To perform their computational function, genetic circuits change states through a symphony of genetic parts that turn regulator expression on and off. Debugging is frustrated by an inability to characterize parts in the context of the circuit and identify the origins of failures. Here, we take snapshots of a large genetic circuit in different states: RNA-seq is used to visualize circuit function as a changing pattern of RNA polymerase (RNAP) flux along the DNA. Together with ribosome profiling, all 54 genetic parts (promoters, ribozymes, RBSs, terminators) are parameterized and used to inform a mathematical model that can predict circuit performance, dynamics, and robustness. The circuit behaves as designed; however, it is riddled with genetic errors, including cryptic sense/antisense promoters and translation, attenuation, incorrect start codons, and a failed gate. While not impacting the expected Boolean logic, they reduce the prediction accuracy and could lead to failures when the parts are used in other designs. Finally, the cellular power (RNAP and ribosome usage) required to maintain a circuit state is calculated. This work demonstrates the use of a small number of measurements to fully parameterize a regulatory circuit and quantify its impact on host.
Debbuging a genetic system is analogous to fixing a watch using only the displayed time, without being able to see the internal mechanisms. Genetic circuits are particularly difficult, as their operation requires coordination between many regulatory parts. Genome-scale technologies are illuminating the inner workings of the cell, including transcription, translation, and chemical composition. Computational tools are needed to extract parameters to inform models and to scan for errant features that can be fixed by making mutations or selecting different parts. These methods need to scale to larger circuits, corresponding to more parameters, without requiring more experiments.

NOR gates have two input promoters, a repressor that inverts their signal, and an output promoter. Sets of repressors have been used to build libraries of NOR gates that can be connected to build circuits. The flux of RNA polymerase on DNA is a critical measurement for genetic circuit construction. The response function of a gate captures how the output changes as a function of the input. The units of JRNAP have been reported in relative promoter units (RPUs) by comparing the strength of a promoter to that of a reference and single-molecule studies have concluded that the RNA-seq data are used to visualize the changes in the circuit is as predicted across all combinations of inputs (states). The RNA-seq data are used to visualize the changes in a promoter activity that occurs in response to an input signal.

Results

RNAP flux and ribosome usage patterns across the circuit. We sought to study a circuit that works well; that is, behaves as designed and does not lead to an obvious cellular phenotype. To this end, a Cello-designed E. coli circuit was selected that encodes a 3-input logic operation as one of the top-performing circuits from a set of 60 (Fig. 1a). It consists of sensors that respond to different stimuli by changing state, representing a different pattern of regulator expression and emulating from the circuit’s promoters. Therefore, the circuit requires continuous cellular power to store its state. Maintaining an RNAP flux requires cellular energy and metabolites, and gates require a futile cycle of regulator expression and degradation. Circuit states that place a larger burden on the cell take these resources away from cellular maintenance, thus decreasing growth and incentivizing evolutionary breakage.

Using RNA sequencing (RNA-seq), JRNAP can be simultaneously calculated for all of the promoters in a genetic circuit. RNA-seq is performed by isolating RNA, converting it to cDNA, deep sequencing the pool, and then mapping the reads to the genome to create a transcript profile. The level of gene expression can be calculated as the average profile height over its length, reported as FPKM (fragments per kilobase of transcript per million mapped reads). FPKM is an estimate of mRNA transcript copy number and has been converted to absolute units using RNA spike-in standards. When there is a sharp increase or decrease in RNA-seq transcription profile, the magnitude can be used to infer the strength of a promoter or terminator, respectively. When the average cDNA fragment size is large (100–500 nt), this complicates the calculation because it obscures nearby promoters and requires a correction factor at the transcript ends. To address this, we use a technique that uses short RNA fragments (<50 nucleotide) and single-end mode sequencing.

Cellular translation can be quantified using ribosome profiling, which takes global snapshot of ribosomes on mRNA transcripts. Ribosome profiling can be used to calculate translational parts, notably ribosome binding sites (RBSSs), and to measure translation start sites, pause sites, coupling in operons, and cryptic translation. It can be performed simultaneously with RNA-seq, where the cellular RNA content is collected, digested, and the ribosome-protected RNA is fractionated, converted to cDNA, and sequenced. A profile is built by mapping ribosome P-sites to the DNA sequence, the height of which is the number of ribosomes at this position (the ribosome occupancy). The average profile height over the gene length (the ribosome density, RD), serves as an estimate for protein expression. It can also be used to calculate the steady-state protein abundance, which can be used to estimate the proteome of the cell.

In this manuscript, all of the parts in a genetic circuit are characterized by combining data from RNA-seq and ribosome profiling. Design automation (Cello) is used to build a 3-input logic gate consisting of seven NOT/NOR gates. The response of the circuit is as predicted across all combinations of inputs (states). The RNA-seq data are used to visualize the changes in a promoter activity that occurs in response to an input signal.
**Fig. 1 Genetic circuit characterization using RNA-seq and ribosome profiling.**

**a** Wiring diagram of the model circuit is shown. The colors correspond to the repressors used for each gate.

**b** Genetic parts for the genetic circuit (left) and output promoter (right), carried on different plasmids ( Supplementary Fig. 1 and Supplementary Table 5).

**c** The measured response of the circuit output to different combinations of inputs (1 mM IPTG, 2 ng/ml (4.32 nM) aTc, and 5 mM Ara). The black distributions are experimental measurements (Methods) compared to Cello predictions (blue if ON, red if OFF).

**d** The RNAP flux across the circuit is shown (blue) with the ribosome occupancies (black). Each circuit state is shown, marked by the diagram (thick black lines indicate corresponding wires that are ON; gates are colored if that repressor is being expressed). The conversions of RNA-seq and ribosome profiling data to absolute units are described in the Methods. Source data are provided as a Source Data file.
observed RNAP bursts\textsuperscript{60–62}. At steady-state, the flux $J_i$ at each nucleotide $i$ is related to the height of the transcript profile $M_i$ by $J_i = \gamma M_i$, where $\gamma$ is the RNA degradation rate. For the circuit, we assume a constant $\gamma = 0.0067 \text{ s}^{-1}$ for all mRNAs, which is reasonable given that they all encode single repressor genes and are approximately the same length with similar 5'– and 3'–UTR (untranslated region) sequences\textsuperscript{63}. The profile is often reported in arbitrary units; however, we can convert it to absolute units in two steps (Methods). First, because we know the promoter in arbitrary units; however, we can convert it to absolute units in reasonable given that they all encode single repressor genes and assume a constant circuit DNA is entirely composed of synthetic parts: promoters, repressors and the design of the gates (Supplementary Fig. 6)\textsuperscript{14,16}. The punctate nature of the synthetic circuit can be seen, analogous to watching the a transcriptional circuit as it performs the computation, flow of electrons in an electronic circuit. The RNAP bursts\textsuperscript{60–62} can be visualized in each state of the circuit using ribosome profiling data (Methods). This shows the changes in the ribosome usage for each gate in their on and off states (Fig. 1d). The quantities of the bars indicate the number of ribosomes that occur at that position in arbitrary units. Similar to RNA-seq profiles, we can convert these profiles to absolute units. First, these occupancies were normalized by the sum of the occupancies at all positions in the circuit and genome, resulting in a profile that represents the fraction of total active ribosomes in the cell that are recruited to each DNA position. To convert this fraction profile to absolute units, we need an estimate of the total number of active ribosomes in the cell, which has been previously measured for E. coli and is a function of cell growth rate\textsuperscript{64}. In our experiments, this corresponds to 20,000 ribosomes that are actively translating mRNAs. The ribosome occupancy pattern across the circuit reflects the one obtained through RNA-seq, as expected (Fig. 1d). However, there is variability in the ribosome usage for each repressor gene, which has implications in the gate function and the cellular resources required to hold its state. The FPKM and RD measurements for the circuit genes are much higher than for those in the genome (Supplementary Fig. 6). The medians are 220-fold and 140-fold higher, respectively. In addition, we analyzed whether the higher estimated protein expression is dominated by higher transcript levels or higher translation per transcript. Van Oudenaarden and co-workers observed that there is more protein expression noise when the transcript level is low and the translation rate is high\textsuperscript{65}. Compared to genomic gene expression, the RD/FPKM levels of circuit proteins are lower, indicating a bias towards lower cell-to-cell variation. Indeed, this was a factor in the selection of these repressors and the design of the gates (Supplementary Fig. 6)\textsuperscript{14,16}.

