello does this by having the user specify the desired operation (in Verilog) and the sensors to serve as inputs (Nielsen *et al.*, 2016). They specify the organism, genetic location of the circuit, and gate technology by selecting a user constraint file (UCF). Using this information, Cello builds the desired DNA sequence and makes quantitative predictions of the circuit performance in each state. The first UCF (Eco1C1G1T1) is for *Escherichia coli* DH10 $\beta$ , the circuit is carried on a p15a plasmid, and the gate technology is based on NOT/NOR gates using orthogonal TetR-family repressors (Stanton *et al.*, 2014). Re-designing a circuit for a new context is as simple as selecting a new UCF.

Cello requires that the gates be transcriptional; in other words, their inputs and outputs have to be promoters. To make a UCF, the gate response functions (how the output changes as a function of the input) have to be measured at the same site where the circuit will be carried. Characterizing the response functions in relative promoter units (RPU) simplifies the prediction of how they can be connected in series to build a circuit (Kelly *et al.*, 2009; Nielsen *et al.*, 2016). Design automation is only as successful as the quality of the gates; they must produce the same response in the context of different circuits. Achieving this has required insulating the gates through the use of strong terminators, long promoters, and ribozymes (Davis *et al.*, 2011; Lou *et al.*, 2012; Cambray *et al.*, 2013; Chen *et al.*, 2013; Mutalik *et al.*, 2013; Brophy & Voigt, 2014; Nielsen *et al.*, 2016). Carrying a circuit in the genome introduces additional modes of potential interference, including RNA polymerases (RNAPs) from neighboring regions and changes to the macrostructure (Guo & Adhya, 2007; Vora *et al.*, 2009; Lasa *et al.*, 2011; Mitschke *et al.*, 2011; Chong *et al.*, 2014; Lybecker *et al.*, 2014; Wade & Grainger, 2014; Brophy & Voigt, 2016; Dorman & Dorman, 2016; Yeung *et al.*, 2017). In addition, different regions of the genome vary in effective copy number, with the highest being near the origin, so a circuit encoded at a locus may behave differently at another (Chandler & Pritchard, 1975; Schmid & Roth, 1987; Sousa *et al.*, 1997; Rocha, 2008; Block *et al.*, 2012; Slager & Veening, 2016). Collectively, these effects can cause the expression of a recombinant gene to vary by up to 300-fold depending on where in the genome it is encoded (Bryant *et al.*, 2014; Scholz *et al.*, 2019).

This work describes the reliable design of genetic circuits for highly insulated “landing pads” in the *E. coli* MG1655 genome. The landing pad positions were identified using random transposon mutagenesis to identify high-expression positions and then confirming that native gene expression is not impacted. The landing pads contain orthogonal phage integration sites so that they can be independently targeted with high efficiency (Datsenko & Wanner, 2000; Choi & Schweitzer, 2006; Sharan *et al.*, 2009; Wang *et al.*, 2009, 2016a; Kuhlman & Cox, 2010; Lambowitz & Zimmerly, 2011; Eneyart *et al.*, 2013; Esvelt & Wang, 2013; Santos *et al.*, 2013; St-Pierre *et al.*, 2013; Gu *et al.*, 2015; Jiang *et al.*, 2015; Pyne *et al.*, 2015;

Bassalo *et al.*, 2016). The landing pads are flanked with new ultra-strong double terminators to block transcription into or out of the sites. A NOR gate architecture is developed for the genome, and the response functions for 6 TetR-family repressors are characterized in a landing pad. These data are used to build a UCF (Eco2C1G3T1) for Cello. Design automation is used to create genome-encoded circuits, and they are found to function quantitatively as predicted. The division of regulatory sensors and circuitry across defined positions in the genome represents a step toward the organized design of synthetic regulatory networks for genome-scale engineering projects.

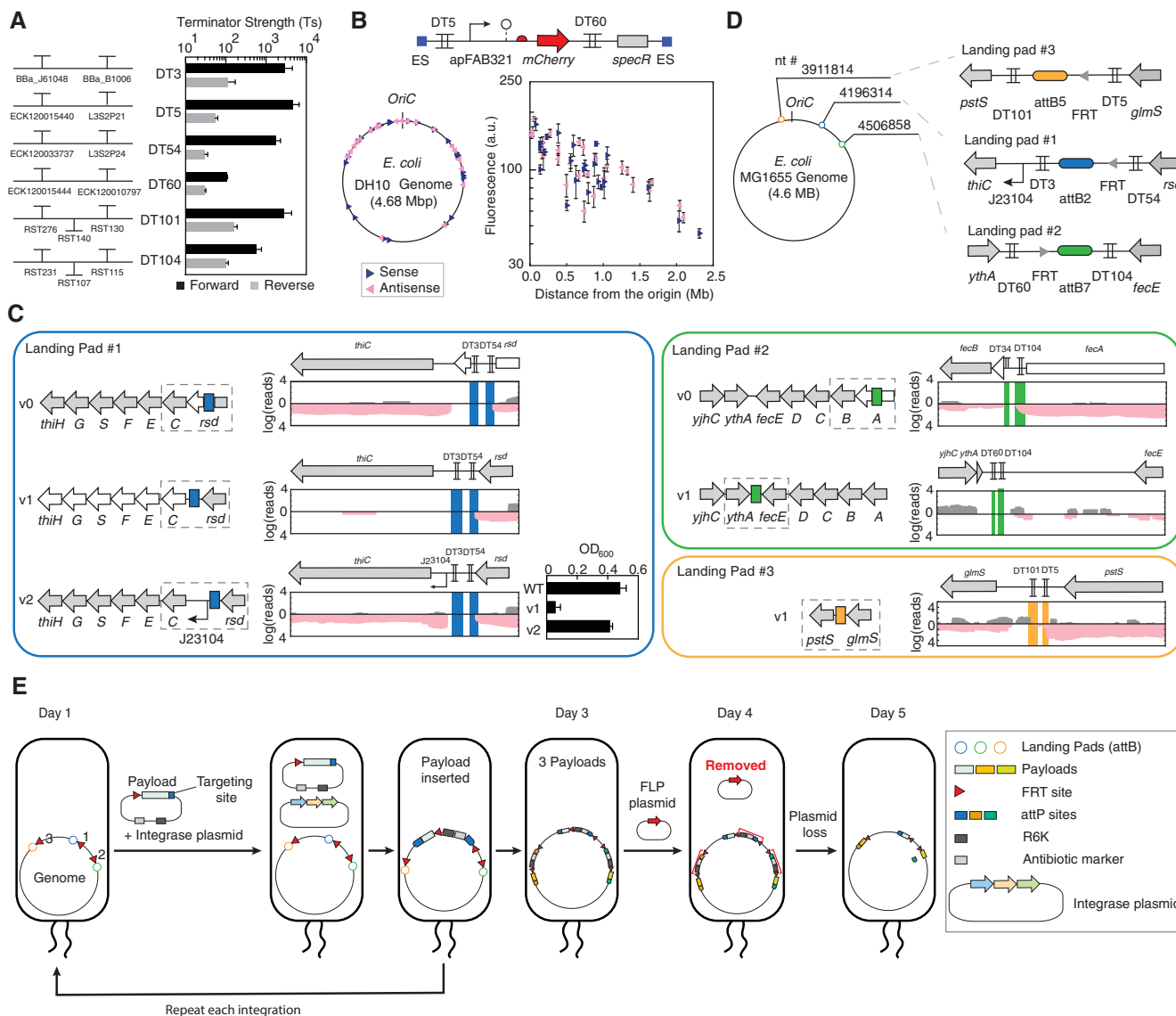
## Results

### Genetic landing pad construction and characterization

Natural terminators are often weak, and pervasive transcription is common in both the sense and antisense directions of the genome (Lasa *et al.*, 2011; Cambray *et al.*, 2013; Chen *et al.*, 2013; Lybecker *et al.*, 2014; Wade & Grainger, 2014). If allowed to enter the circuit DNA, genomic RNAP flux can cause the circuit to malfunction (Lee *et al.*, 2016). The opposite can also be problematic, where the RNAP flux from the circuit can exit and cause genes encoded in the genome to be expressed incorrectly. This effect is worsened by the fact that synthetic circuits typically produce large swings in RNAP flux. To address these issues, we designed very strong terminators to flank the landing pads and block transcription both into and out of the genetic circuit.

The terminator strength ( $T_S$ ) is a metric that captures how much gene expression changes before and after the terminator (see Materials and Methods for definition; Chen *et al.*, 2013). We previously characterized a large library of individual terminators, the majority of which are strong with an average  $T_S = 30$  (97% of transcription is blocked). Some terminators block RNAP flux in both directions (bidirectional), but the average strength in the weak direction is low ( $T_S = 6$  or 17% transcriptional readthrough). To better insulate the landing pads, a stronger set of terminators was designed by concatenating two terminators in series (Appendix Table S1, Appendix Fig S1). The single terminators were sourced from large libraries of characterized parts (Cambray *et al.*, 2013; Chen *et al.*, 2013), the terminator prediction software RNIE (Gardner *et al.*, 2011), and *E. coli* MG1655 transcriptome data (Materials and Methods). This led to a set of 93 double terminators, including variations in spacers and ordering. Then, the strength of each double terminator was measured using an assay based on placing the terminator between red and green fluorescent protein genes downstream of an inducible promoter (Materials and Methods; Appendix Fig S14; Chen *et al.*, 2013). Seventeen were selected that have  $T_S = 105$  to 4744 and sufficiently diverse sequences to avoid homologous recombination (Appendix Table S1). Note that the double terminators are relatively small (< 170 bp). Six bidirectional double terminators ( $T_S > 105$  forward and  $T_S > 25$  reverse) were selected to flank the landing pads (Fig 1A).

The next step was to identify regions of the genome where a landing pad could achieve high levels of expression and not impact growth. Following an approach used by Freddolino and co-workers (Scholz *et al.*, 2019), Tn5 transposon mutagenesis was used to



**Figure 1. Design of insulated genetic landing pads.**

- A The six bidirectional double terminators are shown. The terminator strength is shown, calculated as described in the Materials and Methods. Terminator sequences are provided in Appendix Table S1. The means of three experiments performed on different days are shown, and the error bars are the standard deviation of these measurements.
- B The transposon library screen in *E. coli* DH10β. The construct shown was randomly integrated into the genome (ES, end sequences). Sense and antisense insertions are denoted by the navy right and pink left triangles, respectively. Detail information (location, expression levels, and OD<sub>600</sub>) of the insertion locations is provided in Appendix Table S2. The means of three experiments performed on different days are shown, and the error bars are the standard deviation of these measurements.
- C The genomic impact of the landing pad locations. RNA-seq was performed; sense and antisense transcripts are shown in gray and pink, respectively. The dashed squares show the regions of the genome shown in the transcriptional profiles. Genes colored white are those for which we observed large changes upon the insertion of the landing pads. The growth (OD<sub>600</sub>) of *E. coli* MG1655 strains harboring Landing Pad #1 v1 and v2 in Thamine-free medium is shown. The means of three experiments performed on different days are shown, and the error bars are the standard deviation of these measurements.
- D Final selection of three Landing Pads.
- E Schematic showing the steps and time required to insert multiple payloads into the genome. A detailed protocol and the result of integration are provided in Appendix Note S1.

randomly insert a constitutive promoter driving mCherry expression in the genome (Fig 1B). To avoid the high expression being due to transcriptional activity from the genome, the expression cassette was flanked by a pair of bidirectional double terminators. *E. coli* DH10β was selected as a recipient strain due to its enhanced

tolerance for foreign DNA and high-throughput transposon library construction (Grant *et al*, 1990; Durfee *et al*, 2008). The transposon was randomly integrated into 50 sites, and the expression and growth impact were measured (Appendix Table S2). As observed previously, the highest expression occurs close to the origin, but

there was significant site-to-site variation even in this region (Fig 1b; Schmid & Roth, 1987; Rocha, 2008; Block *et al*, 2012; Bryant *et al*, 2014; Scholz *et al*, 2019). The directionality of the construct did not systematically impact expression. Three sites were initially selected for the landing pads that had high expression and no impact on growth. In addition, we sought sites that did not disrupt an essential gene or for which we identified putative regions within 10 kb where the landing pad could be moved to avoid impacting host gene expression.

After identifying the landing pads in *E. coli* DH10 $\beta$ , they were moved to *E. coli* MG1655 K-12 because it is more commonly used in industry (Xie *et al*, 2003; Sezonov *et al*, 2007). This was simplified by the observation that the relative expression levels between sites are similar when compared between *E. coli* DH10 $\beta$  and *E. coli* MG1655 K-12 (Appendix Fig S2). Three landing pads were designed based on orthogonal *att* sites so that different phage integrases could be used to direct recombinant DNA to a landing pad (Yang *et al*, 2014). To avoid off-target effects, we selected *att* sites that do not share sequence identity with the *E. coli* MG1655 genome (Materials and Methods). Landing Pads #1, #2, and #3 are based on *att* sites specific to Int2, Int7, and Int5, respectively. Finally, FRT sites were added to remove the antibiotic selection markers from each landing pad. After inserting the landing pads using lambda red recombineering, they are transferred to a clean genomic background using P1 transduction (Materials and Methods).

Landing Pad #1 (v0) was initially located within *rsd*, a repressor of sigma factor D involved in the gene expression in stationary phase (Fig 1D; Jishage & Ishihama, 1999). We were concerned about disrupting this gene, so the landing pad was moved downstream of the gene with what we thought would be enough room to avoid impacting the promoter of the next gene (*thiC*) (v1). However, this was found to completely abolish transcription of the *thi* operon, which is responsible for producing thiamine (Vander Horn *et al*, 1993; Xi *et al*, 2001; Leonardi *et al*, 2003). Indeed, while there was no growth defect in M9 supplemented with 0.4% glucose, casaminoacids, and thiamine, there was a severe growth defect when grown in media lacking thiamine (Fig 1C). Noting that in the wild-type genome this operon is constitutively expressed even when grown in media containing thiamine, we selected a medium-strength constitutive promoter (BBa\_J23104; Kelly *et al*, 2009) to be placed downstream of the landing pad. This was found to recover the transcription of the *thi* operon and overcome the growth defect (v2, Fig 1D). The movement of this landing pad and the insertion of the synthetic promoter did not impact the expression of gene expression level from this site (Appendix Fig S3).

Similar experiments were performed to analyze Landing Pads #2 and #3. Landing Pad #2 initially disrupted the first gene of the *fec* operon (Enz *et al*, 1995, 2003). Therefore, we moved it 3.9 kb downstream *fecE* and before the terminator after the *ythA* gene, which is oriented in the opposite direction (Fig 1D). This was found to not impact growth nor transcript levels, while yielding the same level of expression (Appendix Fig S3). The initial site for Landing Pad #3 was found to be only 6.7 kb apart from the well-established Tn7 transposase integrase site (Choi & Schweizer, 2006) that showed no impact on growth and had high level of gene expression (Segall-Shapiro *et al*, 2018). Note that the mCherry constitutive expression cassette yields very similar expression levels across all

three sites (< 10% difference), making them interchangeable in carrying constructs. When empty, all three landing pads are small (407, 386, and 901 bp) and show very strong insulation in both the sense and antisense directions (Fig 1D). The final *E. coli* MG1655 strain (YJP\_MKC173) contains the three landing pads and no selective markers.