**Errors in transcription and translation in the circuit DNA.** The circuit DNA is entirely composed of synthetic parts: promoters are based on well-defined scaffolds, the genes have been codon optimized, the 5'-UTRs contain ribozymes and RBs that were designed computationally. Still, we observe many examples of transcription and translation differing from that expected (Fig. 2a).

Cryptic promoters can adversely impact circuit function\textsuperscript{66,67}. They are identified by calculating the ratio between RNAP flux profiles at neighboring positions $J_{\text{RNAP}}(i+1)/J_{\text{RNAP}}(i)$ to which a threshold is applied to mark a transcription start site (TSS) (Fig. 2b) (Methods). Using this method, all of the TSSs are identified for the synthetic gate promoters, each of which only has one. However, the location of this TSS is upstream of the annotated position for $\text{P}_{\text{srpR}}$, $\text{P}_{\text{Tac}}$, $\text{P}_{\text{BAD,1}}$, and $\text{P}_{\text{BAD,2}}$ (Supplementary Fig. 7). The $\text{srpR}$ and $\text{amtR}$ genes have strong internal promoters in the sense direction. Remarkably, antisense promoters occur in 6/7 of the repressor genes with some stronger than the forward promoters of the gates. There is also a strong antisense promoter in the middle of both $\text{P}_{\text{BAD,1}}$ and $\text{P}_{\text{BAD,2}}$ promoters that corresponds to the known antisense promoter (P) in $\text{P}_{\text{BAD}}$\textsuperscript{68}. Our method identifies the TSS of this promoter occurring at the 5'-end of the $\text{O}_{1L}$ operator. The identified sense and antisense promoters correspond with DNA sequences that are similar to the consensus $\sigma_70$ binding motif (Supplementary Fig. 8).

All of the gates contain a ribosome insulator to normalize the 5'-UTR of the transcript. They are all based on variants of the Ribol insulator, which consists of a hammerhead ribozyme and a 23 nt hairpin added to stabilize ribozyme folding\textsuperscript{59}. While the Ribol sequence varies, this hairpin is always the same sequence, inside of which we discovered a strong antisense promoter at a $\sigma_70$ binding motif (Fig. 2c and Supplementary Fig. 9). Further, after the upstream promoter, there is a consistent drop in transcription (up to 12-fold) at this antisense promoter. Antisense promoters are known to attenuate RNAP flux from a forward promoter and may lead to better gate performance by lowering basal activity in the off state, which may explain why this hairpin emerged from a screen for parts that improve gate function\textsuperscript{67}. Note that this impacts the characterization of the gate response functions and how these data are used by design automation software to connect them (Supplementary Fig. 10).

Transcriptional attenuation is common in long operons, where the early dissociation of RNAP leads to incomplete mRNA transcripts\textsuperscript{47–75}. We define there being evidence in the profile for attenuation when >10-fold continuous decrease in RNAP flux along the gene sequence is observed. Most of the circuit genes do not show evidence for attenuation, the exceptions being $\text{hlyIR}$, $\text{bm3SR1}$, and $\text{yfp}$ (Fig. 2d and Supplementary Fig. 11). In particular, $\text{hlyIR}$ shows a continuous 46-fold decline over the length of the gene. There is an antisense promoter in this gene, but other repressor genes with stronger antisense promoters do not show attenuation. However, the ribosome profiling data show that this promoter also leads to cryptic antisense translation, which is not the case for any of the other cryptic antisense promoters (Fig. 2e). RNAP is coupled to ribosomes and this has been shown to lead to increased collisions with RNAP on the opposite strand, leading to greater attenuation\textsuperscript{76–78}.

The ribosome profiling data was analyzed to determine the impact of ribosome pausing within genes. We looked for a relationship with codon usage, noting that it has been observed that ribosomes can pause when they encounter rare codons or Shine Dalgarno (SD)-like sequences and this can reduce the expression level\textsuperscript{55,79–81}. Rare codons were eliminated when the synthetic repressor genes were codon optimized\textsuperscript{16}. Anecdotally, we noticed a rise in occupancy at the 3'-end of $\text{bm3SR1}$, which corresponds to strong predicted SD and start codon sequences (Supplementary Fig. 12). However, we could not find a systematic relationship between higher ribosome occupancy and codons that participate in SD-like sequences.

\[ J_{\text{RNAP}} = \gamma M_i \]
Parameterization of genetic parts. A complete picture of how parts function in the context of the circuit is illuminated by the RNA-seq and ribosome profiling datasets (Fig. 3, and Tables 1 and 2). Transcriptional parts, including promoters and terminators, can be parameterized using the RNA-seq profile. Parts that operate on the level of mRNA can also be characterized, such as the ribosome insulators. Ribosome occupancy can be used to calculate the RBS strengths. The transcription and translation of genes can also be determined. Collectively, all 54 genetic parts in the 3-input logic circuit can be fully parameterized with these two datasets for the entire circuit (Fig. 3b). There is a weak correlation between the measured and predicted strengths, 7% of the promoters fail in at least one state (are off when they should be on and vice versa), and the HlyIR gate fails completely (discussed below).

Previously, we presented methods to extract the promoter strength from the transcript profile by calculating the magnitude of the change before and after a TSS. However, the use of large fragment sizes caused biases at either end of the transcript—right at the TSS—and blurred the effects of neighboring promoters or promoters next to ribozymes (Supplementary Fig. 3). Data generated using the end-enriching method allow promoters to be identified automatically and their strength determined with fewer assumptions. The promoter strength is the increase in RNAP flux calculated as the difference between the transcript profile before (averaging positions −20 to −10) and after (averaging positions 10–20) the TSS (Fig. 3a).}

Errors in transcription

- Misannotated TSS
- Cryptic promoter (sense)
- Cryptic promoter (antisense)
- Translation attenuation
- Ribozyme malfunction
- Incorrect termination site

Errors in translation

- Translation at 5'UTR
- Alternative start codon
- Cryptic ORF (sense)
- Cryptic ORF (antisense)

Fig. 2 Unexpected transcription and translation. a A summary of transcriptional (top) and translational (bottom) errors, with respect to the designed circuit, are shown. The size of the arrows are accurate to the length of the cryptic ORFs. b The detection of TSSs in the sense (top) and antisense (bottom) directions are shown. When the derivative of the profile is larger than the cutoff required to mark a TSS (dashed lines), it is colored darker. The rpmR and amtR transcription units are shown for state +/−/− (IPTG/aTc/Ara). The dark and light gray filled regions represent transcription on sense and antisense strands, respectively. Cryptic promoter positions are marked by filled (sense strand) and empty (antisense strand) triangles. c The riboJ53 insulator is shown, marking the unexpected TSS (white triangle) and −10 motif. The sense and antisense profiles are shown in dark and light gray, respectively. The transcript profiles for the betI and hlyIR transcription units is shown (state +/+/−). d The dashed line marks the height of the profile at the beginning of the transcript to highlight attenuation. e The antisense cryptic promoter is marked and the antisense translation detected through ribosome profiling is shown. Source data are provided as a Source Data file.
Fig. 3 Complete characterization of genetic circuit parts. 

(a) The data are organized by the eight transcription units in the circuit. For promoters and terminators, the transcript profiles for all circuit states are shown, with the TSS and TTS marked with an arrow or T, respectively. The ribozyme CE and RBS TE values are shown for the states where the repressor is transcribed (average marked by the horizontal line). The estimated transcript and protein (RD) levels for the genes are shown for the eight induction states.

(b) The promoter activities for the promoters as measured by RNA-seq are compared to the values predicted by Cello. Data are shown for all 12 promoters within the circuit for the eight induction states. The points are shown in red if the promoter should be off and blue if it should be on.

(c) The translation initiation rates for the promoters are shown in red if the promoter should be off and blue if it should be on. The estimated transcript and protein (RD) levels for the genes are shown for the eight induction states.

(d) The terminator strengths of all eight terminators as measured using RNA-seq in their most highly transcribed state are compared to their previously measured values in isolation using fluorescence-based assays (data and error bars are from ref. 27, n ≥ 3 independent measurements on different days). Source data are provided as a Source Data file.
Table 1 Genetic part parameters.

| Promoters | Parameter | Units | y_{max} / y_{min} | y_{max} / y_{min} |
|-----------|-----------|-------|------------------|------------------|
| PAmtR     | 67        | 0.3188|                   |                   |
| PAmtR     | 7.4       | 0.0002|                   |                   |
| PBAI      | 4.8       | 0.0002|                   |                   |
| PBAI      | 6.0       | 0.0004|                   |                   |
| PBetI     | 7.5       | 0.0152|                   |                   |
| PBMX3R1   | 9.3       | 0.0235|                   |                   |
| PHlyIIR   | 49        | 0.0133|                   |                   |
| PPhlF     | 0.8       | 0.0002|                   |                   |
| PScrIR    | 0.9       | 0.0002|                   |                   |
| PScrR     | 4.9       | 0.0002|                   |                   |
| P+60      | 171       | 0.0030|                   |                   |
| P+60      | 109       | 0.0002|                   |                   |

Table 2 Empirical parameterization of gate response functions.

| Promoter | Parameter constant | Units | Protein # |
|----------|-------------------|-------|-----------|
| PAmtR    | 0.55              |       |           |
| PAmtR    | 2.24              |       |           |
| PBetI    | 0.12              |       |           |
| PBMX3R1  | 0.43              |       |           |
| PHlyIIR  | 9.80              |       |           |
| PPhlF    | 0.50              |       |           |
| PScrR    | 0.06              |       |           |

Surprisingly, despite these failures, the circuit is able to perform the logic operation for which it was designed.

Ribozymes are used to process the transcripts so that those from different TSSs have the same 5′-UTR and increase stability by decreasing RNase affinity69,82,83. Self-cleavage releases a small RNA that cannot be seen on the transcript profile when it is smaller than the average fragment size sequenced. To address this, we developed a method by which the ribozyme cleavage efficiency (CE) can be calculated from these data (Supplementary Fig. 13) (Methods). The cleavage efficiencies of most of the ribozymes are high and constant across circuit states, as expected, and are close to the values measured in vitro14 (Fig. 3a and Table 1). The exception is riboJ57 in the BetI gate, which we find to be nonfunctional under all conditions. This ribozyme differs from the others in that it is AU-rich and has a smaller folding ΔG and is thus vulnerable to context effects (Supplementary Fig. 14), and ribosome profiling also reveals it contains an open reading frame (Supplementary Fig. 12).

Ribozymes recruit ribosomes to the transcript and, indeed, there is an elevated ribosome occupancy in these regions (Figs. 1d and 3a)82,83. RBS strengths can be measured empirically by fusing them to a reporter gene85,86. However, their strength is highly context-dependent because of sensitivities to 5′-UTR folding and the translation rates of upstream open reading frames86,89. Computational models have been developed that capture these effects to predict RBS strength90,91. The translational efficiency (TE) of a gene is commonly estimated as the ratio of ribosome density and transcript level54. Here, we calculate the TE in absolute units (Methods), which is found to correlate with the translation initiation rate (TIR) predicted by the RBS Calculator (Fig. 3c and Supplementary Fig. 15) (Table 1)90,91.

Terminators block the progression of RNAP and their strength can be quantified as the magnitude of a drop in the transcript profile (Fig. 3a, Supplementary Fig. 16, and Table 1) (Methods). Previous methods to quantify terminators with RNA-seq data were limited by the large fragment lengths that caused a bias at the 3′-end8,35. This is corrected using shorter fragment libraries. These data reveal that the transcription termination site (TTS) does not occur at a precise position; rather, it is distributed across several nucleotides in the poly-T region27. The terminator strength T_{TTS} is defined as the ratio of the transcript level before the TTS to the level after it27. We developed algorithms to identify the TTS region and then calculate T_{TTS} using the same size and location for the windows as defined for the promoter calculation (Supplementary Fig. 16) (Methods). Previously, the T_{TTS}’s were measured for the terminators in isolation using fluorescent reporters27. Their quantitative performance differs and some terminators are much weaker in the circuit context (Fig. 3d).

Gate characterization in the context of a circuit. The circuit was designed by Cello using gate response functions that were measured independently14. This assumes that the gate will produce the same response in the context of the circuit. The response
functions are shown as lines in Fig. 4a. Using the RNA-seq data, this can be compared to the gate responses for the eight circuit states. The response functions are surprisingly similar when measured independently versus in the context of the circuit. Two exceptions are the BetI and HlyIIR gates (Fig. 4a). The RNAP fluxes from the gates in the context of the circuit are generally lower than when measured in isolation. This could be due to the shared ribosome and RNAP usage between gates\(^{25,92,93}\), although it does not appear to impact all of the gates equally.

The HlyIIR gate is noteworthy because it fails in nearly every way. Its response is weak and, even grossly, it produces the incorrect logic in some states (e.g., \(+/-/-\) should be on and \(-/+/-\) should be off) (Figs. 1d, 3a, and 4a). The individual parts show evidence for failures (Fig. 2a). Its RBS is the only one to have an elevated ribosome occupancy, and the terminator TTS is not in the poly-T region in most states. There is also a spike in ribosome occupancy of hlyIIR gene that occurs at a second in-frame ATG internal to hlyIIR that includes a predicted upstream RBS (Fig. 4b and Supplementary Fig. 12). Indeed, when the ATG at +1 is deleted, this has little impact on the response function of the HlyIIR gate (Fig. 4b). Finally, for all the repressors we looked for evidence of their expression impacting host genes. Only the hlyIIR repressor was found to impact host genes, notably aceBAK (glyoxylate cycle), tonB (iron uptake), and mipA (scaffolding protein) (Supplementary Fig. 17). Sequence analysis of these operons showed putative operators that resemble hlyIIR operator sequence (Supplementary Fig. 17). It is remarkable that, despite the complete gate failure, serendipitously this does not impact the circuit function.

Modeling the circuit robustness to parameter variation. We were intrigued that the four failed states of these two gates happen to not impact the performance of the overall circuit function. We first performed simple Boolean logic calculations showing the propagation of the states of the inputs through the network (Supplementary Fig. 18). This is a simple treatment, where the inputs are modeled as 0 or 1 and then each NOR and NOT gate responds digitally without additional parameterization. Using this framework, we modeled the experimentally observed failed states and propagated the effect to the final circuit state. We find that the particular circuit topology we use here is robust to all of these changes and the circuit logic that it is designed for is as expected. This might be interpreted as lucky, but it is important to note that we selected this circuit from a set of 60, of which only 45 worked as designed. One of the reasons that the remaining circuits failed may be that they were not similarly robust to the underlying part failures.

Next, we used a more detailed kinetic modeling to understand the circuit dynamics and evaluate the robustness to changes in the parameters. Typically, the problem with such a model is parameterization, where many of the circuit parameters have to be identified from the literature or estimated. In our case, we can obtain nearly all of the parameters for the parts through the RNA-seq and ribosome profiling experiments (Tables 1 and 2, and Supplementary Fig. 19). The time-dependent mRNA and protein levels for each gene was modeled using a set of ordinary differential equations (ODEs) that have parameters corresponding to the strengths of promoters, ribozymes, RBSs, and terminators as...
Fig. 5 Dynamic modelling of genetic circuit using -omics extracted parameters. a An example of the dynamics predicted when the circuit changes state. The differential equations used for the simulations are provided in the Methods. At t = 0, the inputs are switched from +/+/- to +/-/+ (IPTG/aTc/Ara) and the response is shown. b The predicted steady-state promoter activities of all seven gates in the circuit are compared against their measured activities across all eight induction states of the circuit (R² = 0.73, p value = 5.5 × 10⁻¹⁸). A line is drawn at x = y to guide the eye. c Sensitivity analysis. A circuit diagram showing all genetic parts used in the circuit is shown at the top. When each parameter is varied, all of the remaining parameters are held constant (Methods).