## Genome engineering methodology

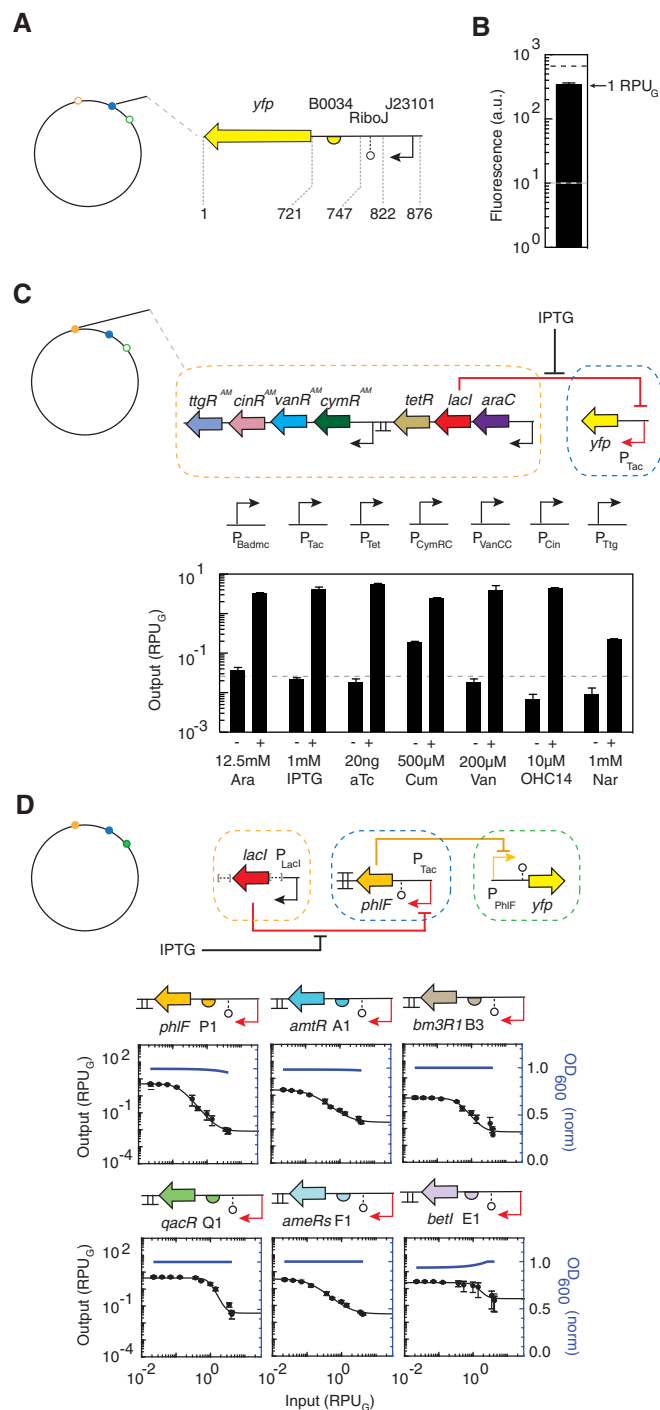
Our objective was to simplify the process of genome engineering so that it approaches the ease of plasmid manipulations (Fig 1E). The DNA payloads are cloned into three plasmids (pLYJP064, pLYJP066, and pLYJP070), each containing an *att* site directing it to a landing pad, an antibiotic marker and a single FRT site. A single plasmid (pLYJP053) constitutively expresses all three integrases. The payload and integrase plasmids all contain the R6K origin that can be amplified in a *pir*<sup>+</sup> strain, but cannot replicate in *E. coli* MG1655 (Shafferman *et al*, 1982; Metcalf *et al*, 1994). Electroporation is typically performed for the transformation steps, but with the addition of the origin of transfer (OriT) (Fu *et al*, 1991), the plasmids can also be delivered via conjugation by mixing a recipient strain and two donor strains (Materials and Methods). Each payload plasmid has a different antibiotic resistance, which can be removed in a single step by transforming a plasmid containing the FLP recombination, which can be subsequently cured. We attempted to co-transform multiple plasmids to integrate into different landing pads simultaneously, but this approach was unsuccessful. Appendix Note S1 contains a detailed protocol for performing serial integrations.

## RPU reference promoter in the genome

Cello designs circuits by connecting transcriptional sensors and circuits, whose signal carrier is defined as RNAP flux (Canton *et al*, 2008). In previous works we selected a plasmid-borne constitutive promoter (BBa\_J23101) for this purpose, which had been identified by others as a good reference promoter to obtain more reliable promoter measurements across laboratories (Canton *et al*, 2008; Kelly *et al*, 2009; Nielsen *et al*, 2016). We moved this promoter to the genome (Landing Pad #1) with a stronger RBS (BBa\_B0034) and measured its fluorescence (Fig 2A). This value is used to normalize the promoters of the sensors and gates such that they can be reported in relative promoter units (RPU<sub>G</sub>; Fig 2B). Note that correcting for the RBS strength and reporting RPU on a per DNA basis would result in the same value as that measured for the plasmid-borne RPU reference promoter. The value of RPU<sub>G</sub> in absolute units is estimated to be 0.067 RNAP/s (Materials and Methods).

## Sensor array in Landing Pad #3

To modularize the organization of synthetic regulatory networks, we assigned Landing Pad #3 to carry the sensors and the circuit to be divided between Landing Pads #1 and #2. The sensor array is composed of a set of 7 small molecular sensors. These sensors were selected from a larger set of 12 developed for the “Marionette” *E. coli* strains (Moon *et al*, 2012; Stanton *et al*, 2014; Meyer *et al*, 2019), with several removed because their repressors appear in our



gate library. Note that this work places the sensor array in a different genomic location and we made several genetic changes to the sensors to optimize their dynamic range in this context (Materials and Methods). The seven regulatory genes are organized as two constitutively transcribed operons (Fig 2C). The output promoter for each sensor was used to drive YFP expression from Landing Pad #1. These sensors produce a 12- to 640-fold induction (Fig 2C, Appendix Fig S5, Appendix Table S3) with low off states and no evidence of crosstalk (Fig 2c and Appendix Fig S5).

### Figure 2. Characterization of genetic sensors and gates in the genome.

The genomic landing pads are shown as open circles when empty and full when carrying a genetic design.

- A** The reference promoter and construct used for the calculation of  $\text{RPU}_G$  are shown.
- B** The strength of the  $\text{RPU}_G$  reference promoter is shown. Gray dotted line shows autofluorescence level, and the black dotted line shows the  $\text{RPU}_G$  reference characterized on a p15a plasmid (Nielsen *et al.*, 2016; Shin *et al.*, 2020). The means of three experiments performed on different days are shown, and the error bars are the standard deviation of these measurements.
- C** The sensor array is shown; genetic parts are provided in Appendix Table S5. The output promoters associated with each sensor is shown at the top of the bar graph. The detailed characterization of the sensors, including ON/OFF values and response functions, is shown in Appendix Fig S5 and Appendix Table S3. The means of three experiments performed on different days are shown, and the error bars are the standard deviation of these measurements.
- D** Characterization of NOT gates. The [...] in the construct represent portions of the sensor array not including in the schematic for clarity. Response functions were measured for the following 12 concentrations of IPTG: 0, 2, 5, 10, 20, 40, 60, 80, 100, 200, 400, and 500  $\mu\text{M}$ . The x-axis was converted to  $\text{RPU}_G$  by separately measuring the activity of the IPTG-inducible promoter for these inducer concentrations (Appendix Fig S12). Each response function was fit to the Hill equation (Materials and Methods), and the fit parameters are provided in Table 1. The means of three experiments performed on different days are shown, and the error bars are the standard deviation of these measurements. The blue lines show the impact on cell growth ( $\text{OD}_{600}$ ) when the input promoter to the gate is turned on (Materials and Methods). The smooth line is a linear regression to data from twelve inducer concentrations repeated over 3 days (Appendix Fig S8).

### NOT gates in Landing Pads #1 and #2

The NOT gate response functions need to be measured in the same genetic context used to carry the circuits. The genome-encoded IPTG-inducible system in Landing Pad #3 is used as the input ( $P_{\text{Tac}}$ ), and the gate is carried in Landing Pad #1 (Fig 2d). The corresponding output promoter driving the expression of YFP is carried in Landing Pad #2. The response function is measured by changing the concentration of IPTG and measuring the output at steady state using flow cytometry (Materials and Methods). However, this results in a function whose y-axis is in units of inducer concentration. To change the units to the activity of the input promoter, separate measurements are made for the induction of  $P_{\text{Tac}}$  in Landing Pad #1 (Appendix Fig S12).

Initially, we characterized a set of 10 gates based on orthogonal TetR-family repressors that had been previously designed to build circuits on a p15a plasmid (SrpR, PhlF, QacR, AmtR, LitR, BM3R1, PsrA, AmeR, and BetI; Stanton *et al.*, 2014; Nielsen *et al.*, 2016). From the initial set of repressors, we removed SprR, LitR, and PsrA because of crosstalk or toxicity. Several changes were made to the remaining gates to improve their dynamic ranges when carried in the genome: Alternative RBSs were selected for PhlF/BetI, and mutations were made to AmeR to improve binding to its operator (AmeRs) (Appendix Fig S7). To better insulate the gates, we replaced their single terminators with double terminators not already used as part of the landing pads (Appendix Fig S6 and S14, Appendix Table S4). The response functions of the six gates are shown in Fig 2D (Table 1). The impact of each gate on cell growth is minimal (Fig 2D, Appendix Fig S8; Shin *et al.*, 2020).

## NOR gate design for the genome

To build circuits, the NOT gates have to be convertible to multi-input NOR gates. On plasmids, we found that it was efficient to place two promoters in tandem upstream of the repressor gene (Tamsir *et al.*, 2011). Our first genome-encoded gates were based on this compact design, but we found that they did not function well (Fig 3D and F, and Appendix Fig S9). Indeed, this design has several problems that are exacerbated when carried on the genome. The first is that the upstream promoter is inhibited by the binding of repressor to the downstream promoter (“roadblocking”), which is more problematic at lower copy number (Nielsen *et al.*, 2016; Shin *et al.*, 2020). In addition, maximum repressor expression is less than the sum of the input promoter activities.

A new NOR gate design was developed that splits the gate such that there are two copies of the repressor gene, each of which is driven by a different input promoter (Fig 3B). To avoid homologous recombination, the two genes are encoded in different landing pads. Each gene contains the same ribozyme, RBS, codon usage, and terminator in order to produce the same repressor expression levels in response to a given input promoter activity. Two versions of the PhIF repressor NOR gate were constructed based on the old (tandem) and new (split) gate designs (Fig 3A and B). Both gates were evaluated for 64 combinations of inducers (aTc and ara), and the activity of the output promoters was compared (Fig 3C). The gates perform comparably, noting that the maximum repression is higher for the split design, as expected. The two responses are close enough where the NOT gate response functions can be used in the UCF for circuit design without having to include data for the 2D response functions associated with NOR gates.

We found Cello’s prediction of circuit response improved with the use of split gates. An example is shown in Fig 3D for a 2-input AND gate where the signals from the two sensors are inverted with NOT gates before being integrated by a NOR gate. The design based on the tandem gates shows the on state is far from the predicted response. In contrast, when the split-gate design is used, the data closely match the predicted responses (Fig 3E). In addition to the AND gate, we constructed two versions of a 3-input circuit that contains four NOT/NOR gates and compared its output to that predicted for all eight combinations of inputs (Fig 3F and Appendix Fig S9). The circuits based on split gates consistently outperformed those based on tandem gates, sometimes by orders of magnitude.

**Table 1. Response function parameters of NOT gates in the genome<sup>a</sup>**

Gate name	$y_{max}$ (RPU)	$y_{min}$ (RPU)	$K$ (RPU)	$n$
P1_PhIF	5.12	0.01	0.15	2.4
Q1_QacR	4.52	0.04	0.97	4.3
A1_AmtR	2.08	0.03	0.15	1.7
B3_BM3R1	0.65	0.01	0.40	2.7
F2_AmeRs	3.69	0.03	0.13	1.7
E1_BetI	2.25	0.25	1.25	3.8

<sup>a</sup>The fit parameters are shown for a Hill equation ( $y = y_{min} + ((y_{max} - y_{min}) / (1 + (x/K)^n))$ ).

## Automated genetic circuit design for the genome

Cello is design automation software that allows a user to define the desired circuit function using the Verilog language and specify the sensors to serve as inputs. The design is mapped to a particular species, genetic location, and gate technology, all of which are contained in the user constraint file (UCF). The software designs the DNA sequence containing the circuit and predicts the response of the outputs for different combinations of input activity. The software also predicts the impact of the circuit states on the growth rates of the cell.

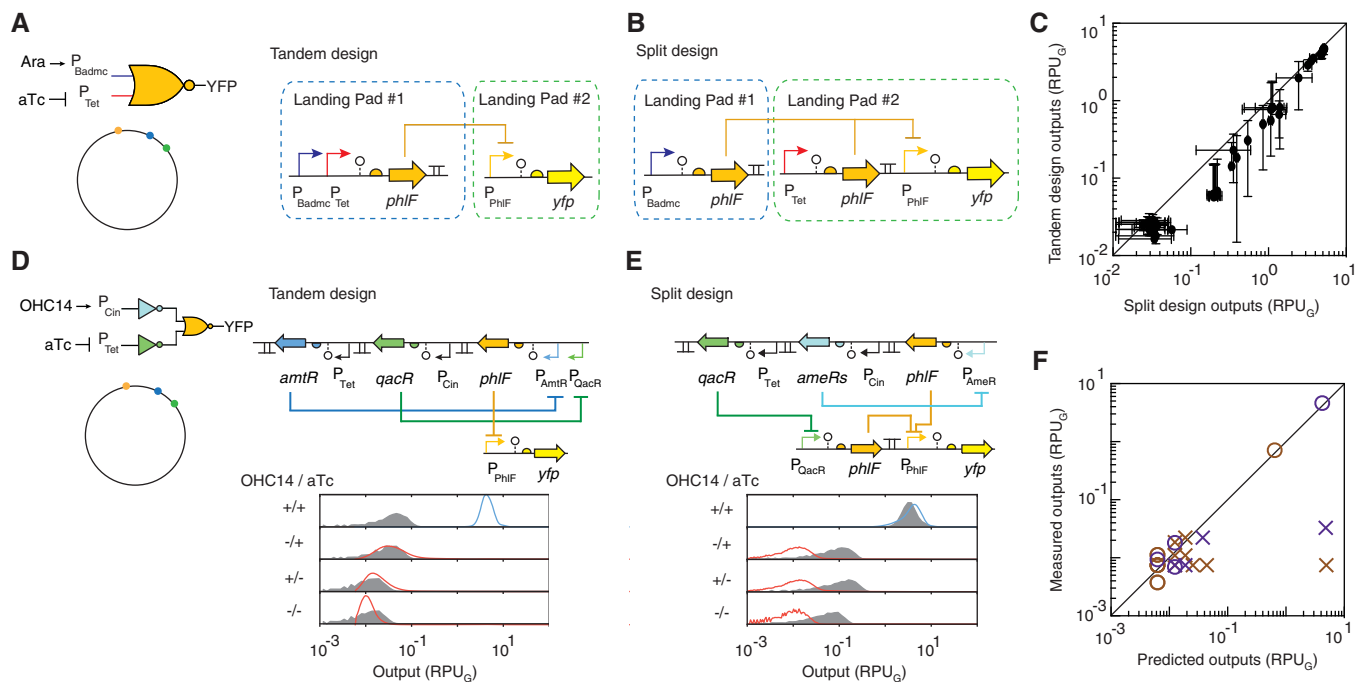
We constructed a new UCF (Eco2C1G3T1) to design circuits for the three landing pads in *E. coli* MG1655 that works with Cell version 2.0 (provided as Appendix File S1). The response functions for the gates, including cytometry distributions, are included along with the impact on growth ( $OD_{600}$ ) (Fig 2D). Design constraints (gene order, orientation, and locations in the landing pads) are encoded as EUGENE rules (Oberortner *et al.*, 2014; Nielsen *et al.*, 2016). The split-gate designs required modifying the Cello code (Materials and Methods).

Cello was used to design five genetic circuits for the genome with up to 6 gates and 4 sensors (Fig 4). Each design was constructed as predicted with no additional DNA modifications (Materials and Methods). The circuits were constructed using the same sensor array in Landing Pad #3. All the circuits were characterized and found to closely match the predicted responses, with the exception of two high off states in 0x0B that were still fivefold lower than the lowest on state.

## Evolutionary stability and total RNAP flux

Circuits draw on different levels of host resources depending on their state. For example, each state of a logic circuit, defined by a combination of inputs, corresponds to a different combination of active promoters and expressed repressors (Canton *et al.*, 2008). Therefore, each state requires a different amount of the host’s RNAPs and ribosomes. It has been shown that the expression of heterologous genes decreases the growth rate of the host and provides evolutionary incentive to remove the offending DNA (Klump *et al.*, 2009; Scott *et al.*, 2010). We have observed a higher probability of circuit breakage under conditions where the total RNAP flux used by a plasmid-encoded circuit is high (Shin *et al.*, 2020).

Two circuits were designed by Cello to compare the impact of carrying the circuit in the genome or a plasmid (Fig 5A). Each circuit was designed to produce the same 3-input logic operation, encoded using Verilog, but different UCFs were used to map the circuit to a DNA sequence. The first was designed using the Eco1C2G2T2 UCF for the p15a plasmid in *E. coli* DH10 $\beta$  and was published previously (Nielsen *et al.*, 2016; Shin *et al.*, 2020). The second was designed using the Eco2C1G3T1 for the *E. coli* MG1655 genome. Note that despite encoding the same logic operation, the circuits have different repressor assignments and DNA sequences. RPU is a surrogate for the RNAP flux exiting a promoter; thus, the total RNAP usage of a circuit can be estimated by summing the activities of all circuit promoters (including the sensor output promoters) (Materials and Methods). There is up to a fourfold decrease in total RNAP flux required by the



**Figure 3. Comparison of tandem and split NOT gates.**

- A Tandem NOR gate encoded on the genome. Tandem refers to the  $P_{Badmc}$  and  $P_{Tet}$  being encoded in series before the  $PhIF$  repressor gene, all of which are encoded within one landing pad.
- B Split NOR gate encoded on the genome. The  $P_{Badmc}$  and  $P_{Tet}$  promoters each drive the expression of different  $PhIF$  repressor genes and each are encoded within different landing pads.
- C A comparison of the output of the tandem and split NOR gates to changes in the input promoters. Each point represents a combination of inducers leading to the induction of the input promoters. For  $P_{Bad}$  induction, 8 different concentrations of L-arabinose (0, 1  $\mu$ M, 31.25  $\mu$ M, 62.5  $\mu$ M, 0.25 mM, 1 mM, 10 mM, and 25 mM) were used. For  $P_{Tet}$  induction, 8 different concentrations of aTc (0, 0.01, 0.1, 0.5, 1, 5, 10, and 20 ng/ml) were used. The line is drawn at  $x = y$ . The means of three experiments performed on different days are shown, and the error bars are the standard deviation of these measurements.
- D An AND gate built with a tandem NOR gate carried in Landing Pad #1. The blue and red distributions show the Cello predictions, and the gray distributions show the experimental data. The inducer concentrations are 10  $\mu$ M OHC14 and 20 ng/ $\mu$ l aTc.
- E An AND gate built with a split NOR gate divided across Landing Pads #1 and #2.
- F The comparison between the predicted and measured outputs for different combinations of inducers for the AND gates in parts d, e (purple) and a second 3-input logic gate (Circuit Ox08, brown). Circuits based on Tandem NOR gate design are shown as symbol "X", and the circuits based on Split NOR gate design are shown as symbol "O". The additional inducer concentration is 200 mM vanillic acid (Tandem Ox08) and the genetic design, and full response of Ox08 is shown in Appendix Fig S9. The line is drawn at  $x = y$ . The means of three experiments performed on different days are shown, and the error bars are the standard deviation of these measurements. The error bars are often smaller than the data points.

genome-encoded circuit to function, depending on the state (Fig 5B).

The differences in resource usage lead to different impacts on cell growth between the two circuits (Fig 5C). Under our growth conditions, cells harboring genetic circuits showed lower cell growth than that of cells without the genetic circuits and that of cells only harboring a *yfp* reporter. The impact on the cell growth was lower when the circuit is carried in the genome (45% decrease in cell growth compared to cells without circuits, averaged across all eight states). When cells harbored plasmid-encoded circuits, 67% decrease in cell growth (averaged across all eight states) compared to *E. coli* DH10 $\beta$  with empty plasmid was observed.

The evolutionary stability of a genome-encoded circuit was then evaluated (Fig 5D). Circuits carried on plasmids can break quickly, even under antibiotic selection and when carried in recombinase-deficient strains (*E. coli* DH10 $\beta$ ; Chen et al, 2013; Fernandez-Rodriguez et al, 2015; Shin et al, 2020). To replicate conditions without antibiotic selection, a p15a plasmid containing the OxF1 circuit (Shin

et al, 2020) was carried for 2 weeks in *recA-E. coli* DH10 $\beta$  in M9 media without antibiotics. The plasmid was rapidly lost from the population, and no circuit response can be observed after 3–5 days (Fig 5A and Appendix Fig S13). In contrast, the genome-encoded circuit is stable over weeks without antibiotics while being cycled between different input states with no sign of reduced performance. After the 12-day experiment, the genome-encoded circuits were Sanger-sequenced by amplifying junctions of the circuit and genomic DNA and no mutations were observed to the circuit DNA.

## Discussion

We demonstrate that design automation can be applied to create stable regulatory networks carried in the genome of *E. coli*. This requires re-optimizing gates for the genome and re-measuring their response functions in this context. But once this is done and the data used to populate a UCF file, then designing a circuit for either a

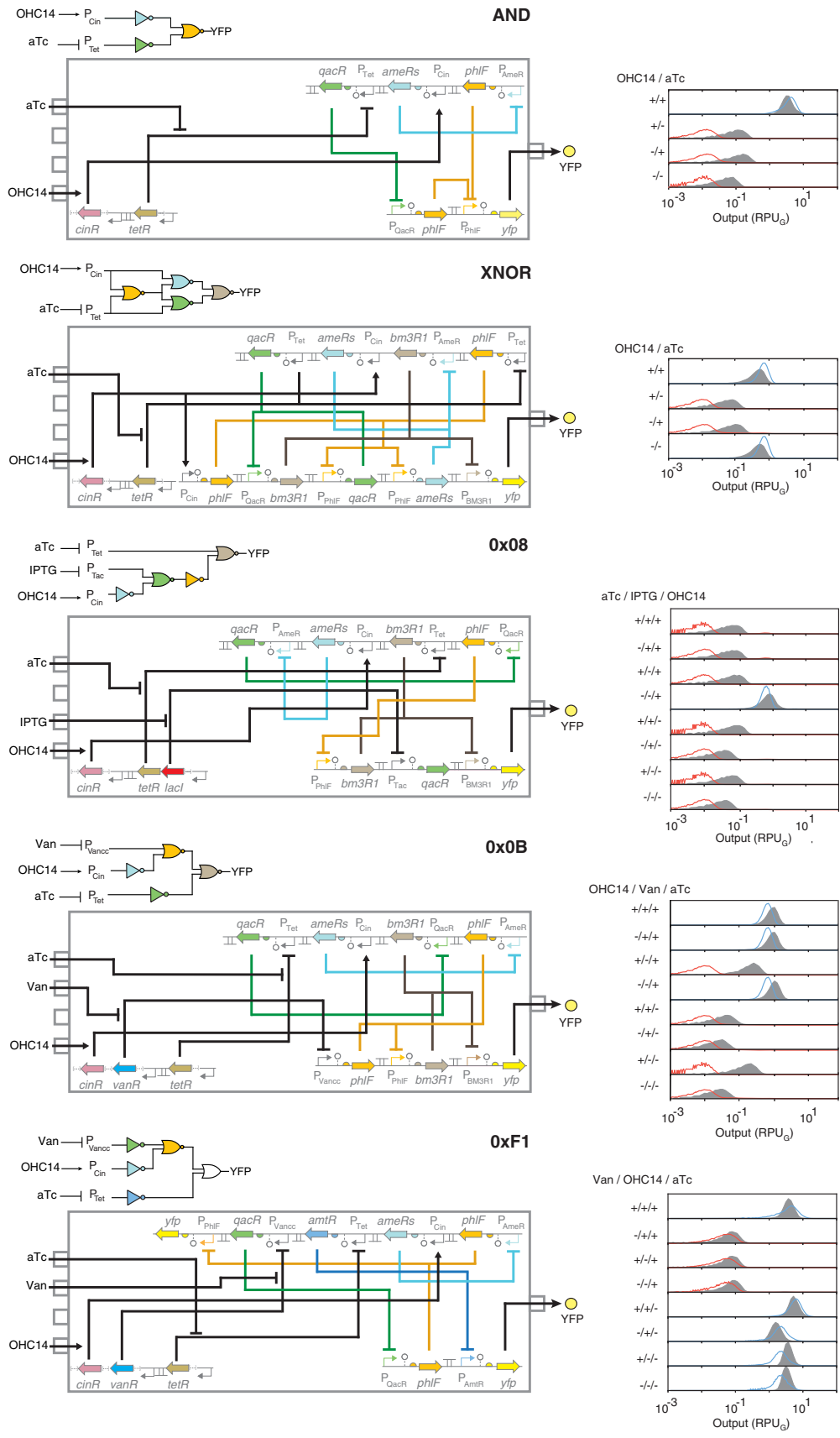
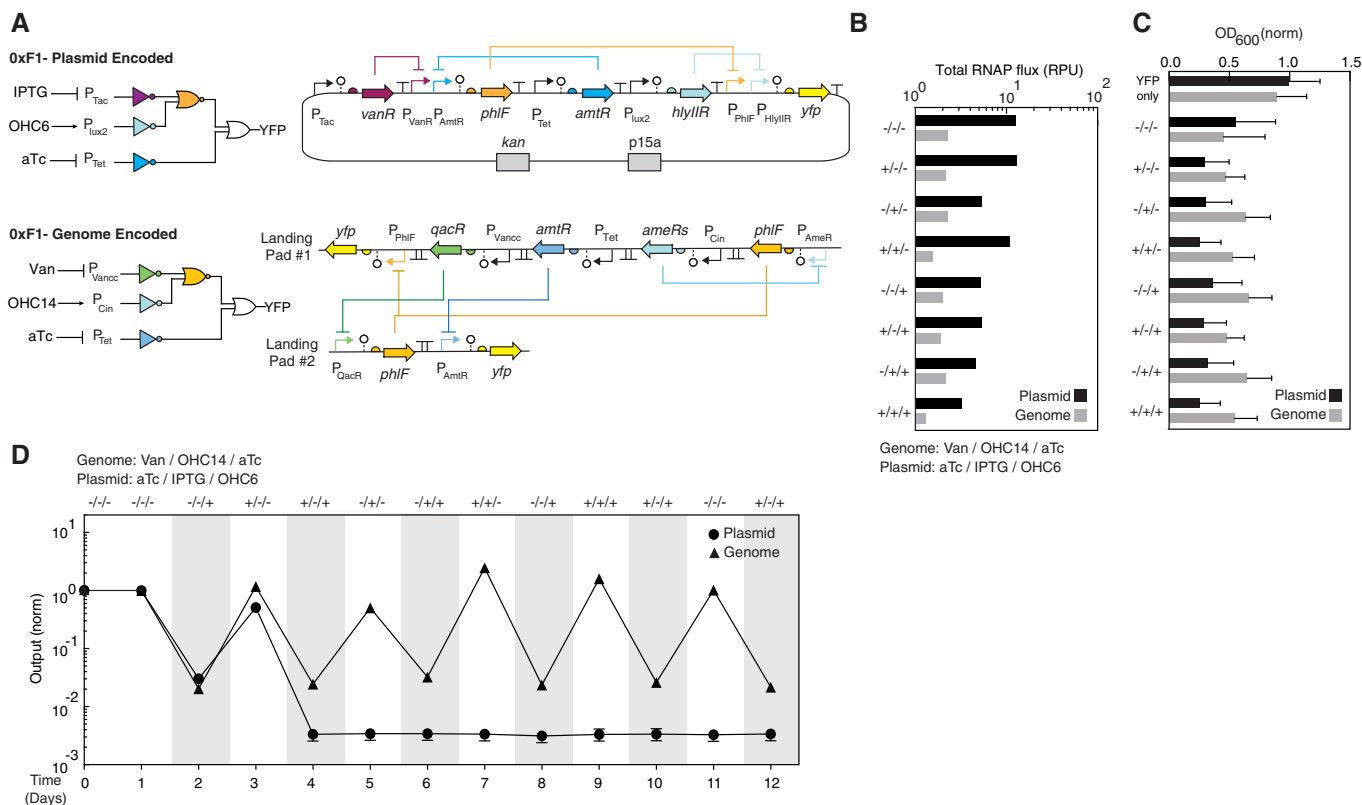


Figure 4.



**Figure 4. Genome-encoded genetic circuits designed by Cello.**

The wiring diagrams are shown with gates colored by the repressor identity. The gray boxes show the genetic circuit, with the construct in Landing Pad #1 at the top and Landing Pad #2 at the bottom. The genetic sensors used for the circuit are shown to the bottom left; all circuits use the same sensor array (Landing Pad #3), and [...] indicates portions of it not visualized for clarity. The concentrations of inducers used are 1 mM IPTG, 20 ng/μl aTc, 200 μM vanillic acid, and 10 μM of OHC14. The distributions predicted by Cello are shown in blue or red, and the gray distributions are the experimental measurements. The experiments were repeated three times on different days with similar results (error bars shown in Appendix Fig S10).