Regions in pink show the parameter range that lead to circuit that no longer performs the intended logic operation (circuit score <1). The parameters of the circuit extracted from -omics data are shown as dashed lines (Tables 1 and 2). Source data are provided as a Source Data file.

input parameters (Methods). The only parameters taken from the literature are the mRNA degradation rate and plasmid copy number, which are assumed to be constant, and the only non-omics derived parameter from this work is the growth rate μ. Despite the parameterization being based on static measurements, the equations can be used to predict the circuit dynamics; the response to a change in inputs that leads to a glitch is shown in Fig. 5a. The steady-state solutions are determined and found to match all the circuit states, including all of the internal sensor/gate output promoters (Fig. 5b and Supplementary Fig. 20). This demonstrates that the parameters extracted from -omics experiments can be used in conjunction with ODE modeling to predict system performance. Note, however, that discrepancies in the circuit prediction (Fig. 5b) cause the circuit score predicted by the kinetic simulations (2) to be much lower than the values predicted by Cello (80) or measured experimentally (21).

A sensitivity analysis was conducted for each genetic parameter to identify the operational boundaries in which the circuit remains functional, inspired by earlier work on the parameter robustness of natural regulatory networks (Fig. 5c)⁹⁵–⁹⁷. The circuit score was defined as the lowest expected on state, divided by the highest expected off state in the simulation (Methods). When this score is <1, the circuit is no longer performing the logic operation for which it was designed. Each parameter was varied individually, the eight induction states simulated, and the circuit score calculated. The plots for the robustness of each parameter are shown in Fig. 5c. The starting values are shown as vertical dashed lines for reference.

There are striking differences in the robustness of the circuit design to perturbations in the underlying parts. For example, the PBAD promoter can vary in strength over orders of magnitude with little impact, but the system is very sensitive to the strength of the PTac promoter. Similarly, the gate promoters show variability in sensitivity, both in the maximum of their response functions and their thresholds. The circuits are insensitive to ribozyme efficiencies, which is not surprising given their primary function is to standardize the mRNAs against errors due to S²⁻UTR differences, which are not captured in the model. More surprisingly, there is large robustness to terminator efficiency, where most can be varied over order of magnitude with little impact. The particular terminators that are robust versus sensitive are a function of the circuit topology as well as the particular organization of gates on the DNA sequence.

The cellular impact of circuit RNAP/ribosome usage. The RNA-seq and ribosome profiling datasets are then used to calculate the transcriptional and translational costs of holding a gate in a particular state (Fig. 4a). The input to a gate is in units of RNAP/s, representing the usage of transcriptional resources. The number of ribosomes used in each gate can also be calculated by summing all ribosomes used in each gate can also be calculated by summing all ribosomes used in each gate. The predicted steady-state promoter activities of all seven gates in the circuit are compared against their measured activities across all eight induction states of the circuit (R² = 0.73, p value = 5.5 × 10⁻¹⁸). A line is drawn at x = y to guide the eye. c Sensitivity analysis. A circuit diagram showing all genetic parts used in the circuit is shown at the top. When each parameter is varied, all of the remaining parameters are held constant (Methods). Regions in pink show the parameter range that lead to circuit that no longer performs the intended logic operation (circuit score <1). The parameters of the circuit extracted from -omics data are shown as dashed lines (Tables 1 and 2). Source data are provided as a Source Data file.

Fig. 5 Dynamic modelling of genetic circuit using -omics extracted parameters. a An example of the dynamics predicted when the circuit changes state. The differential equations used for the simulations are provided in the Methods. At t = 0, the inputs are switched from +/+/- to +/-/+ (IPTG/aTc/Ara) and the response is shown. b The predicted steady-state promoter activities of all seven gates in the circuit are compared against their measured activities across all eight induction states of the circuit (R² = 0.73, p value = 5.5 × 10⁻¹⁸). A line is drawn at x = y to guide the eye. c Sensitivity analysis. A circuit diagram showing all genetic parts used in the circuit is shown at the top. When each parameter is varied, all of the remaining parameters are held constant (Methods). Regions in pink show the parameter range that lead to circuit that no longer performs the intended logic operation (circuit score <1). The parameters of the circuit extracted from -omics data are shown as dashed lines (Tables 1 and 2). Source data are provided as a Source Data file.

There are striking differences in the robustness of the circuit design to perturbations in the underlying parts. For example, the PBAD promoter can vary in strength over orders of magnitude with little impact, but the system is very sensitive to the strength of the PTac promoter. Similarly, the gate promoters show variability in sensitivity, both in the maximum of their response functions and their thresholds. The circuits are insensitive to ribozyme efficiencies, which is not surprising given their primary function is to standardize the mRNAs against errors due to S²⁻UTR differences, which are not captured in the model. More surprisingly, there is large robustness to terminator efficiency, where most can be varied over order of magnitude with little impact. The particular terminators that are robust versus sensitive are a function of the circuit topology as well as the particular organization of gates on the DNA sequence.

The cellular impact of circuit RNAP/ribosome usage. The RNA-seq and ribosome profiling datasets are then used to calculate the transcriptional and translational costs of holding a gate in a particular state (Fig. 4a). The input to a gate is in units of RNAP/s, representing the usage of transcriptional resources. The number of ribosomes used in each gate can also be calculated by summing all the ribosome occupancy across the gate’s repressor gene (Fig. 4a). While the gates have a similar design, they vary in the resources required. For example, the PhIF gate (in our hands, a reliable one) requires low input JRNAP to turn off and <30 ribosomes to operate. In contrast, the SrpR and AmeR gates require high JRNAP to turn off
and up to 300 ribosomes. A variety of biophysical factors could explain these differences, including the repressor affinities and translational efficiency of transcripts36,99.

Cells possess a finite pool of cellular resources, and transcription and translation of circuit proteins compete for these resources with the cell’s native genes100–102. Carrying the circuit slows growth, the magnitude of which depends on the circuit state (Fig. 6a).

RNAP flux is analogous to the total power usage by the circuit and this varies based on the pattern of transcriptional activity of a state, ranging from 107 to 816 RNAP/s (Figs. 1d and 6a). The estimated fraction of the cellular proteome occupied by circuit proteins can be estimated using the ribosome profiling data (Methods). Depending on the state, up to 4.3% of the proteome is comprised of circuit proteins. This fraction can be estimated by comparing the amount of RNA polymerase and ribosome genes for the eight states of the circuit based on the estimated fraction of the proteome occupied by circuit proteins for each state (Fig. 6f).

Next, we measured the impact on host gene expression as the circuit consumes more resources. First, we hypothesized that the cell could try to compensate by upregulating RNAP or ribosome expression. We found no upregulation of ribosome or RNAP genes, individually or in combination (Fig. 6c). We then systematically analyzed all chromosomal genes and pathways, looking for positive or negative correlations in their expression (RDs) with the estimated fraction of the proteome occupied by RNAP or ribosome genes for the eight states of the circuit based on the estimated fraction of the proteome occupied by circuit proteins for that state (part a). Data for the control are also shown for comparison. The RNAP values represent the sum of the rpoABCDD genes and the ribosome values include the sum of rpsABCDGHIJKLMNOPQRSTU, rpmABCDEFGHJKLMNOPQRSTUVWXYZ, and rpmABCDEFGHJUI. The responses of two housekeeping proteases are shown. The responses of two housekeeping proteases are shown. The response of RNase III is shown. The line is the best fit ($R^2 = 0.67$, $p$-value = 0.007). The nine data points in d, e, and f represent the eight induction states of the circuit as well as the control, as described in c. Source data are provided as a Source Data file.
Discussion

Computer aided design (CAD) helps the management of large genetic engineering projects that require the selection of many genetic parts to build a system\textsuperscript{3,5,14,103–105}. Parts can be selected from databases of previously characterized sequences or designed de novo with computational tools. However, once parts are used in a genetic system, it has been difficult to discern whether they are functioning as intended. Systems are evaluated based on their gross function, such as a circuit function or the production of a small molecule. Tools to interrogate genetic systems have to be able to scale to larger designs, with more parts, without incurring additional cost or experimentation.

Deep sequencing scales appropriately and here we have combined methods that enumerate mRNAs and ribosome usage to fully characterize the parts in a system in a single experiment. While genetic circuits require the evaluation of multiple states, these samples can easily be pooled in a single sequencing lane. Solely using RNA-seq and ribosome profiling requires some parameters to be pulled from other sources, including degradation rates, plasmid copy number, and total number of active ribosomes, all of which are not constant and depend on environmental conditions and the growth rate\textsuperscript{28,30,64,106,107}. Beyond these techniques, the rapid decline in deep sequencing costs has led to a suite of techniques that could be performed simultaneously to offer further insight into the workings of an engineered system in a cell\textsuperscript{108}. New sequencing methods have been developed for the detection of TSSs, mRNA degradation, RNA structures, RNAS elongation, transcription dynamics, enzyme activity, metabolomics, and translational initiation/elongation/termination\textsuperscript{109–123}. The single-cell resolution obtained from fluorescent reporters and flow cytometry and microscopy\textsuperscript{124} is currently lost, but this could be recovered as single-cell RNA-seq experiments become more practical and they are able to provide the full transcript profiles, rather than just the mRNA levels of genes\textsuperscript{125–127}. Automating these techniques in experimentation and data handling will be a powerful debugging tool.

Here, we chose one of the most well-behaved circuits we have constructed to date. We were shocked at the number of internal failures. Many parts do not function as designed, bring unintended regulatory sequences in their DNA (e.g., cryptic promoters) or introduce new functions due to the new sequences formed when two parts are connected. Remarkably, a gate (HlyIIIR) fails entirely. Indeed, there is some luck in our selection of a circuit whose topology is robust to the gate and parameter variations observed experimentally. However, if these same parts are used in a different system, they could cause it to fail, the probability of which increases as the size of the project grows. This is reminiscent of an earlier observation by Leibler and co-workers when they randomly combined regulatory parts and noticed that the circuit often does not behave as can be explained by the known part functions\textsuperscript{2}. Similar effects are seen in the combinatorial assembly of metabolic pathways\textsuperscript{35,128}.

Haphazardly finding that systems work or fail for unknown reasons leads to the mystic of biology being labeled unpredictable when in reality it is only our inability to observe how a design actually works in a cell that inhibits precision engineering. This work focuses on the analysis of a complete circuit where the individual parts were not pre-characterized using—omics tools and the design algorithm did not consider this type of information. One can imagine going in after the fact and mutating problems, like cryptic promoters, to fix them or to replace malfunctioning parts. However, it would be more valuable to run—omics experiments routinely on the subsystems to correct problems prior to their use, particularly when the intention is to reuse them in many permutations (e.g., in the Cello User Constraint File). Gates could also be removed from this list completely, as we have now done for HlyIIIR. The careful integrative steps of design and debugging will lead to their more reliable assembly later. In addition, the characterization of the total RNAS and ribosomes used by a gate allows for the design of low power variants that minimize the impact on the cell. This will lead to more reliable CAD in genetic engineering that can predict failure modes and warn a user if there is a potential growth defect or the time expected before there is evolutionary breakdown.

Methods

Strain, media, and inoculants. The Escherichia coli DH10B derivative NEB 10-beta (Str-lev = 7679 araD139 hisG46 lacX74 galK16 galE15 c18-480lacZ2M15 recA1 relA1 endA1 nupG rpsL (StrR) rph spoT1) was used as the host (New England Biolabs, MA, C3019). Cells were grown in M9 minimal media, consisting of 1× M9 media salts (Sigma–Aldrich, MO, M6030), 0.34 g/l thiamine hydrochloride (Sigma–Aldrich, MO, T4625), 0.4% D-glucose (Sigma–Aldrich, MO, G8270), 0.2% Casamino acids (Acros, NJ, AC612045000), 2 mM MgSO\textsubscript{4} (Sigma–Aldrich, MO, 230391), and 0.1 mM CaCl\textsubscript{2} (Sigma–Aldrich, MO, 449709). The inducers used in this study were isopropyl β-D-1-thiogalactopyranoside (IPTG; Sigma–Aldrich, MO, I6758), and hydroxytretonine hydrochloride (αTc; Sigma–Aldrich, MO, 37919), and L-arabinose (Ara; Sigma–Aldrich, MO, A3256). Antibiotic selections were performed with 50 μg/ml kanamycin (Gold Biotechnology, MO, K-120-5) and 50 μg/ml spectinomycin (Gold Biotechnology, MO, S-140-5). Phosphate buffered saline (1x PBS) solution was prepared from a 10x solution purchased from Omnipur (MA, 6505).

Circuit induction. \textit{E. coli} was co-transformed with the pAN3945 circuit (circuit) and pAN4023 (output) plasmids\textsuperscript{129} (Supplementary Fig. 1). Individual colonies were picked and inoculated into 5 ml M9 minimal media with kanamycin and spectinomycin in a 14 ml culture tube (Corning, MA, 352059). The culture was grown overnight for 16 h at 37 °C and 250 rpm in an Innova 44 shaker (Eppendorf, CT). The following day, cultures were diluted 23 μl into 4 ml fresh M9 media with antibiotics and grown under the same conditions for 3 h. For induction, cells were diluted a second time 380 μl into 250 ml M9 media with antibiotics in 3 L Erlenmeyer flasks (Pyrex) in the presence or absence of 1 mM IPTG, 2 ng/ml (4.32 nM) αTc, and 5 mM Ara. Culture flasks were incubated at 37 °C and 200 rpm for 5–7 h. At OD\textsubscript{590} = 0.3, cells were collected, rapidly filtered at room temperature by passing through a 0.22 μm pore size nitrocellulose filter (Sigma–Aldrich, MO, N8645), and carefully scraped from the surface of the filter using a stainless-steel scoopula (Thermo Fisher Scientific, USA, 14-357Q) and flash frozen immediately in liquid nitrogen. Cell pellets were saved in −80 °C until use in RNA-seq and ribosome profiling steps.

Fluorescence measurement. Individual colonies of wild-type \textit{E. coli} as well as \textit{E. coli} cells harboring circuit plasmids (pAN3945 and pAN4023) and \textit{E. coli} cells harboring a reference plasmid (pBBR1SUP) (Supplementary Fig. 1) were picked and inoculated into 200 µl M9 minimal media with kanamycin and spectinomycin in a 1 ml culture tube (Corning, MA, 352059). The cell activity of a circuit \textit{s} output was measured using flow cytometry.

Flow-cytometry analysis. Fluorescence was measured using an LSRII Fortessa flow cytometer (BD Biosciences, San Jose, CA) using BD FACSDiva software, with the following parameters: 80% 488 nm, 3% 440 nm, 0.3% 540 nm, 0.2% 570 nm, 0.1% 620 nm, 0.1% 670 nm. The flow cytometry software FlowJo, version 9 (TreeStar, Inc., Ashland, OR) was used to gate the raw data and calculate the median YFP fluorescence values for the gated events (Supplementary Fig. 23).

Conversion to relative promoter units (RPU). The activity of a circuit’s output promoter (P\textsubscript{DNA17}) can be calculated in RPU by comparing circuit’s fluorescence to the fluorescence of a reference plasmid (pBBR1SUP). The representative fluorescent reference promoter (P\textsubscript{DNA17}) in RPU: <YFP > = <YFP >\textsubscript{0} + <YFP >\textsubscript{circuit} - <YFP >\textsubscript{0}. <YFP >\textsubscript{circuit} is the median fluorescence of circuit at each induction state, <YFP >\textsubscript{0} is the median fluorescence of circuit at each induction state.
Cell predictions. Cello version 1.0 was used with the Eco1C1G1T1 UCF (http://cellocad.org/). The sensor activities used were (on and off): PTac (0.25 μM), Cytosine (10 μM), and mCherry (1μM). The total RNA-seq was amplified with a Konka (VT-1) dye PCR system with a 10 cycles program (72°C for 15 s, 70°C for 15 s, and 68°C for 3 min). The purified RNA was labeled with a sequencing primer C6 (CS010012), and the libraries were generated using the Illumina TruSeq PCR-free Library Kit (Illumina, USA). The libraries were sequenced on an Illumina HiSeq 2500 in rapid-run mode using sequencing primer CGACAAGGTTTCTACATCGTATC (15 cycles). The reads were then mapped to the reference genome using Bowtie (v2.2.1). The raw reads were filtered by removing adapter sequences and poor quality reads. The reads were then filtered by removing reads with high percentages of mismatches.