**Figure 5. Evolutionary stability of a circuit encoded on a genome versus a plasmid.**

- A** The wiring diagrams and designs are shown. The sequences of genetic parts are provided in Appendix Table S5.
- B** The total RNAP flux for the genome-encoded circuit (gray) is compared to the circuit designed for a p15a plasmid (black) (Materials and Methods). The inducers for the genome-encoded circuit and plasmid-encoded circuit are 200 μM vanillic acid, 10 μM OHC14/20 ng/μl aTc and 2 ng/μl aTc, 0.2 mM IPTG and 0.1 pg/μl OHC6, respectively.
- C** Cell density measurements for cultures grown in the different circuit states (combinations of inducers). The means of three experiments performed on different days are shown, and the error bars are the standard deviation of these measurements. The cell densities are normalized by the OD<sub>600</sub> measured using cells without genetic circuits (Materials and Methods).
- D** The activities of the circuit output promoters are shown as a function of time. The data are normalized by the fluorescence of the first time point (time = 0). The shading indicates the periods where cells are grown in the presence of the combinations of inducers shown at the top. The inducers for the genome-encoded circuit (triangles) and plasmid-encoded circuit (circles) are 200 μM vanillic acid, 10 μM OHC14, 20 ng/μl aTc and 2 ng/μl aTc, 0.2 mM IPTG and 0.1 pg/μl OHC6, respectively. The dashed line represents the autofluorescence of wild-type *E. coli* MG1655. The circuit designs and their responses to different combinations of inducers are shown in Appendix Fig S11. The autofluorescence was measured as 0.05 (a.u.).

plasmid or the genome is as simple as selecting a different UCF. Mapping of the circuit function to DNA is completely different in these cases, resulting in different assignments of repressors to the gates. Circuits designed using this approach are found to function as predicted, but this requires (i) insulation from context effects by carrying the circuit in an insulated landing pad, and (ii) a new gate architecture that functions more reliably, albeit requiring more DNA to encode.

This allows circuits to be designed for different genetic locations, depending on the needs of the application. A responsive therapeutic or diagnostic application may require the regulation be maximally sensitive and produce the largest possible response for a limited period of time (Chowdhury *et al.*, 2019). Some agricultural applications require very high expression of enzymes, for example, up to 25% of cell mass is composed of nitrogenase (Ryu *et al.*, 2020). These cases benefit from plasmids stabilized with additional

systems (Easter *et al*, 1998; Prell *et al*, 2002; Fedorec *et al*, 2019). When cells have to be maintained over long periods in a competitive environment, for example, in the microbiota in the human gut or in soil in agriculture, then it is important to reduce the resource utilization (Klump *et al*, 2009; Liao *et al*, 2017; Riglar *et al*, 2017; Ceroni *et al*, 2018; Bloch *et al*, 2020). Similarly, in bio-production applications that are sensitive to titers and yields from feedstock, the circuit needs to have essentially no growth impact and minimal draw on carbon or energy resources. These require maintaining the minimal copy number possible, either in the genome or from a bacterial artificial chromosome (BAC).

With the ease of DNA synthesis and genome editing, entire genomes can be built (Annaluru *et al*, 2014). Increasingly ambitious efforts have been undertaken to reorganize these genomes to simplify their further engineering, for example, the removal of transposons and problematic sequences, the reorganization of genes into clusters defined by function, or even the joining of entire chromosomes to create a single sequence (Datsenko & Wanner, 2000; Baba *et al*, 2006; Hutchison *et al*, 2016; Mitchell *et al*, 2017; Wu *et al*, 2017; Xie *et al*, 2017; Shao *et al*, 2018). Here, we start to think about the industrial design behind the organization of synthetic systems into the chromosome. Placing the sensors together into one landing pad helps organize the design. Similarly, allocating landing pads for regulatory circuits has an anthropomorphic purpose. One can imagine extending this to defining dedicated spaces for metabolic pathways, stress response, or combinations of functionality required for different stages of an application. Insulating these systems from the background processes required for growth and basic cellular function, as well as predicting and measuring this impact, will be increasingly important when designing genomes for increasingly complex applications, whether it be in a fermenter, the body, or the environment.

## Materials and Methods

### Strains and media

Unless otherwise noted, the strain used for the characterization of genetic circuits is *E. coli* MG1655 (numbering is based on the genome sequence, NCBI U00096.3; Blattner *et al*, 1997). Plasmid engineering was performed using *E. coli* DH10 $\beta$  (New England Biolabs, USA, C3019H) or *E. coli* DH5 $\alpha$  (New England Biolabs, USA, C2988J). *E. coli* TransforMax™ EC100D™ pir<sup>+</sup> (Lucigen, USA, CP09500) and *E. coli* JTK164A (Kittleson *et al*, 2011) were used for plasmids containing the R6K origin of replication. For conjugation, *E. coli* S17-1  $\lambda$ pir strain (TpR, SmR, *recA*, *thi*, *pro*, *hsdR*-M+RP4: 2-Tc:Mu: Km Tn7  $\lambda$ pir) was used as a donor strain to deliver plasmids with R6K origin of replication and OriT. LB media (BD Biosciences, USA, BD244610) were used for cell growth and cloning. 2xYT media (BD Biosciences, USA, DF0440-17) were used to grow cells for plasmid extraction with the QIAprep Spin Miniprep Kit (Qiagen, USA, 27104). Electrocompetent cells were prepped in SOB media (Teknova, USA, S0210). SOC recovery media (New England Biolabs, USA, B9020S) were used to recover cells after transformation. M9 media consist of M9 minimal salt (Sigma-Aldrich, USA, M6030) supplemented with 0.034% thiamine (Fisher Scientific, USA, BP892-100), 0.4% glucose (Fisher Chemical, USA, M-10046U), and 0.2%

casaminoacids (BD Biosciences, USA, 223050). This media (hereafter referred to as “M9 media”) were used for all measurements and characterizations, unless noted otherwise. Antibiotics used are as follows: ampicillin (100  $\mu$ g/ml, Amp) (GoldBio, USA, A-301-5), chloramphenicol (34  $\mu$ g/ml, Cm) (Alfa Aesar, USA, AAB20841-14), kanamycin (50  $\mu$ g/ml, Kan) (GoldBio, USA, K-120-10), spectinomycin (40  $\mu$ g/mL, Sp) (MP Biomedicals LLC, USA, 158993), and tetracycline (5  $\mu$ g/ml, Tet) (GoldBio, USA, T-101-25). All oligonucleotides, Gblocks, and oligos were ordered from IDT (Integrated DNA Technologies, USA; Appendix Fig S14 and Appendix Tables S4–S8).

### Terminator identification from genome sequences

Candidate terminators were identified from the *E. coli* MG1655 genome (NCBI RefSeq: NC\_000913) using RNIE version 0.01 with default settings and the “genome.cm” terminator model (Gardner *et al*, 2011). This produced an output GFF file containing the location (start and end base pair), orientation (sense or antisense strand), and scoring statistics for each putative terminator part. To assess the strength of each putative terminator, transcription profiles were generated for both sense and antisense strands of the MG1655 genome from the raw RNA-seq reads (Gorochowski *et al*, 2017). Next, for each putative terminator the appropriate transcription profile for the sense or antisense strand (depending on the orientation of the terminator) was selected. The terminator strength was then estimated by measuring the average transcription profile height for the 25-bp region before and after the position of the terminator (to smooth localized fluctuations) and calculating the ratio of the average transcription profile height directly after the terminator to directly before. Finally, those terminators that had been used previously in other work (Chen *et al*, 2013) were filtered out and a final ranked list of terminators by termination strength produced.

### Measurement of terminator strength

Terminators were characterized following a previously published assay (Chen *et al*, 2013). They were cloned into the pGR plasmid (Appendix Fig S14, Appendix Fig S1, Appendix Tables S1 and S5). *Escherichia coli* DH5 $\alpha$  harboring pGR plasmids with a terminator (pGR-DT#) were cultured overnight in 200  $\mu$ l of LB medium with ampicillin (100  $\mu$ g/ml). Cells were cultured using Nunc™ 96-well plates (Thermo Scientific, USA, 249662) in an ELMI Digital Thermo Microplate Shaker Incubator (ELMI Ltd, Latvia; hereafter “ELMI plate shaker”). The next day, cells were 200-fold diluted into 200  $\mu$ l of fresh LB medium with ampicillin (100  $\mu$ g/ml) and 12.5 mM L-arabinose (Sigma-Aldrich, USA, A3256). Cells were induced for 3 h at 37°C and 1,000 rpm in an ELMI plate shaker. After the induction, fluorescence levels were analyzed with flow cytometry and the geometric mean of GFP (FITC-A) and RFP (PE-Texas-RED-A) was calculated using FlowJo (TreeStar, Inc., USA) software. These geometric means were then used to calculate the terminator strength,  $T_S = [(<GFP>_{term}/<RFP>_{term})/(<GFP>_{ref}/<RFP>_{ref})]$ . The  $<RFP>_{term}$  and  $<GFP>_{term}$  denote geometric mean calculated for cells containing the plasmid terminator after subtracting the autofluorescence. The  $<RFP>_{ref}$  and  $<GFP>_{ref}$  denote geometric mean calculated with the pGR plasmid without a terminator between GFP and RFP after subtracting the autofluorescence.

### Flow cytometry analysis

Cytometry was performed using a LSRII Fortessa flow cytometer (BD Biosciences, USA). Upon harvesting, cells were diluted into 200  $\mu$ l of 1 $\times$  PBS ([NaCl]: 137 mM, [KCl]: 2.7 mM, [Na<sub>2</sub>HPO<sub>4</sub>]: 10 mM and [KH<sub>2</sub>PO<sub>4</sub>]: 1.8 mM) with 2 mg/ml kanamycin. An FSC voltage of 437 V, SSC voltage of 289 V, a green-laser (488 nm) voltage of 425 V, and a red-laser (561 nm) voltage of 489 V were used. For each sample, > 30,000 events were recorded. The recorded flow cytometry data were further analyzed with the software FlowJo. FITC-A and PE-Texas Red-A median values were used to represent the expression level distribution within a population.

### Construction of Tn5 transposon library

The pLYJP017 plasmid that contains a constitutively expressed mCherry probe was constructed by modifying the pBAMD1-4 plasmid (Martinez-Garcia *et al*, 2014). To prevent constant *tnpA* expression from backbone integration, *sfGFP* was added to the backbone as a counter-selection marker (Fig 1b). *Escherichia coli* S17-1  $\lambda$ pir electrocompetent cells were transformed with the pLYJP017 plasmid. The strain harboring pLYJP017 plasmid was then used for conjugation with *E. coli* DH10 $\beta$  carrying a tetracycline (Tet) resistance marker (*tetA*) in the genome (YJP\_DHC404; Appendix Fig S14). Donor strains (*E. coli* S17-1  $\lambda$ pir) and recipient strains (*E. coli* DH10 $\beta$ ) were separately grown overnight in 200  $\mu$ l LB media with antibiotics at 37°C and 1,000 rpm using Nunc™ 96-well plates (Thermo Scientific, USA, 249662) in an ELMI plate shaker. The next day, cells were 200-fold diluted into 4 ml LB media with antibiotics and incubated for 2.5 h at 37°C at 250 rpm in a New Brunswick Innova 44 Shaker (Eppendorf, USA). When cells reached OD<sub>600</sub> = 0.4, 250  $\mu$ l of both donor and recipient cells were mixed. Cells were then gently centrifuged (8,000 g, 25°C) and washed with 1 ml of LB without antibiotics four times at room temperature. Finally, cell pellets were resuspended with 50  $\mu$ l of SOC recovery media and were spotted on a plain LB agar plate (7  $\mu$ l per spot). Spotted LB agar plates were incubated 5 h at 37°C. Cells were then collected and resuspended into 1 ml SOC recovery media, followed immediately by two additional dilutions (100-fold and 10,000-fold). Then, 75  $\mu$ l of each diluted cell suspension was plated on LB agar plates with spectinomycin (40  $\mu$ g/ml) and tetracycline (5  $\mu$ g/ml). The next day, single colonies were picked from plates and grown in 200  $\mu$ l M9 media for 5.5 h at 37°C and 1,000 rpm in an ELMI plate shaker using Nunc™ 96-well plates (Thermo Scientific, USA, 249662). After the incubation, cells were analyzed with flow cytometry by measuring GFP (FITC-A) and mCherry (PE-Texas RED) expression levels. Only cells that have mCherry but not GFP were used for further analysis to determine the insertion location. For each colony, the insertion location was determined by amplifying the junction between inserted DNA and the neighboring genomic DNA. A randomized primer (oYJP1741: GGCACGCGTCTGACTAG-TACNNNNNNNNNNACGCC) and an insertion-specific primer (oYJP1745: CTTGGCCTCGCGCAGATCAG; Martinez-Garcia *et al*, 2014) were used to amplify the junction using two consecutive rounds of PCR amplifications. Each colony was suspended in water and was incubated at 95°C for 10 min to completely lyse cells. A 1.25  $\mu$ l aliquot of the colony suspension was added as a PCR template to the PCR premix that has 12.5  $\mu$ l of 2 $\times$  Phusion High-

Fidelity Master Mix (New England Biolabs, USA, M0531), 10  $\mu$ l of water, 1.25  $\mu$ l of 10  $\mu$ M oYJP1741 primer, and 0.5  $\mu$ l of 10  $\mu$ M oYJP1745 primers. The PCR products were then Sanger-sequenced with an internal primer oYJP1746 (CACCAAGGTAGTCCGGCAAAT) and aligned using NCBI nucleotide blast. Only the insertions with a unique hit were selected for further characterization. The mCherry expression levels and growth phenotypes were characterized for each member of the Tn5 transposon library. Each clone was streaked on LB agar plates with spectinomycin (40  $\mu$ g/ml) and Tet (5  $\mu$ g/ml). Single colonies were picked from plates and were inoculated overnight in M9 media without antibiotics for 16 h at 37°C and 1,000 rpm in an ELMI plate shaker using Nunc™ 96-well plates (Thermo Scientific, USA, 249662). Cells were diluted 185-fold into 200  $\mu$ l of fresh M9 media and incubated for 3 h. Cells were diluted again 700-fold into 200  $\mu$ l of fresh M9 media and grown for 6 h. After incubating for 6 h, 30  $\mu$ l of cells was added to 200  $\mu$ l of 1 $\times$  PBS solution with 2 mg/ml kanamycin (Kan) and fluorescence measured using flow cytometry. The optical density (OD) at 600 nm of the cultures was measured. To do this, 150  $\mu$ l of the culture was transferred to an optically transparent Nunc™ 96-well plates (Thermo Scientific, USA, 165305). The Hybrid Microplate Reader BioTek Synergy H1 (BioTek Instruments Inc, USA) was used to measure the final absorbance of the culture. The relative growth for each member of the library was calculated by relative growth = ((OD<sub>600: Tn5</sub> - (OD<sub>600: blank</sub>))/(OD<sub>600: DH10 $\beta$  Tet</sub> - (OD<sub>600: blank</sub>))), where OD<sub>600: Tn5</sub>, OD<sub>600: blank</sub> and OD<sub>600: DH10 $\beta$  Tet</sub> refer to OD<sub>600</sub> of Tn5 library member, blank M9 media, and *E. coli* DH10 $\beta$  with the tetracycline marker on the genome, respectively.

### Computational search for putative off-target integrase sites

To identify potential off-target sites for integrase 2, 5, and 7, the published attB/P sites (Yang *et al*, 2014) (Appendix Tables S5 and S6) were searched for in the *E. coli* MG1655 genome (NCBI U00096.3) using megablast (Zhang *et al*, 2000) and blastn (Altschul *et al*, 1997). Each site was tested whether it (i) has any matches to the genome via megablast, (ii) covers > 30% of the query sequence via blastn, and (iii) has a match with significant overlap (E-value < 0.1). If any of these criteria were true, the site was rejected.

### Construction of genomic landing pads

Two methods were used to insert the landing pads into the genome. Two landing pads (#1 and #2) were introduced sequentially using  $\lambda$ -RED recombineering (Datsenko & Wanner, 2000). *E. coli* MG1655 cells harboring arabinose-inducible  $\lambda$ -RED recombinase on a plasmid with a temperature-sensitive origin (pKD46; Datsenko & Wanner, 2000) were grown overnight using Falcon 14-ml round-bottom polypropylene tubes (Corning, USA, 352059). Cells were grown in 4 ml LB with Amp (100  $\mu$ g/ml) at 30°C and 250 rpm in a New Brunswick Innova 44 Shaker (Eppendorf, USA). The next day, cells were diluted 200-fold into 25 ml of fresh SOB medium in a nicked-bottom Erlenmeyer flask with ampicillin (100  $\mu$ g/ml) and 10 mM L-arabinose. Cells were induced for three hours at 30°C and 250 rpm in a New Brunswick Innova 44 Shaker (Eppendorf, USA) until they reached OD<sub>600</sub> = 0.4. Cells were then centrifuged (4,700 g, 4°C, 10 min) using Legend XFR centrifuge (Thermo Scientific, USA). Cell pellets were then washed with 5 ml of chilled 10%

glycerol and spun down again at 4,700 g and 4°C for 10 min. The washing step was repeated two more times using 1 ml of chilled 10% glycerol, and cells were pelleted using a refrigerated benchtop centrifuge (Eppendorf, USA) at 21,000 g and 4°C for 30 s. Finally, cell pellets were resuspended in 500 µl of chilled 10% glycerol. Next, 200 µl of these electrocompetent cells was transformed with 150 ng of landing pad DNA using electroporation (2,500 mA) (Eppendorf, USA) and recovered by adding 1 ml SOC recovery media and incubated at 30°C for 3 h. Transformed cells were then plated on LB agar (2%) with Cm (35 µg/ml) or Kan (50 µg/ml) antibiotics and incubated at 30°C overnight. Integration of landing pads was confirmed by PCR amplification and sequencing of the genomic regions that include landing pads. For Landing Pad #1, amplifying primers oYJP3436 (CCTGATCAGGTTCCGCGATCCC-GAATAACGGTC) and oYJP3437 (AGGCGCTGGAAGCGGCTTTG TGCTGGAAGATAAG) were used. For Landing Pad #2, amplifying primers oYJP3525 (ACCAATTGGCGCGCTTCGCAATAAAATTCC CTTCG) and oYJP3526 (TGCCAAAGCGGATAGGTGAAATAATGTC GGCGACAGCGG) were used. After integrating these two landing pads, the temperature-sensitive plasmid harboring λ-RED recombinase was removed by growing cells overnight at 37°C in LB without Amp (100 µg/ml). After the Landing Pads #1 and #2 were successfully integrated, Landing Pad #3 was constructed and integrated using a site-specific mini-Tn7 transposase (Choi & Schweizer, 2006). First, the Landing Pad #3 was cloned into plasmid pYJP072 (Appendix Fig S14). Next, *E. coli* S17-1 λpir electrocompetent cells were transformed with pYJP072 and the resulting strain was conjugated with two different strains for biparental mating. These two strains consist of a strain harboring Tn7 transposase-encoding plasmids and the strain containing Landing Pads #1 and #2 (described above). The conjugated cells were plated on LB agar plate (2%) with Cm (35 µg/ml) or Kan (50 µg/ml) and Tet (5 µg/ml) antibiotics. The insertion of Landing Pad #3 was confirmed by PCR-amplifying and sequencing genomic regions that include the landing pad by using PCR primers oYJP2826 (AGAGATGACAGAAAATTTT-CATTCTGTGACAGAGAAAAAGTAGCCGAAGATG) and oYJP2827 (CCGCGTAACCTGGCAAATCGGTTACGGTTGAGTAA). After the landing pads were inserted, phage transduction was used to move them into a clean genomic background. Transduction using P1 phage followed a previously published protocol, unless otherwise noted (Thomason *et al.*, 2007). To prepare P1 lysate, *E. coli* MG1655 cells harboring landing pads with three antibiotic markers (Kan, Cm, and Tet) (YJP\_MKC172) were cultured overnight at 37°C in LB media with antibiotics. The next day, cells were diluted 100-fold into 5 ml LB supplemented with 0.2% glucose and 5 mM calcium chloride (CaCl<sub>2</sub>) without antibiotics. After 45 min of incubation at 37°C, 250 rpm in a New Brunswick Innova 44 Shaker (Eppendorf, USA), 100 µl of P1 phage stock harvested from *E. coli* MG1655 was added to the culture and incubated for 3 h at 37°C. Once the culture was cleared, a few drops of chloroform (CHCl<sub>3</sub>) were added. Cell debris was spun down at 9,200 g for 10 min at 4°C using a refrigerated benchtop centrifuge (Eppendorf, USA). The resulting P1 lysate was further purified through 0.45-µm syringe filter (VWR international USA, 28145-481). To perform P1 transduction, wild-type *E. coli* MG1655 was grown in LB overnight at 37°C from a single colony. The next day, 1.5 ml of overnight culture was harvested and resuspended in 0.75 ml of P1 salt solution (10 mM CaCl<sub>2</sub> and 5 mM MgSO<sub>4</sub>) (Fisher Scientific, USA). Varying volumes of P1 lysate (100,

10, and 1 µl) were added to the 100 µl of the resuspended cells and were incubated for 30 min at 25°C. The mixtures were transferred into 1 ml of LB supplemented with 200 µl of sodium citrate and were incubated for 1 h at 37°C and 250 rpm in a New Brunswick Innova 44 Shaker (Eppendorf, USA). After the incubation, cells were centrifuged at 21,000 g and 25°C for 30 s and were plated on LB agar (2%) plates with antibiotics and 5 mM sodium citrate. Colonies were verified by PCR-amplifying genomic DNA with primers as described above (oYJP3436 and oYJP3437 for Landing Pad #1, oYJP3525 and oYJP3526 for Landing Pad #2, and oYJP2826 and oYJP2827 for Landing Pad #3). After the genomic insertion of landing pads, each landing pad contained a unique antibiotic resistance marker (Cm, Kan, and Tet for Landing Pads #1, #2, and #3, respectively). These antibiotic resistance markers were located between a pair of unidirectional flippase recognition target (FRT) sites.

### RNA-seq

RNA-seq libraries were prepared following a previously described method (Gorochowski *et al.*, 2017). *Escherichia coli* MG1655 strains with integrated landing pads (YJP\_MKC173) were first streaked on LB agar (2%) plates without antibiotics and incubated at 37°C. Single colonies were selected and grown overnight in M9 media without antibiotics at 37°C. The next day, cells were diluted 185-fold into 200 µl M9 media without antibiotics and grown for 3 h at 37°C and 1,000 rpm in an ELMI plate shaker using Nunc™ 96-well plates (Thermo Scientific, USA, 249662). After 3 h, cells were diluted 700-fold by adding 4.28 µl of the culture into 3 ml of M9 media without antibiotics. Cells were grown using Falcon 14-ml round-bottom polypropylene tubes (Corning, USA, 352059) in an Innova 44 Shaker (Eppendorf, USA) at 37°C, 250 rpm. After 5 h, cells were spun down at 4°C and 21,000 g for 3 min to collect the cell pellets for RNA-seq library preparation. After discarding the supernatants, cell pellets were flash-frozen in liquid nitrogen for storage at -80°C. Cell pellets were lysed by adding 1 mg of lysozyme, in 10 mM Tris-HCl (pH 8.0) with 0.1 mM EDTA, and total RNA was extracted using PureLink RNA Mini Kit (Life Technologies, CA, 12183020). RNA samples were further purified and concentrated with RNA Clean & Concentrator-5 Kit (Zymo Research, R1015), which was verified by Bioanalyzer (Agilent, CA). Ribosomal RNAs were depleted from RNA samples using Ribo-Zero rRNA Removal Kit for bacteria (Illumina, CA, MRZMB126). RNA-integrity numbers (RIN) were obtained for each sample, and only those samples with RIN > 8.5 were selected for library preparation. Strand-specific RNAtag-seq libraries were created by the Broad Technology Labs Microbial Omics Core (MOC) where uniquely barcoded samples were pooled together to run on two separate lanes of an Illumina HiSeq 2500. After the sequencing runs, reads from both lanes were combined, and the pooled mixture was de-multiplexed into original samples, followed by trimming the barcode tag from each read. Lysozyme, Tris-HCl (pH 8.0), and EDTA that were used for RNA-seq library preparation were purchased from Sigma-Aldrich (L6871), USB (75825), and USB (15694), respectively. Raw sequencing reads were aligned to the reference genomes, and transcription profiles were generated following a previously developed in-house Python script (Gorochowski *et al.*, 2017). Briefly, the first step of the process was to generate new reference files for the genome of strains with integrated landing pads. Therefore, for each strain, all integrated DNA

sequences were inserted into their corresponding locations on the reference genome of *E. coli* MG1655 (NCBI RefSeq: NC\_000913.3). These new FASTA and GFF files were then used to perform the alignment of raw reads using BWA version 0.7.4 with default settings, resulting in corresponding SAM and BAM files. Next, BAM files were filtered using the “view” command of SAMtools (Li et al, 2009; Barnett et al, 2011), and filter codes 83 and 163 were applied to select the reads mapping to the sense strand, and filter codes 99 and 147 were applied to select reads mapping to the antisense strands. Finally, filtered read coverage at each position along the reference sequence was normalized by the total mapped nucleotides across the genome and multiplied by  $10^9$  to generate the transcription profiles in both forward and reverse directions across the genome.

### Evaluation of thiamine-dependent growth

*Escherichia coli* MG1655 and two strains harboring Landing Pad #1 v1 and v2 were grown in M9 media (with thiamine) overnight. The next day, all three cultures were centrifuged (15,000 g, 25°C, 3 min) and resuspended into DI water three times to remove residual thiamine in the media. The  $OD_{600}$  of resuspended cells was measured, and the cells were diluted to  $OD_{600} = 0.01$ . Each dilution was then inoculated into thiamine-free M9 medium, consisting of M9 minimal salt (Sigma-Aldrich, USA, M6030) supplemented with 0.4% glucose and 0.2% casaminoacids (BD Biosciences, USA, 223050). After 6 h of incubation at 37°C and 250 rpm in a New Brunswick Innova 44 Shaker (Eppendorf, USA), the  $OD_{600}$  of three samples and a blank sample containing only thiamine-free M9 media were measured using a Cary 50 Bio Spectrophotometer (Agilent, USA).

### Insertion of payloads into landing pads

Note that an easy-to-follow detailed protocol is provided as Appendix Note S1. The strain containing the empty landing pads was co-transformed with a plasmid encoding three integrases (pLYJP053) and a plasmid containing the DNA payloads (pLYJP066-KanR, pLYJP070-CmR, and pLYJP064-TetR). To prepare electrocompetent cells, a single colony was inoculated into 2 ml LB without antibiotics and grown for 12 h at 37°C and 250 rpm in a New Brunswick Innova 44 Shaker (Eppendorf, USA) using Falcon 14-ml round-bottom polypropylene tubes (Corning, USA, 352059). The next day, 125  $\mu$ l of the overnight culture was added to 25 ml SOB medium in a nicked-bottom Erlenmeyer flask. Cells were grown for 2 h at 37°C and 250 rpm in a New Brunswick Innova 44 Shaker (Eppendorf, USA). When the early exponential phase was reached ( $OD_{600} = 0.3$ – $0.5$ ), cells were centrifuged (4,700 g, 4°C, 10 min) and washed with chilled 10% glycerol three times. After the third wash, cells were resuspended with 200  $\mu$ l of chilled 10% glycerol. Cells were electroporated with 500 ng of pLYJP053 plasmid and 500 ng of the payload plasmids. Immediately after the transformation, 1 ml of SOC recovery media was added to the cells. Cells were then incubated at 30°C for 3 h and plated on LB agar plates (2%) with necessary antibiotics. The insertion into the Landing Pads #1, #2, and #3 was selected with Kan (50  $\mu$ g/ml), Cm (35  $\mu$ g/ml), and Tet (5  $\mu$ g/ml), respectively. To confirm the integration with colony PCR, primers that can amplify the junction between integrated constructs and the adjacent genomic DNA were used. For Landing Pads #1 and #2, oYJP2164 (AATAAACAAATAGGCATGGTCTAAGAAACCATT) was

used as a primer that binds to the integrated construct. oYJP3436 (CCTGATCAGGTTCCGCGGATCCCGAATAAACGGTC) and oYJP3526 (TGCCAAAGCGGATAGGTGAAATAATGTCGGCGACAGCGG) were used as primers that bind genomic DNA adjacent to Landing Pad #1 and Landing Pad #2, respectively. For Landing Pad #3, a forward primer that binds to the end of integrated construct in a forward direction was used with primer oYJP2826 (AGAGATGACAGAAAAATTTTCATTCTGTGACAGAGAAAAAGTAGCCGAAGATG) to amplify the junction between Landing Pad #3 and the adjacent genomic DNA. The amplicon size was confirmed using gel electrophoresis. All three markers can be removed by transforming strains harboring the landing pads with inserted payloads with the pE-FLP plasmid containing a temperature-sensitive origin of replication (St-Pierre et al, 2013). To do this, each strain was cultured overnight in 2 ml of LB media with corresponding antibiotics (Cm (35  $\mu$ g/ml), Kan (50  $\mu$ g/ml), and Tet (5  $\mu$ g/ml)) at 37°C and 250 rpm in a New Brunswick Innova 44 Shaker (Eppendorf, USA) using Falcon 14-ml round-bottom polypropylene tubes (Corning, USA, 352059). The next day, cells were 200-fold diluted into 4 ml of LB media without antibiotics and incubated at 37°C and 250 rpm in a New Brunswick Innova 44 Shaker using Falcon 14-ml round-bottom polypropylene tubes (Corning, USA, 352059) for 2 h until reaching  $OD_{600} = 0.4$ . Cells were then harvested with centrifugation (4,700 g, 4°C, 10 min) using a Legend XFR centrifuge. Cell pellets were washed with 1 ml chilled 10% glycerol and spun down using a refrigerated benchtop centrifuge (Eppendorf, USA) at 21,000 g and 4°C for 30 s for three times. Finally, cell pellets were resuspended into 75  $\mu$ l of chilled 10% glycerol, yielding electrocompetent cells that were then transformed with 20 ng of pE-FLP (St-Pierre et al, 2013), followed by recovery in 1 ml SOC media and incubation at 30°C for 30 min. Cells were then plated on LB agar plates (2%) with Amp (100  $\mu$ g/ml) and incubated at 30°C overnight. The next day, three individual colonies were streaked on LB agar plates (2%) with no antibiotics and incubated at 37°C overnight. Three colonies from each streak were then grown in LB media with four antibiotics (Amp (100  $\mu$ g/ml), Cm (35  $\mu$ g/ml), Kan (50  $\mu$ g/ml), and Tet (5  $\mu$ g/ml)) at 37°C and 1,000 rpm for 16 h on an ELMI plate shaker using Nunc™ 96-well plates (Thermo Scientific, USA, 249662). Cells that did not grow in all four antibiotics were streaked and then used as the final strains.