Conversion of RPU to RNAP/s. The RNAP flux can be reported as either the total RNAP flux for all copies of a promoter, or the average RNAP flux for all the copies of a promoter, or the average RNAP flux for all the copies of a promoter, or the average RNAP flux for all the copies of a promoter.

RiboSequencing. As described above, the lysate cell was spliced to be used for both RNA-seq and ribosome profiling steps. The following steps are performed on the portion saved for ribosome profiling. The total RNA was extracted from the lysate using TRIzol reagents. The RNA-seq libraries were generated using the Illumina TruSeq PCR-free Library Kit (Illumina, USA). The libraries were sequenced on an Illumina HiSeq 2500 in rapid-run mode using sequencing primer CGACAAGGTTTCTACATCGTATC. The reads were then mapped to the reference genome using Bowtie (v2.2.1). The raw reads were filtered by removing adapter sequences and poor quality reads. The reads were then filtered by removing reads with high percentages of mismatches.

Conversion of RPU to RNAP/s. The RNAP flux can be reported as either the total RNAP flux for all copies of a promoter, or the average RNAP flux for all the copies of a promoter, or the average RNAP flux for all the copies of a promoter.

RiboSequencing. As described above, the lysate cell was spliced to be used for both RNA-seq and ribosome profiling steps. The following steps are performed on the portion saved for ribosome profiling. The total RNA was extracted from the lysate using TRIzol reagents. The RNA-seq libraries were generated using the Illumina TruSeq PCR-free Library Kit (Illumina, USA). The libraries were sequenced on an Illumina HiSeq 2500 in rapid-run mode using sequencing primer CGACAAGGTTTCTACATCGTATC. The reads were then mapped to the reference genome using Bowtie (v2.2.1). The raw reads were filtered by removing adapter sequences and poor quality reads. The reads were then filtered by removing reads with high percentages of mismatches.

Conversion of RPU to RNAP/s. The RNAP flux can be reported as either the total RNAP flux for all copies of a promoter, or the average RNAP flux for all the copies of a promoter, or the average RNAP flux for all the copies of a promoter.

RiboSequencing. As described above, the lysate cell was spliced to be used for both RNA-seq and ribosome profiling steps. The following steps are performed on the portion saved for ribosome profiling. The total RNA was extracted from the lysate using TRIzol reagents. The RNA-seq libraries were generated using the Illumina TruSeq PCR-free Library Kit (Illumina, USA). The libraries were sequenced on an Illumina HiSeq 2500 in rapid-run mode using sequencing primer CGACAAGGTTTCTACATCGTATC. The reads were then mapped to the reference genome using Bowtie (v2.2.1). The raw reads were filtered by removing adapter sequences and poor quality reads. The reads were then filtered by removing reads with high percentages of mismatches.

Conversion of RPU to RNAP/s. The RNAP flux can be reported as either the total RNAP flux for all copies of a promoter, or the average RNAP flux for all the copies of a promoter, or the average RNAP flux for all the copies of a promoter.

RiboSequencing. As described above, the lysate cell was spliced to be used for both RNA-seq and ribosome profiling steps. The following steps are performed on the portion saved for ribosome profiling. The total RNA was extracted from the lysate using TRIzol reagents. The RNA-seq libraries were generated using the Illumina TruSeq PCR-free Library Kit (Illumina, USA). The libraries were sequenced on an Illumina HiSeq 2500 in rapid-run mode using sequencing primer CGACAAGGTTTCTACATCGTATC. The reads were then mapped to the reference genome using Bowtie (v2.2.1). The raw reads were filtered by removing adapter sequences and poor quality reads. The reads were then filtered by removing reads with high percentages of mismatches.

Conversion of RPU to RNAP/s. The RNAP flux can be reported as either the total RNAP flux for all copies of a promoter, or the average RNAP flux for all the copies of a promoter, or the average RNAP flux for all the copies of a promoter.

RiboSequencing. As described above, the lysate cell was spliced to be used for both RNA-seq and ribosome profiling steps. The following steps are performed on the portion saved for ribosome profiling. The total RNA was extracted from the lysate using TRIzol reagents. The RNA-seq libraries were generated using the Illumina TruSeq PCR-free Library Kit (Illumina, USA). The libraries were sequenced on an Illumina HiSeq 2500 in rapid-run mode using sequencing primer CGACAAGGTTTCTACATCGTATC. The reads were then mapped to the reference genome using Bowtie (v2.2.1). The raw reads were filtered by removing adapter sequences and poor quality reads. The reads were then filtered by removing reads with high percentages of mismatches.

Conversion of RPU to RNAP/s. The RNAP flux can be reported as either the total RNAP flux for all copies of a promoter, or the average RNAP flux for all the copies of a promoter, or the average RNAP flux for all the copies of a promoter.

RiboSequencing. As described above, the lysate cell was spliced to be used for both RNA-seq and ribosome profiling steps. The following steps are performed on the portion saved for ribosome profiling. The total RNA was extracted from the lysate using TRIzol reagents. The RNA-seq libraries were generated using the Illumina TruSeq PCR-free Library Kit (Illumina, USA). The libraries were sequenced on an Illumina HiSeq 2500 in rapid-run mode using sequencing primer CGACAAGGTTTCTACATCGTATC. The reads were then mapped to the reference genome using Bowtie (v2.2.1). The raw reads were filtered by removing adapter sequences and poor quality reads. The reads were then filtered by removing reads with high percentages of mismatches.

Conversion of RPU to RNAP/s. The RNAP flux can be reported as either the total RNAP flux for all copies of a promoter, or the average RNAP flux for all the copies of a promoter, or the average RNAP flux for all the copies of a promoter.

RiboSequencing. As described above, the lysate cell was spliced to be used for both RNA-seq and ribosome profiling steps. The following steps are performed on the portion saved for ribosome profiling. The total RNA was extracted from the lysate using TRIzol reagents. The RNA-seq libraries were generated using the Illumina TruSeq PCR-free Library Kit (Illumina, USA). The libraries were sequenced on an Illumina HiSeq 2500 in rapid-run mode using sequencing primer CGACAAGGTTTCTACATCGTATC. The reads were then mapped to the reference genome using Bowtie (v2.2.1). The raw reads were filtered by removing adapter sequences and poor quality reads. The reads were then filtered by removing reads with high percentages of mismatches.

Conversion of RPU to RNAP/s. The RNAP flux can be reported as either the total RNAP flux for all copies of a promoter, or the average RNAP flux for all the copies of a promoter, or the average RNAP flux for all the copies of a promoter.

RiboSequencing. As described above, the lysate cell was spliced to be used for both RNA-seq and ribosome profiling steps. The following steps are performed on the portion saved for ribosome profiling. The total RNA was extracted from the lysate using TRIzol reagents. The RNA-seq libraries were generated using the Illumina TruSeq PCR-free Library Kit (Illumina, USA). The libraries were sequenced on an Illumina HiSeq 2500 in rapid-run mode using sequencing primer CGACAAGGTTTCTACATCGTATC. The reads were then mapped to the reference genome using Bowtie (v2.2.1). The raw reads were filtered by removing adapter sequences and poor quality reads. The reads were then filtered by removing reads with high percentages of mismatches.