### Calculation of RPU (relative promoter units)

*Escherichia coli* MG1655 containing the RPU<sub>C</sub> reference promoter (YJP\_MKC254) was streaked on a LB agar plate without antibiotics and incubated overnight at 37°C. Single colonies picked from the plate were then inoculated into 200  $\mu$ l of M9 media without antibiotics and were incubated overnight. All culturing steps were carried out at 37°C and 1,000 rpm in an ELMI plate shaker using Nunc™ 96-well plates (Thermo Scientific, USA, 249662). The next day, cells were diluted 185-fold into 200  $\mu$ l of M9 media without antibiotics and incubated for 3 h. Cells were then diluted again 700-fold into 200  $\mu$ l of M9 media without antibiotics and were incubated for 5 h. Then, a 30  $\mu$ l aliquot was transferred to 200  $\mu$ l of 1× PBS solution with 2 mg/ml kanamycin and evaluated using flow cytometry. To convert the fluorescence of a promoter from au ( $\langle YFP \rangle_{\text{measured}}$ ) to RPU<sub>C</sub>, the following equation is used:  $[(\langle YFP \rangle_{\text{measured}}) - (\langle YFP \rangle_{\text{blank}})] / [(\langle YFP \rangle_{\text{RPU}}) - (\langle YFP \rangle_{\text{blank}})]$ , where  $\langle YFP \rangle_{\text{blank}}$  is autofluorescence of wild-type *E. coli* MG1655.

### RPU<sub>G</sub>-to-RNAP flux conversion

RPU<sub>G</sub> was converted into RNAP flux by multiplying a previously calculated conversion factor 1 RPU = 0.019 RNAP/s per DNA (B. Shao, J. Rammohan, D.A. Anderson, N. Alperovich, D. Ross & C.A. Voigt, unpublished data) and the copy number of DNA where the Landing Pad #1 is located. The copy number of the Landing Pad #1 was estimated to 3.5 by comparing the Tn5 expression data for site 7 where Landing Pad #1 is located and the site #3, a site adjacent to the single-copy region of the genome. Note that the same promoter is used for the RPU and RPU<sub>G</sub> standard cassette, and it is assumed that this promoter produces the same constitutive flux in both locations. Therefore, 1 RPU<sub>G</sub> was converted into the 0.067 RNAP/s.

### Sensor characterization

The strain containing the seven sensors in Landing Pad #3 (YJP\_MKC174) was transformed with a reporter plasmid containing a promoter fused to *yfp* (pLYJP067-(promoter name)) that is responsive to the seven regulators (AraC, LacI, TetR, CymR<sup>AM</sup>, VanR<sup>AM</sup>, CinR<sup>AM</sup>, and TtgR<sup>AM</sup>) and streaked on LB agar (2%) plates with Kan (50 µg/ml). Single colonies were inoculated into 200 µl M9 media with Kan (50 µg/ml) were grown overnight at 37°C and 1,000 rpm in an ELMI plate shaker using Nunc™ 96-well plates. The next day, cells were diluted 185-fold into 200 µl fresh M9 media without any antibiotics and were incubated for three hours at 37°C and 1,000 rpm in an ELMI plate shaker using Nunc™ 96-well plates (Thermo Scientific, USA, 249662). Cells were then diluted 700-fold into 200 µl fresh M9 media (no antibiotics) with appropriate inducers and were incubated for 5.5 hours in an ELMI plate shaker at 37°C and 1,000 rpm using Nunc™ 96-well plates (Thermo Scientific, USA, 249662). Then, either 12.5 mM L-arabinose (Sigma-Aldrich, USA, A3256), 1 mM IPTG (GoldBio, USA, I2481C), 20 ng/µl aTc (Sigma-Aldrich, USA, 37919), 500 µM 4-isopropylbenzoic acid (Sigma-Aldrich, USA, 268402), 200 µM vanillic acid (Sigma-Aldrich, USA, H36001), 10 µM OHC14 (Sigma-Aldrich, USA, 51481), or 1 mM naringenin (Sigma-Aldrich, USA, N5893) was used to induce the sensors. After 5.5 h, 30 µl of cells was added to 200 µl 1× PBS with 2 mg/ml Kan for flow cytometry analysis.

### NOT/NOR gate characterization

Each strain containing a NOT gate was streaked on the LB agar (2%) plates with Kan (50 µg/ml) and Cm (35 µg/ml) antibiotics. A single colony was picked and inoculated into M9 media with Kan (50 µg/ml) and Cm (35 µg/ml) for overnight culture in an ELMI plate shaker at 37°C and 1,000 rpm. Cell cultures were performed using Nunc™ 96-well plates (Thermo Scientific, USA, 249662). The next day, cells were 185-fold diluted into 200 µl fresh M9 media with no antibiotics and incubated for 3 h in an ELMI plate shaker at 37°C and 1,000 rpm. After the 3 h, cells were then diluted again 700-fold into 200 µl fresh M9 media with no antibiotics and were incubated with inducers for additional 5.5 h in an ELMI plate shaker at 37°C and 1,000 rpm. After the 5.5 h, 30 µl of cells was added to 200 µl 1× PBS with 2 mg/ml Kan for flow cytometry analysis. To measure OD, 150 µl aliquots were

transferred to an optically transparent Nunc™ 96-well plates (Thermo Scientific, USA, 165305) to measure OD<sub>600</sub> using a Hybrid Microplate Reader BioTek Synergy H1 (BioTek Instruments Inc, USA). To measure the gate response functions, input and output promoter activities measured as median YFP fluorescence were converted into RPU<sub>G</sub>. The response functions were fit to a Hill equation  $y = y_{\min} + ((y_{\max} - y_{\min}) / (1 + (x/K)^n))$ , using an in-house Python script.

### Circuit design automation using Cello 2.0

The Cello 2.0 software (cellocad.org) and code are available open source on GitHub (github.com/CIDARLAB/Cello-v2). Note that the UCF in this paper is designed for Cello 2.0 and will not run with the old Web-based Cello interface. A UCF file (Eco2C1G3 T1.UCF.json) (Appendix File S1) was created to encode Hill parameters for NOT gate response functions in the gate library (Table 1), cytometry distributions of each gate, Eugene rules, landing pad location information, and growth assay conditions. Sensor output promoter activities ( $P_{\text{Badmc}}$ ,  $Y_{\min} = 0.04$ ,  $Y_{\max} = 3.33$ ;  $P_{\text{Tac}}$ ,  $Y_{\min} = 0.02$ ,  $Y_{\max} = 4.20$ ;  $P_{\text{Tet}}$ ,  $Y_{\min} = 0.02$ ,  $Y_{\max} = 5.41$ ;  $P_{\text{CymRC}}$ ,  $Y_{\min} = 0.19$ ,  $Y_{\max} = 2.39$ ;  $P_{\text{VanCC}}$ ,  $Y_{\min} = 0.02$ ,  $Y_{\max} = 3.79$ ;  $P_{\text{Cin}}$ ,  $Y_{\min} = 0.01$ ,  $Y_{\max} = 4.38$  and  $P_{\text{TtgR}}$ ,  $Y_{\min} = 0.01$ ,  $Y_{\max} = 0.22$  (in RPU<sub>G</sub>)), a UCF file, a truth table formulated as a Verilog file, and the growth score cutoff (set to 0.75) were used for each circuit design. Cello 2.0 was run locally using the Cello 2.0 API.

### Genetic circuit construction

Genetic circuits were first split and cloned into two plasmids, pLYJP066 (KanR) and pLYJP070 (CmR), that target Landing Pads #1 and #2, respectively, using Type II assembly as previously described (Nielsen *et al*, 2016; Shin *et al*, 2020). The order of transcription units within a circuit was assigned by Cello 2.0 based on EUGENE rules. In brief, each transcription unit of the circuit was sub-cloned into pLYJP080 (AmpR) (Appendix Fig S14 and Appendix Table S6) backbones with p15a origin using Type II assembly with BsaI (New England Biolabs, USA, R3733) and T4 ligase HC (Promega, USA, M1794). The reaction mix was prepared by adding 40 fmol of each DNA fragment, 1 µl of BsaI, 0.5 µl of T4 ligase HC, 1.5 µl of T4 ligase buffer (Promega, USA, M1794), and water up to 15 µl. The reaction mix was then cycled between 37°C (5 min) and 16°C (3 min) for 30 times, resulting in transcription units (in pLYJP080 plasmids) that are ready to be used in genetic circuit construction. For every designed genetic circuit, all the used transcription units were assembled into the full circuit using BbsI (New England Biolabs, USA, R3539) and T4 ligase HC (Promega, USA, M1794). Reaction mixture included 40 fmol of each pLYJP080 plasmids containing the transcription unit, 1 µl of BbsI, 0.5 µl of T4 ligase HC, 1.5 µl of T4 ligase buffer (Promega, USA, M1794), and water up to 15 µl. The reaction mix was then cycled between 37°C (5 min) and 16°C (3 min) for 50 times for assembly reaction, resulting in two new plasmid backbones pLYJP066 and pLYJP070 plasmids. These two plasmids were then integrated into the genome as described above. Antibiotic markers were removed by transforming pE-FLP (St-Pierre *et al*, 2013). pE-FLP plasmid was cured by incubating cells at 37°C as previously described (Appendix Fig S14).

## Genetic circuit characterization

Strains harboring genetic circuit were streaked on the LB agar (2%) plate without antibiotics. *E. coli* MG1655 wild-type and RPU<sub>C</sub> standard strains were streaked as controls. Individual colonies were picked from plates and were inoculated into M9 media without antibiotics. Cells were incubated overnight at 37°C and 1,000 rpm in an ELMI plate shaker using Nunc™ 96-well plates (Thermo Scientific, USA, 249662). The next day, cells were then diluted 185-fold into 200 µl of fresh M9 media without antibiotics and incubated for three hours at 37°C, 1,000 rpm in an ELMI plate shaker using Nunc™ 96-well plates (Thermo Scientific, USA, 249662). After three hours, cells were 700-fold diluted into M9 media without antibiotics and were induced with appropriate combinations of inducers such as 1 mM IPTG, 12.5 mM L-arabinose, 20 ng/µl aTc, 10 µM OHC14, and 200 µM vanillic acid were used as indicated in Cello 2.0 prediction. After 5.5 h, 30 µl of cells was added to 200 µl 1× PBS with 2 mg/ml Kan for flow cytometry analysis. Median YFP fluorescence from each sample was analyzed using FlowJo (TreeStar, Inc., USA) software and was converted into RPU<sub>C</sub>.

## Calculation of total RNAP flux

The input promoter activity (RNAP flux) of every NOT and NOR gate in the 0xF1 circuit was calculated for each state using the Cello 2.0 software package. The total RNAP flux for a circuit was calculated by summing the promoter activities across all the gates in the circuit, including the output promoters of the sensors. For the genome-encoded 0xF1 circuit, we used the UCF file Eco2C1G3T1, the truth table (0xF1.v), and the genome-encoded sensor output promoter activities (described above). The UCF was based on an ordinary additive model for tandem promoter activity and was not encoded with tandem roadblocking rules. For the plasmid-encoded 0xF1 circuit, we used the UCF file Eco1C2G2T2, the truth table (0xF1.v), and the plasmid-encoded sensor output promoter activities ( $P_{lux2}$ ,  $Y_{min} = 0.030$ ,  $Y_{max} = 2.234$ ;  $P_{Tac}$ ,  $Y_{min} = 0.018$ ,  $Y_{max} = 1.689$ ;  $P_{Tet}$ ,  $Y_{min} = 0.040$ ,  $Y_{max} = 1.967$ ; and  $P_{Cin}$ ,  $Y_{min} = 0.005$ ,  $Y_{max} = 3.178$  (in RPU)). The UCF was encoded with a non-additive tandem promoter model. Once the circuits were designed, the design output files (0xF1\_A000\_logic\_circuit.txt (genome), 0xF1\_A001\_logic\_circuit.txt (plasmid)) were used to calculate total RNAP flux used by the circuit. From each file, the input promoter activity of every NOT and NOR gate in the circuit (second to the last column) was summed to calculate the total RNAP flux. The total RNAP flux calculated for genome-encoded circuits (in RPU<sub>C</sub>) was then divided by 6.33 (Appendix Fig S4) to convert the RPU<sub>C</sub> into RPU. The total RNAP flux for a plasmid-encoded circuit was calculated in RPU.

## Long term stability test

The plasmid-encoded 0xF1 circuit was designed using the UCF for the p15a plasmid (Eco1C2G2T2) and was constructed in *E. coli* DH10β (Nielsen *et al*, 2016; Shin *et al*, 2020). The *E. coli* MG1655 strain harboring the genome-encoded 0xF1 circuit was streaked on LB agar (2%) plates with Kan (50 µg/ml) and Cm (35 µg/ml) antibiotics and grown overnight. The *E. coli* DH10β strain harboring the plasmid-encoded 0xF1 circuit was streaked on LB agar (2%) plates containing Kan (50 µg/ml) and grown overnight. For

each, three colonies were picked and grown in M9 media without antibiotics. Every day, each culture was diluted 10<sup>4</sup>-fold into 500 µl of fresh M9 media in 96-deep well plates (USA Scientific, USA, 1,896–2,000) with the inducer combination indicated in Fig 5. After incubation for 8 h at 37°C and 900 rpm in a Multitron Pro Incubator Shaker (In Vitro Technologies, VIC, Australia), 30 µl of cells was added to 200 µl 1× PBS with 2 mg/ml Kan for flow cytometry analysis and 100 µl of cells was mixed with 80% autoclaved glycerol (VWR chemical BDH1172-1LP) and stored at –80°C. Another aliquot was diluted 100-fold into fresh media with the same inducer combinations and incubated overnight. The cycle continued for 12 days by repeating this protocol. To measure the OD<sub>600</sub> of genome and plasmid-encoded circuits, individual colonies from the streak were inoculated into M9 media and incubated overnight at 37°C and 1,000 rpm in an ELMI plate shaker. Cells harboring plasmid-encoded genetic circuits were incubated overnight with Kan (50 µg/ml). Genome-encoded circuits were incubated overnight without antibiotics. The next day, each culture was diluted 185-fold into 3 ml of fresh M9 media and was grown for 3 h at 37°C and 250 rpm in an Innova Shaker. Three hours later, the OD<sub>600</sub> of each culture (OD<sub>600</sub>) was measured using a Cary 50 Bio Spectrophotometer (Agilent, USA). The measured OD<sub>600</sub> was used to calculate the amount of the sample required to transfer the same number of cells to the second dilution. Cells were diluted ~700-fold into 3 ml M9 media with appropriate inducers and were grown for 5.5 h at 37°C and 250 rpm in an Innova Shaker. The growth of cells without a circuit was determined using either *E. coli* MG1655 with empty landing pads (YJP\_MKC173) or *E. coli* DH10β harboring a p15a plasmid with a Kan (50 µg/ml) resistance gene (pYJP018).

## Data availability

The datasets and computer code produced in this study are available in the following databases: UCF information (Eco2C1G3T.UCF, Eco2C1G3T.input, and Eco2C1G3T.output) is available in Dataset EV1. Codes used to process the data are available on GitHub ([www.github.com/CIDARLAB/Cello-v2](http://www.github.com/CIDARLAB/Cello-v2)).