Conversion of RPU to RNAP/s. The RNAP flux can be reported as either the total RNAP flux for all copies of a promoter, or the average RNAP flux for all the copies of a promoter, or the average RNAP flux for all the copies of a promoter.
rapid-run mode using sequencing primer CGACAGGTTCCAGAATCCTCAGCGACGAGATC. Similar to RNA-seq data processing steps, adaptors (CTGATGACGACATCAATATCTGATACACGCTCTTAGCTCAGTGC) were first trimmed off the raw reads acquired from ribosome profiling runs using Cutadapt[30] (read length > 10 nt). These trimmed reads were then aligned to the same indexed reference sequences generated for RNA-seq. Alignments were performed using the ‘bowtie’ function of Bowtie 1.2.3 with parameters ‘-v1 -m2 -k1’. The output files stored all necessary information about the aligned reads (reference, position, strand, and mapping quality) to be used for the ribosome occupancy profile generation.  

Construction of ribosome occupancy profile. A center-weighting approach was taken to map the P-site position of ribosomes on all aligned footprint reads ranging from 23 to 42 nucleotides in length. For each aligned footprint read, 11 nucleotides on either end were removed, and the nucleotide counts in the remaining region were given the same score, normalized by the length of this central region[34]. Therefore, the ribosome occupancy at each nucleotide position is the sum of all the center-weighted scores from all the reads aligned to that nucleotide. The ribosome occupancy at each nucleotide was then normalized by the sum of all ribosome occupancies across DNA sequences in that sample (including genome and circuit), and then multiplied by the total number of active ribosomes in E. coli which is a function of cell growth rate (20,000 at 45 min doubling time[36]). The resulting ribosome occupancy profile has an absolute unit which is the number of ribosomes at each position. Contrary to RNA-seq, applying tRNA correction was not needed, as ribosome occupancies and therefore tRNA contamination is negligible. We also generated a profile of ribosome occupancy per transcript by dividing the ribosome occupancy at each nucleotide position by the amount of RNAP flow (in unit of RNAP/s) at the same position, multiplied by mRNA degradation rate (γ = 0.0067 s−1[36]). The resulting profile has normalized (number of ribosomes) at each nucleotide position x is shown in Supplementary Fig. 25 for all states of the circuit on both sense and antisense strands. The ribosome density (RD) of each gene was calculated by averaging all ribosome occupancies over the length of the gene, in absolute units. However, three additional normalizations were needed to calculate RDs (following approaches by Li and of translation initiation and termination on ribosome density, the first and last five codons of each gene were excluded from averaging calculation, (2) to account for the elevation of ribosome occupancies at the beginning of the genes, an exponential decay function was fitted to a genome-wide occupancy profile and this fitted function was used to normalize the ribosome occupancy at each nucleotide of a given gene, and (3) to reduce the effect of outlier ribosome occupancies, the top and bottom 5% of the ribosome occupancies over the length of the gene were removed (90% winsorization) if the average read density on a gene was higher than one.

TSS/TTS identification. Transcriptional start sites (TSSs) across DNA sequences are identified by calculating the ratio of RNAP flow JRNAP between each neighboring positions, JRNAP(f+1)/JRNAP(f). For TSSs in the reverse strand, the inverse of the above ratio is calculated. We also applied a threshold of ratio > 5 to call a TSS. As a limitation to our method, for promoters located immediately upstream of a ribozyme, the identified TSS is the same as ribozyme cleavage site, since the cleaved RNA product is ~10 nt long and is lost during RNA-seq sample preparation, creating a sharp increase in RNAP flow profile at the cleavage site.

Transcriptional termination sites (TTSs) are identified by calculating an averaged-window ratio AWR(x) profile for each nucleotide position x

\[
AWR(x) = \frac{\sum_{n=1}^{k} J(n)}{\sum_{n=1}^{k} J(n)}
\]

where k = 10 is the length of the averaging window. For reverse strand, AWR(x) is the inverse of Eq. 2. Similar to promoters, we applied a threshold AWR(x) > 5. Because of the gradual decrease in RNAP flux at the termination sites, a continuous span of multiple nucleotides can pass this threshold. To identify the most probable termination site within this span, position x with maximum AWR(x) value is marked as transcriptional termination site TTS.

Characterization of the Δ(ATG) HlyIIR gate. The response function was characterized as previously described[35]. A version of the pHlyIIR-H1 plasmid[14] was constructed (pHlyIIR-H1-ΔATG, Supplementary Fig. 1 and Supplementary Table 4) where Gibson assembly was used to delete the start codon Δ(ATG) of hlyIIR. The original and mutated hlyIIR plasmids were transformed separately into E. coli DH10B. In addition, a pAN1818 plasmid[44] carrying an IPTG-inducible promoter Prps, for yfp transcription (Supplementary Fig. 1) was also used to transform E. coli DH10B. Single colonies of these three strains as well as an E. coli strain carrying the reference RPU plasmid (pSBSRPU, Supplementary Fig. 1) were inoculated into 200 μl of M9 media with kanamycin in Nunco™96-well plates and grown for 16 h at 37°C and 1000 rpm in an ELMi plate shaker. On the next day, cells were diluted 178-fold (two serial dilutions of 15 μl into 185 μl) into M9 media with kanamycin, and were incubated for 3 h at 37°C and 1000 rpm in an ELMi plate shaker using Nunco™96-well plates. Cells were then diluted 68-fold (two serial dilutions of 15 μl into 185 μl, and then 3 μl into 145 μl) to M9 media with kanamycin, and were induced with varying concentrations of IPTG (0, 5, 10, 20, 30, 40, 50, 70, 100, 150, 200, and 1000 μM). After induction, cells were grown in Nunco™96-well plates for 5 h at 37°C and 1000 rpm in an ELMi plate shaker, and then 10 μl of culture was added to 120 μl of 1x PBS solution with 2 mg/ml kanamycin. This mixture was incubated for 1 h and fluorescence was measured using flow cytometry and converted to RPU (above). At each IPTG concentration, the measured RPU of pAN1818 plasmid represents the input promoter activity s to the both original and mutated HlyIIR gates, while the measured RPU of either hlyIIR plasmids is the output promoter activity y of the gates. The output activity y of each gate was plotted as a function of the input activity s for all IPTG inductions, and a Hill function was fit to data according to Eq. 1. Fitting was performed in Python with minimize function of the SciPy module.

Promoter strength calculation. To characterize promoter strength Δf, the first TSS of each promoter was identified (above). Using the identified TSS, the promoter activity is calculated as

\[
\delta f = \frac{1}{\pi} \sum_{n=1}^{\Delta X_{TSS}} J(n) - \frac{1}{\pi} \sum_{n=\Delta X_{TSS}}^{\Delta X_{YSS}} J(n)
\]

where j(x) is RNAP flux at position i, n = 10 is the length of averaging window, and A = 10 is the gap between TSS and averaging window. For promoters in reverse strand, the δf value calculated from Eq. 3 is multiplied by −1.

Calculation of ribosome activity. The cleavage site of each ribozyme was identified as a result of TSS calculation (above) for the promoter located immediately upstream of each ribozyme. Next, for each ribozyme, all the cut and uncut fragments around the cleavage site were identified. These are the raw RNA-seq fragment sequences used to generate transcription profiles with. Cut fragments are those that either end exactly at the cleavage site or begin with it. Uncut fragments are those that overlap the cleavage site (Supplementary Fig. 14). The total number of cut fragments downstream of the cleavage site Funcut was quantified as well as the total number of uncut fragments Fmature. Ribosome cleavage efficiency is calculated as CE = Funcut/(Funcut + Fmature).

Calculation of RBS strength. The translation efficiency (TE) of an RBS was calculated using the ribosome density (RD) and the steady-state mRNA expression level of the gene associated with the RBS. The number of mRNA transcripts at steady-state (mRNA) is calculated by averaging RNAP flux profile (in unit of RNAP/s) over a 10 nt window at the end of each gene (~10 to 0), divided by mRNA degradation rate (0.0067 s−1). Note that for amrB gene, the averaging window was shifted 150 nucleotides into the coding section (~160 to ~150) to avoid the RNAP flux by an internal cryptic promoter. Next, translation efficiency is calculated as TE = (RD × ω)/mRNA, where ω = 15 s−1 is the ribosome elongation rate[107]. TE represents the number of ribosomes that finish translation and produce a new molecule of protein per mRNA transcript per second.