**Expanded View** for this article is available online.

## Acknowledgements

This work was supported by US Defense Advanced Research Projects Agency (DARPA) 1KM and SD2 awards HR0011-15-C-0084 and FA8750-17-C-0229 (C.A.V., A.E.B., and Y.P.), National Science Foundation Award CCF-1807575 (C.A.V.), U Colorado-Boulder/Dept of Energy subaward DE-SC0018368 (C.A.V., J.S.), and a Samsung Scholarship (Y.P.).

## Author contributions

YP and CAV conceived the study and designed the experiments; YP performed all the experiments; JS designed and constructed plasmid-based genetic circuits used in the study; YP and TEG developed new terminator prediction pipeline; YP and AEB performed the RNA sequencing data analysis; and YP, AEB, and CAV wrote the manuscript with input from all the authors.

## Conflict of interest

The authors declare that they have no conflict of interest.

## References

- Altschul SF, Madden TL, Schäffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res* 25: 3389–3402
- Annaluru N, Muller H, Mitchell LA, Ramalingam S, Stracquadiano G, Richardson SM, Dymond JS, Kuang Z, Scheifele LZ, Cooper EM et al (2014) Total synthesis of a functional designer eukaryotic chromosome. *Science* 344: 55–58
- Antunes MS, Morey KJ, Smith JJ, Albrecht KD, Bowen TA, Zdunek JK, Troupe JF, Cuneo MJ, Webb CT, Hellinga HW et al (2011) Programmable ligand detection system in plants through a synthetic signal transduction pathway. *PLoS One* 6: e16292
- Atkinson MR, Savageau MA, Myers JT, Ninfa AJ (2003) Development of genetic circuitry exhibiting toggle switch or oscillatory behavior in *Escherichia coli*. *Cell* 113: 597–607
- Baba T, Ara T, Hasegawa M, Takai Y, Okumura Y, Baba M, Datsenko KA, Tomita M, Wanner BL, Mori H (2006) Construction of *Escherichia coli* K-12 in-frame, single-gene knockout mutants: the Keio collection. *Mol Syst Biol* 2: 2006.0008
- Barnett DW, Garrison EK, Quinlan AR, Střimberg MP, Marth GT (2011) Bamtools: a C++ API and toolkit for analyzing and managing BAM files. *Bioinformatics* 27: 1691–1692
- Bassalo MC, Garst AD, Halweg-Edwards AL, Grau WC, Domaille DW, Mutalik VK, Arkin AP, Gill RT (2016) Rapid and efficient one-step metabolic pathway integration in *E. coli*. *ACS Synth Biol* 5: 561–568
- Basu S, Mehreja R, Thiberge S, Chen M-T, Weiss R (2004) Spatiotemporal control of gene expression with pulse-generating networks. *Proc Natl Acad Sci USA* 101: 6355–6360
- Blattner FR, Plunkett G III, Bloch CA, Perna NT, Burland V, Riley M, Collado-Vides J, Glasner JD, Rode CK, Mayhew GF et al (1997) The complete genome sequence of *Escherichia coli* K-12. *Science* 277: 1453–1462
- Bloch SE, Clark R, Gottlieb SS, Wood LK, Shah N, Mak S-M, Lorigan JG, Johnson J, Davis-Richardson AG, Williams L et al (2020) Biological nitrogen fixation in maize: optimizing nitrogenase expression in a root-associated diazotroph. *J Exp Bot* 71: 4591–4603
- Block DH, Hussein R, Liang LW, Lim HN (2012) Regulatory consequences of gene translocation in bacteria. *Nucleic Acids Res* 40: 8979–8992
- Borkowski O, Ceroni F, Stan G-B, Ellis T (2016) Overloaded and stressed: whole-cell considerations for bacterial synthetic biology. *Curr Opin Microbiol* 33: 123–130
- Brophy JA, Voigt CA (2014) Principles of genetic circuit design. *Nat Methods* 11: 508–520
- Brophy JA, Voigt CA (2016) Antisense transcription as a tool to tune gene expression. *Mol Syst Biol* 12: 854
- Brophy JAN, Triassi AJ, Adams BL, Renberg RL, Stratis-Cullum DN, Grossman AD, Voigt CA (2018) Engineered integrative and conjugative elements for efficient and inducible DNA transfer to undomesticated bacteria. *Nat Microbiol* 3: 1043–1053
- Bryant JA, Sellars LE, Busby SJW, Lee DJ (2014) Chromosome position effects on gene expression in *Escherichia coli* K-12. *Nucleic Acids Res* 42: 11383–11392
- Cai Y, Hartnett B, Gustafsson C, Peccoud J (2007) A syntactic model to design and verify synthetic genetic constructs derived from standard biological parts. *Bioinformatics* 23: 2760–2767
- Cambray G, Guimaraes JC, Mutalik VK, Lam C, Mai QA, Thimmaiah T, Carothers JM, Arkin AP, Endy D (2013) Measurement and modeling of intrinsic transcription terminators. *Nucleic Acids Res* 41: 5139–5148
- Canton B, Labno A, Endy D (2008) Refinement and standardization of synthetic biological parts and devices. *Nat Biotechnol* 26: 787–793
- Ceroni F, Boo A, Furini S, Gorochowski TE, Borkowski O, Ladak YN, Awan AR, Gilbert C, Stan GB, Ellis T (2018) Burden-driven feedback control of gene expression. *Nat Methods* 15: 387–393
- Chandler MG, Pritchard RHJM (1975) The effect of gene concentration and relative gene dosage on gene output in *Escherichia coli*. *Mol Gen Genet* 138: 127–141
- Chang MM, Gaidukov L, Jung G, Tseng WA, Scarcelli JJ, Cornell R, Marshall JK, Lyles JL, Sakorafas P, Chu A-HA et al (2019) Small-molecule control of antibody N-glycosylation in engineered mammalian cells. *Nat Chem Biol* 15: 730–736
- Chen YJ, Liu P, Nielsen AA, Brophy JA, Clancy K, Peterson T, Voigt CA (2013) Characterization of 582 natural and synthetic terminators and quantification of their design constraints. *Nat Methods* 10: 659–664
- Chiang CS, Bremer H (1988) Stability of pBR322-derived plasmids. *Plasmid* 20: 207–220
- Choi KH, Schweizer HP (2006) mini-Tn7 insertion in bacteria with single attTn7 sites: example *Pseudomonas aeruginosa*. *Nat Protoc* 1: 153–161
- Chong S, Chen C, Ge H, Xie XS (2014) Mechanism of transcriptional bursting in bacteria. *Cell* 158: 314–326
- Chowdhury S, Castro S, Coker C, Hinchliffe TE, Arpaia N, Danino T (2019) Programmable bacteria induce durable tumor regression and systemic antitumor immunity. *Nat Med* 25: 1057–1063
- Danino T, Prindle A, Kwong GA, Skalak M, Li H, Allen K, Hasty J, Bhatia SN (2015) Programmable probiotics for detection of cancer in urine. *Sci Transl Med* 7: 289ra284
- Datsenko KA, Wanner BL (2000) One-step inactivation of chromosomal genes in *Escherichia coli* K-12 using PCR products. *Proc Natl Acad Sci USA* 97: 6640–6645
- Davidsohn N, Beal J, Kiani S, Adler A, Yaman F, Li Y, Xie Z, Weiss R (2015) Accurate predictions of genetic circuit behavior from part characterization and modular composition. *ACS Synth Biol* 4: 673–681
- Davis JH, Rubin AJ, Sauer RT (2011) Design, construction and characterization of a set of insulated bacterial promoters. *Nucleic Acids Res* 39: 1131–1141
- Din MO, Danino T, Prindle A, Skalak M, Selimkhanov J, Allen K, Julio E, Atolia E, Tsimring LS, Bhatia SN et al (2016) Synchronized cycles of bacterial lysis for in vivo delivery. *Nature* 536: 81–85
- Dorman CJ, Dorman MJ (2016) DNA supercoiling is a fundamental regulatory principle in the control of bacterial gene expression. *Biophys Rev* 8: 89–100
- Duportet X, Wroblewska L, Guye P, Li Y, Eyquem J, Rieders J, Rimchala T, Batt G, Weiss R (2014) A platform for rapid prototyping of synthetic gene networks in mammalian cells. *Nucleic Acids Res* 42: 13440–13451
- Durfee T, Nelson R, Baldwin S, Plunkett G, Burland V, Mau B, Petrosino JF, Qin X, Muzny DM, Ayele M et al (2008) The complete genome sequence of *Escherichia coli* DH10B: insights into the biology of a laboratory workhorse. *J Bacteriol* 190: 2597–2606
- Easter CL, Schwab H, Helinski DR (1998) Role of the *parCBA* operon of the broad-host-range plasmid RK2 in stable plasmid maintenance. *J Bacteriol* 180: 6023–6030
- Elowitz MB, Leibler S (2000) A synthetic oscillatory network of transcriptional regulators. *Nature* 403: 335–338
- Enyeart PJ, Chirieleison SM, Dao MN, Perutka J, Quandt EM, Yao J, Whitt JT, Keatinge-Clay AT, Lambowitz AM, Ellington AD (2013) Generalized bacterial genome editing using mobile group II introns and Cre-lox. *Mol Syst Biol* 9: 685



- Enz S, Braun V, Crosa JH (1995) Transcription of the region encoding the ferric dicitrate-transport system in *Escherichia coli*: similarity between promoters for fecA and for extracytoplasmic function sigma factors. *Gene* 163: 13–18
- Enz S, Mahren S, Menzel C, Braun V (2003) Analysis of the ferric citrate transport gene promoter of *Escherichia coli*. *J Bacteriol* 185: 2387–2391
- Esvelt KM, Wang HH (2013) Genome-scale engineering for systems and synthetic biology. *Mol Syst Biol* 9: 641
- Fedorec AJH, Ozdemir T, Doshi A, Ho YK, Rosa L, Rutter J, Velazquez O, Pinheiro VB, Danino T, Barnes CP (2019) Two new plasmid post-segregational killing mechanisms for the implementation of synthetic gene networks in *Escherichia coli*. *iScience* 14: 323–334
- Fernandez-Rodriguez J, Yang L, Gorochowski TE, Gordon DB, Voigt CA (2015) Memory and combinatorial logic based on DNA inversions: dynamics and evolutionary stability. *ACS Synth Biol* 4: 1361–1372
- Fernandez-Rodriguez J, Moser F, Song M, Voigt CA (2017) Engineering RGB color vision into *Escherichia coli*. *Nat Chem Biol* 13: 706–708
- Fu YH, Tsai MM, Luo YN, Deonier RC (1991) Deletion analysis of the F plasmid oriT locus. *J Bacteriol* 173: 1012–1020
- Gaidukov L, Wroblewska L, Teague B, Nelson T, Zhang X, Liu Y, Jagtap K, Mamo S, Tseng WA, Lowe A et al (2018) A multi-landing pad DNA integration platform for mammalian cell engineering. *Nucleic Acids Res* 46: 4072–4086
- Gander MW, Vrana JD, Voje WE, Carothers JM, Klavins E (2017) Digital logic circuits in yeast with CRISPR-dCas9 NOR gates. *Nat Commun* 8: 15459
- Gardner TS, Cantor CR, Collins JJ (2000) Construction of a genetic toggle switch in *Escherichia coli*. *Nature* 403: 339–342
- Gardner PP, Barquist L, Bateman A, Nawrocki EP, Weinberg Z (2011) RNIE: genome-wide prediction of bacterial intrinsic terminators. *Nucleic Acids Res* 39: 5845–5852
- Gorochowski TE, Espah Borujeni A, Park Y, Nielsen AA, Zhang J, Der BS, Gordon DB, Voigt CA (2017) Genetic circuit characterization and debugging using RNA-seq. *Mol Syst Biol* 13: 952
- Grant SG, Jessee J, Bloom FR, Hanahan D (1990) Differential plasmid rescue from transgenic mouse DNAs into *Escherichia coli* methylation-restriction mutants. *Proc Natl Acad Sci USA* 87: 4645–4649
- Gu P, Yang F, Su T, Wang Q, Liang Q, Qi Q (2015) A rapid and reliable strategy for chromosomal integration of gene(s) with multiple copies. *Sci Rep* 5: 9684
- Guiziuo S, Ulliana F, Moreau V, Leclere M, Bonnet J (2018) An automated design framework for multicellular recombinase logic. *ACS Synth Biol* 7: 1406–1412
- Guo F, Adhya S (2007) Spiral structure of *Escherichia coli* HUalpha provides foundation for DNA supercoiling. *Proc Natl Acad Sci USA* 104: 4309–4314
- Gupta A, Reizman IMB, Reisch CR, Prather KLJ (2017) Dynamic regulation of metabolic flux in engineered bacteria using a pathway-independent quorum-sensing circuit. *Nat Biotechnol* 35: 273
- Gyorgy A, Jiménez José I, Yazbek J, Huang H-H, Chung H, Weiss R, Del Vecchio D (2015) Isocost lines describe the cellular economy of genetic circuits. *Biophys J* 109: 639–646
- Hutchison CA, Chuang R-Y, Noskov VN, Assad-Garcia N, Deerinck TJ, Ellisman MH, Gill J, Kannan K, Karas BJ, Ma L et al (2016) Design and synthesis of a minimal bacterial genome. *Science* 351: aad6253
- Isabella VM, Ha BN, Castillo MJ, Lubkowitz DJ, Rowe SE, Millet YA, Anderson CL, Li N, Fisher AB, West KA et al (2018) Development of a synthetic live bacterial therapeutic for the human metabolic disease phenylketonuria. *Nat Biotechnol* 36: 857
- Jiang Y, Chen B, Duan C, Sun B, Yang J, Yang S (2015) Multigene editing in the *Escherichia coli* genome via the CRISPR-Cas9 system. *Appl Environ Microbiol* 81: 2506–2514
- Jishage M, Ishihama A (1999) Transcriptional organization and in vivo role of the *Escherichia coli* *rsd* gene, encoding the regulator of RNA polymerase sigma D. *J Bacteriol* 181: 3768–3776
- Jusiak B, Jagtap K, Gaidukov L, Duportet X, Bandara K, Chu J, Zhang L, Weiss R, Lu TK (2019) Comparison of integrases identifies Bxb1-GA mutant as the most efficient site-specific integrase system in mammalian cells. *ACS Synth Biol* 8: 16–24
- Kelly JR, Rubin AJ, Davis JH, Ajo-Franklin CM, Cumbers J, Czar MJ, de Mora K, Gliberman AL, Monie DD, Endy D (2009) Measuring the activity of BioBrick promoters using an in vivo reference standard. *J Biol Eng* 3: 4
- Kittleson JT, Cheung S, Anderson JC (2011) Rapid optimization of gene dosage in *E. coli* using DIAL strains. *J Biol Eng* 5: 10
- Klumpp S, Zhang Z, Hwa T (2009) Growth rate-dependent global effects on gene expression in bacteria. *Cell* 139: 1366–1375
- Kong W, Wanda S-Y, Zhang X, Bollen W, Tinge SA, Roland KL, Curtiss R (2008) Regulated programmed lysis of recombinant Salmonella in host tissues to release protective antigens and confer biological containment. *Proc Natl Acad Sci USA* 105: 9361–9366
- Kotula JW, Kerns SJ, Shaket LA, Siraj L, Collins JJ, Way JC, Silver PA (2014) Programmable bacteria detect and record an environmental signal in the mammalian gut. *Proc Natl Acad Sci USA* 111: 4838–4843
- Kuhlman TE, Cox EC (2010) Site-specific chromosomal integration of large synthetic constructs. *Nucleic Acids Res* 38: e92
- Lambowitz AM, Zimmerly S (2011) Group II introns: mobile ribozymes that invade DNA. *Cold Spring Harb Perspect Biol* 3: a003616
- Lasa I, Toledo-Arana A, Dobin A, Villanueva M, de los Mozos IR, Vergara-Irigaray M, Segura V, Fagegaltier D, Penades JR, Valle J et al (2011) Genome-wide antisense transcription drives mRNA processing in bacteria. *Proc Natl Acad Sci USA* 108: 20172–20177
- Lee JW, Gyorgy A, Cameron DE, Pyenson N, Choi KR, Way JC, Silver PA, Del Vecchio D, Collins JJ (2016) Creating single-copy genetic circuits. *Mol Cell* 63: 329–336
- Leonardi R, Fairhurst SA, Kriek M, Lowe DJ, Roach PL (2003) Thiamine biosynthesis in *Escherichia coli*: isolation and initial characterisation of the ThiGH complex. *FEBS Lett* 539: 95–99
- Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R (2009) The sequence alignment/Map format and SAMtools. *Bioinformatics* 25: 2078–2079
- Liao C, Blanchard AE, Lu T (2017) An integrative circuit-host modelling framework for predicting synthetic gene network behaviours. *Nat Microbiol* 2: 1658–1666
- Liu Q, Schumacher J, Wan X, Lou C, Wang B (2018) Orthogonality and burdens of heterologous AND gate gene circuits in *E. coli*. *ACS Synth Biol* 7: 553–564
- Lou C, Stanton B, Chen Y-J, Munsky B, Voigt CA (2012) Ribozyme-based insulator parts buffer synthetic circuits from genetic context. *Nat Biotechnol* 30: 1137
- Lybecker M, Bilusic I, Raghavan R (2014) Pervasive transcription: detecting functional RNAs in bacteria. *Transcription* 5: e944039
- Martinez-Garcia E, Aparicio T, de Lorenzo V, Nikel PI (2014) New transposon tools tailored for metabolic engineering of gram-negative microbial cell factories. *Front Bioeng Biotechnol* 2: 46
- Mayo AE, Setty Y, Shavit S, Zaslaver A, Alon U (2006) Plasticity of the cis-regulatory input function of a gene. *PLoS Biol* 4: e45

- McAdams H, Shapiro L (1995) Circuit simulation of genetic networks. *Science* 269: 650–656
- Metcalfe WW, Jiang W, Wanner BL (1994) Use of the rep technique for allele replacement to construct new *Escherichia coli* hosts for maintenance of R6K gamma origin plasmids at different copy numbers. *Gene* 138: 1–7
- Meyer AJ, Segall-Shapiro TH, Glassey E, Zhang J, Voigt CA (2019) *Escherichia coli* “Marionette” strains with 12 highly optimized small-molecule sensors. *Nat Chem Biol* 15: 196–204
- Mimee M, Tucker AC, Voigt CA, Lu TK (2015) Programming a human commensal bacterium, bacteroides thetaiotaomicron, to sense and respond to stimuli in the murine gut microbiota. *Cell Syst* 1: 62–71
- Mitchell LA, Wang A, Stracquandano G, Kuang Z, Wang X, Yang K, Richardson S, Martin JA, Zhao Y, Walker R et al (2017) Synthesis, debugging, and effects of synthetic chromosome consolidation: synVI and beyond. *Science* 355: eaaf4831
- Mitschke J, Georg J, Scholz I, Sharma CM, Dienst D, Bantscheff J, Voss B, Steglich C, Wilde A, Vogel J et al (2011) An experimentally anchored map of transcriptional start sites in the model cyanobacterium *Synechocystis* sp. PCC6803. *Proc Natl Acad Sci USA* 108: 2124–2129
- Moon TS, Lou C, Tamsir A, Stanton BC, Voigt CA (2012) Genetic programs constructed from layered logic gates in single cells. *Nature* 491: 249–253
- Moser F, Broers NJ, Hartmans S, Tamsir A, Kerkman R, Roubos JA, Bovenberg R, Voigt CA (2012) Genetic circuit performance under conditions relevant for industrial bioreactors. *ACS Synth Biol* 1: 555–564
- Moser F, Borujeni AE, Ghodasara AN, Cameron E, Park Y, Voigt CA (2018) Dynamic control of endogenous metabolism with combinatorial logic circuits. *Mol Syst Biol* 14: e8605
- Mutalik VK, Guimaraes JC, Cambray G, Lam C, Christoffersen MJ, Mai QA, Tran AB, Paull M, Keasling JD, Arkin AP et al (2013) Precise and reliable gene expression via standard transcription and translation initiation elements. *Nat Methods* 10: 354–360
- Naydich AD, Nangle SN, Bues JJ, Trivedi D, Nissar N, Inniss MC, Niederhuber MJ, Way JC, Silver PA, Riglar DT (2019) Synthetic gene circuits enable systems-level biosensor trigger discovery at the host-microbe interface. *mSystems* 4: e00125-00119
- Nielsen AA, Der BS, Shin J, Vaidyanathan P, Paralanov V, Strychalski EA, Ross D, Densmore D, Voigt CA (2016) Genetic circuit design automation. *Science* 352: aac7341
- Oberortner E, Bhatia S, Lindgren E, Densmore D (2014) A rule-based design specification language for synthetic biology. *ACM J Emerg Technol Comput Syst* 11: 1–19
- Prell J, Boesten B, Poole P, Priefer UB (2002) The *Rhizobium leguminosarum* bv. viciae VF39  $\gamma$ -aminobutyrate (GABA) aminotransferase gene (*gabT*) is induced by GABA and highly expressed in bacteroids The GenBank accession number for the sequence determined in this work is AF335502. *Microbiology* 148: 615–623
- Pyne ME, Moo-Young M, Chung DA, Chou CP (2015) Coupling the CRISPR/Cas9 system with lambda red recombineering enables simplified chromosomal gene replacement in *Escherichia coli*. *Appl Environ Microbiol* 81: 5103–5114
- Riglar DT, Giessen TW, Baym M, Kerns SJ, Niederhuber MJ, Bronson RT, Kotula JW, Gerber GK, Way JC, Silver PA (2017) Engineered bacteria can function in the mammalian gut long-term as live diagnostics of inflammation. *Nat Biotechnol* 35: 653–658
- Rocha EP (2008) The organization of the bacterial genome. *Annu Rev Genet* 42: 211–233
- Roybal KT, Rupp LJ, Morsut L, Walker WJ, McNally KA, Park JS, Lim WA (2016) Precision tumor recognition by T cells with combinatorial antigen-sensing circuits. *Cell* 164: 770–779
- Ryu M-H, Zhang J, Toth T, Khokhani D, Geddes BA, Mus F, Garcia-Costas A, Peters JW, Poole PS, Ané J-M et al (2020) Control of nitrogen fixation in bacteria that associate with cereals. *Nat Microbiol* 5: 314–330
- Santos CN, Regitsky DD, Yoshikuni Y (2013) Implementation of stable and complex biological systems through recombinase-assisted genome engineering. *Nat Commun* 4: 2503
- Schmid MB, Roth JR (1987) Gene location affects expression level in *Salmonella typhimurium*. *J Bacteriol* 169: 2872–2875
- Scholz SA, Diao R, Wolfe MB, Fivenson EM, Lin XN, Freddolino PL (2019) High-resolution mapping of the *Escherichia coli* chromosome reveals positions of high and low transcription. *Cell Syst* 8: 212–225.e219
- Scott M, Gunderson CW, Mateescu EM, Zhang Z, Hwa T (2010) Interdependence of cell growth and gene expression: origins and consequences. *Science* 330: 1099–1102
- Segall-Shapiro TH, Sontag ED, Voigt CA (2018) Engineered promoters enable constant gene expression at any copy number in bacteria. *Nat Biotechnol* 36: 352–358
- Sezonov G, Joseleau-Petit D, D’Ari R (2007) *Escherichia coli* physiology in luria-berthani broth. *J Bacteriol* 189: 8746–8749
- Shafferman A, Kolter R, Stalker D, Helinski DR (1982) Plasmid R6K DNA replication: III. Regulatory properties of the  $\pi$  initiation protein. *J Mol Biol* 161: 57–76
- Shao Y, Lu N, Wu Z, Cai C, Wang S, Zhang LL, Zhou F, Xiao S, Liu L, Zeng X et al (2018) Creating a functional single-chromosome yeast. *Nature* 560: 331–335
- Sharan SK, Thomason LC, Kuznetsov SG, Court DL (2009) Recombineering: a homologous recombination-based method of genetic engineering. *Nat Protoc* 4: 206–223
- Shin J, Zhang S, Der BS, Nielsen AA, Voigt CA (2020) Programming *Escherichia coli* to function as a digital display. *Mol Syst Biol* 16: e9401
- Singh JS, Abhilash PC, Singh HB, Singh RP, Singh DP (2011) Genetically engineered bacteria: an emerging tool for environmental remediation and future research perspectives. *Gene* 480: 1–9
- Slager J, Veening J-W (2016) Hard-wired control of bacterial processes by chromosomal gene location. *Trends Microbiol* 24: 788–800
- Sleight SC, Bartley BA, Lieviant JA, Sauro HM (2010) Designing and engineering evolutionary robust genetic circuits. *J Biol Eng* 4: 12
- Sleight SC, Sauro HM (2013) Visualization of evolutionary stability dynamics and competitive fitness of *Escherichia coli* engineered with randomized multigene circuits. *ACS Synth Biol* 2: 519–528
- Sousa C, de Lorenzo V, Cebolla A (1997) Modulation of gene expression through chromosomal positioning in *Escherichia coli*. *Microbiology* 143: 2071–2078
- Stanton BC, Nielsen AA, Tamsir A, Clancy K, Peterson T, Voigt CA (2014) Genomic mining of prokaryotic repressors for orthogonal logic gates. *Nat Chem Biol* 10: 99–105
- Stirling F, Bitzan L, O’Keefe S, Redfield E, Oliver JWK, Way J, Silver PA (2017) Rational design of evolutionarily stable microbial kill switches. *Mol Cell* 68: 686–697.e683
- Stoebel DM, Dean AM, Dykhuizen DE (2008) The cost of expression of *Escherichia coli* lac operon proteins is in the process, not in the products. *Genetics* 178: 1653–1660
- St-Pierre F, Cui L, Priest DG, Endy D, Dodd IB, Shearwin KE (2013) One-step cloning and chromosomal integration of DNA. *ACS Synth Biol* 2: 537–541

- Stricker J, Cookson S, Bennett MR, Mather WH, Tsimring LS, Hasty J (2008) A fast, robust and tunable synthetic gene oscillator. *Nature* 456: 516
- Summers DK, Sherratt DJ (1984) Multimerization of high copy number plasmids causes instability: *colE1* encodes a determinant essential for plasmid monomerization and stability. *Cell* 36: 1097–1103
- Summers DK (1991) The kinetics of plasmid loss. *Trends Biotechnol* 9: 273–278
- Tabor JJ, Salis HM, Simpson ZB, Chevalier AA, Levskaya A, Marcotte EM, Voigt CA, Ellington AD (2009) A synthetic genetic edge detection program. *Cell* 137: 1272–1281
- Tamsir A, Tabor JJ, Voigt CA (2011) Robust multicellular computing using genetically encoded NOR gates and chemical ‘wires’. *Nature* 469: 212–215
- Thomason LC, Costantino N, Court DL (2007) *E. coli* genome manipulation by P1 transduction. *Curr Protoc Mol Biol* Chapter 1: Unit 1 17
- Vaidyanathan P, Der BS, Bhatia S, Roehner N, Silva R, Voigt CA, Densmore D (2015) A framework for genetic logic synthesis. *Proc IEEE* 103: 2196–2207
- Vander Horn PB, Backstrom AD, Stewart V, Begley TP (1993) Structural genes for thiamine biosynthetic enzymes (thiCEFGH) in *Escherichia coli* K-12. *J Bacteriol* 175: 982–992
- Vora T, Hottes AK, Tavazoie S (2009) Protein occupancy landscape of a bacterial genome. *Mol Cell* 35: 247–253
- Wade JT, Grainger DC (2014) Pervasive transcription: illuminating the dark matter of bacterial transcriptomes. *Nat Rev Microbiol* 12: 647–653
- Wang HH, Isaacs FJ, Carr PA, Sun ZZ, Xu G, Forest CR, Church GM (2009) Programming cells by multiplex genome engineering and accelerated evolution. *Nature* 460: 894–898
- Wang B, Kitney RI, Joly N, Buck M (2011) Engineering modular and orthogonal genetic logic gates for robust digital-like synthetic biology. *Nat Commun* 2: 508
- Wang K, Fredens J, Brunner SF, Kim SH, Chia T, Chin JW (2016a) Defining synonymous codon compression schemes by genome recoding. *Nature* 539: 59
- Wang Y, Penkul P, Milstein JN (2016b) Quantitative localization microscopy reveals a novel organization of a high-copy number plasmid. *Biophys J* 111: 467–479
- Wu Y, Li B-Z, Zhao M, Mitchell LA, Xie Z-X, Lin Q-H, Wang X, Xiao W-H, Wang Y, Zhou X et al (2017) Bug mapping and fitness testing of chemically synthesized chromosome X. *Science* 355: eaaf4706
- Xi J, Ge Y, Kinsland C, McLafferty FW, Begley TP (2001) Biosynthesis of the thiazole moiety of thiamin in *Escherichia coli*: identification of an acylsulfide-linked protein–protein conjugate that is functionally analogous to the ubiquitin/E1 complex. *Proc Natl Acad Sci USA* 98: 8513–8518
- Xie L, Hall D, Eiteman MA, Altman EJAM (2003) Optimization of recombinant aminolevulinate synthase production in *Escherichia coli* using factorial design. *Appl Microbiol Biotechnol* 63: 267–273
- Xie Z-X, Li B-Z, Mitchell LA, Wu Y, Qi X, Jin Z, Jia B, Wang X, Zeng B-X, Liu H-M et al (2017) ‘Perfect’ designer chromosome V and behavior of a ring derivative. *Science* 355: eaaf4704
- Yang L, Nielsen AA, Fernandez-Rodriguez J, McClune CJ, Laub MT, Lu TK, Voigt CA (2014) Permanent genetic memory with >1-byte capacity. *Nat Methods* 11: 1261–1266
- Yeung E, Dy AJ, Martin KB, Ng AH, Del Vecchio D, Beck JL, Collins JJ, Murray RM (2017) Biophysical constraints arising from compositional context in synthetic gene networks. *Cell Syst* 5: 11–24.e12
- Zhang Z, Schwartz S, Wagner L, Miller W (2000) Greedy algorithm for aligning DNA sequences. *J Comput Biol* 7: 203–214



**License:** This is an open access article under the terms of the Creative Commons Attribution 4.0 License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.