RBS Calculator and mRNA secondary structure prediction. The translation initiation rate (TIR) for each RBS sequence was predicted using RBS Calculator (v2.0) (https://salislab.net/software/predict_rbs_calculator)[90,91]. The total Gibbs free energy change ΔGtotal is calculated and used to determine TIR = Kρ−ΔGtotal, where ρ = 0.45 mol/kcal is the Boltzmann constant, K = 2500 is a constant, and ΔGtotal is the total energy difference between ribosome-free and ribosome-bound mRNA. For all calculations, 35 nucleotides before and after start codon was used as mRNA sequence and E. coli’s 16 S RNA (5’-ACCUCCUA-3’) was used as anti-Shine-Dalgarno sequence.

RNA folding Gibbs free energy (ΔG) and RNA secondary structure of ribozyme insulators were calculated using ViennaRNA software (version 1.8.5, default options)[132]. For ΔG calculations, the full sequence of each ribozyme was used. RNA secondary structures were also predicted for two ribozymes (sarf and ribo57) where their transcription started from an upstream promoter adding 63 extra nucleotides to the 5’-UTR sequence (Supplementary Fig. 14).

Calculation of mRNA abundance. The number of full length mRNA molecules for each gene at steady state (mRNA) is calculated as the ratio of mRNA transcription rate (in unit of RNAP/s) and mRNA degradation rate (above).

Calculation of terminator strength. First, the TTS was identified following the averaged-window ratio (AWR) method described above. Δf is defined as the fold-decrease in RNAP flux before and after a TTS as

\[
\delta f = \frac{\sum_{n=1}^{\Delta X_{TSS}} J(n)}{\sum_{n=\Delta X_{TSS}}^{\Delta X_{YSS}} J(n)}
\]

where j(x) is RNAP flux at position i, n = 10 is the length of averaging window, and A = 10 is the gap between TTS and averaging window. For terminators in reverse strand, the δf value calculated from Eq. 4 is inverted.
Calculation of ribosome usage. The ribosome usage is defined by the number of ribosomes sequenced by a gene at steady state, and is calculated by summing the height of ribosome occupancy profile over the length of the gene.

Calculation of proteome fraction. Ribosome profiling data are used to estimate the fraction of total cellular proteome $\Phi(i)$ that is occupied by a particular gene product $i$ as

$$\Phi(i) = \frac{\text{RD}_{i} \cdot \text{MW}_{i}}{\sum \text{RD}_{i} \cdot \text{MW}_{i}}$$

where $\text{RD}_{i}$ is ribosome density of gene $i$ and $\text{MW}_{i}$ is the molecular weight of its protein product. The estimated proteome fraction of the circuit at each induction state is calculated by summing the proteome fraction of its regulatory proteins (repressors and sensors) and $y$-ff reporter. This value was compared against the estimated total proteome fraction of all DNA-binding regulatory proteins in E. coli genome. A list of 302 unique DNA-binding proteins in E. coli MG1655 were identified using EcoCyc database (https://ecocyc.org/), from which 283 were found to be present in E. coli DH10B. In addition, the impact of estimated circuit proteome fraction on cellular metabolism (below), RNAP and ribosome expression, and cellular maintenance for proteins and mRNAs was studied. To calculate the RD of ribosomes, the RD of all 54 ribosomal proteins in E. coli DH10B (pAABCDEFGHJKLMNOPQRSTUVWXYZ in 30 S and pAABCDEFGLMNOQP- STUVWXYZ and pmpABCDEFGHIJ) in 50 S subunits were summed as an estimate for the RD of ribosomes. Similarly, RD of RNAP was estimated as the sum of the RD of rpoBC, RD of rpoD (σ70), and half of the RD of rpoH due to its 2:1 stoichiometry with RNAP. Finally, the average RD of proteins involved in the unfolded protein response of E. coli (hspG, dnaK, groES/L, grpE, hulOUV, yegD, ybbN, secB, taken from EcoCyc) was calculated and correlated in the change in the estimated circuit proteome fraction across induction states.

EdgeR and gene ontology analysis. To study the impact of estimated circuit proteome fraction on cellular metabolism, first, a list of up and downregulated genes was generated by comparing the ribosome density (RD) of each gene in the cell containing the circuit with that of the control cell. RD was used as the estimate of protein expression level. We calculated these differences for all eight induction states. To do that, differential gene expression analysis was performed using EdgeR, where its ‘exact test’ analysis was performed in R (version 3.4.4) and adjusted $p$-values were calculated using build-in false discovery rate (FDR) correction. EdgeR requires biological replicates for its exact test analysis in order to estimate the dispersion in data (dispersion = BCV) where BCV is the biological coefficient of variation). A fixed BCV = 0.1 was used in all exact test analysis. Genes >2-fold change in expression and with $p$ value and FDR less than 0.05 were selected as differentially expressed genes. These up and downregulated genes were used as the input for Gene Ontology analysis (GO) and identifications a list of regulated pathways based on their co-regulated gene members. Each regulated pathway was evaluated for potential relationship with estimated circuit proteome fraction by cross induction state correlation between RD of each gene member in the pathway and the estimated circuit proteome fraction across induction states. Not all pathways were correlated. Only those pathways/gens with continuous decrease or increases in RD as a function of estimated circuit proteome fraction are reported in Supplementary Tables 2 and 3.

Prediction of repressor off-target binding. RD values of all native genes in E. coli were tested for correlation against the RD of individual repressors in the circuit (RD was used as the estimate of protein expression level). We were specifically searching for sigmoidal relationships, resembling off-target repression by repressors. To do that, for each repressor, the eight induction states of the circuit were divided into two subgroups, in which the repressor was either on or off. Next, for all the genes across the genome, RD values within each subgroup were averaged, resulting in $\text{RD}_{\text{ON,}}$ and $\text{RD}_{\text{OFF,}}$ for each native gene $i$. We then calculated a fold-repression for each native gene $i$ using $\text{RD}_{\text{ON,}}$/ $\text{RD}_{\text{OFF}}$, and selected those native genes with $>5$-fold repression for visual inspection of sigmoidal relationship. Only hlyIIR repressor showed significant off-target activity, with aceB, tonB, and mfpA being the most repressed native genes. To identify potential binding site of hlyIIR repressor for its genomic off-targets, 100 nucleotides upstream or downstream (depending on the strand orientation) of the annotated TSS of each off-target gene was taken (upstream for aceB and tonB and downstream for mfpA). The operator sequence of hlyIIR repressor (5'-ATATTAAAAAATCCTGTTAAAA-3') was then aligned to each DNA sequence using Needleman–Wunsch aligning algorithm with mismatch penalty = 0, gap penalty = -2, and gap extension penalty = 0.

Promoter motif identification. Cryptic promoters were examined to determine whether they have motifs consistent with a $70\beta$ promoter. First, the TSSs were identified (above), for which we found 228 forward and reverse cryptic promoters across the circuit DNA sequence and across the induction states. These peaks associated with ribozyme cleavage were manually removed. Promoter strength was calculated for each cryptic promoter using Eq. 3, and those promoters with activity >10^{-5} RD per promoter were selected for motif identification (27 promoters passed the threshold, Supplementary Table 1), for which 50 nucleotides upstream of the TSS was taken for sequence analysis. Next, the −10 and −35 boxes of each promoter were manually identified. Not all promoters had both of the two boxes, and the spacing between them varied between 16 and 18 nucleotides. Finally, WebLogo (version 2.8.2) was used to find common motifs across these promoter sequences (Supplementary Fig. 8).

Calculation of off-target activity. The off-target activity, with $\text{OD}_{\text{off}}$, was calculated and correlated with change in the total proteome fraction on cellular metabolism, RNAP and ribosome expression, and cellular maintenance for proteins and mRNAs. The ribosome usage is defined as the number of ribosomes sequenced by a gene at steady state, and is calculated by summing the height of ribosome occupancy profile over the length of the gene. The ribosome density (RD) of each gene in the cell containing the circuit with that of the control cell. RD was used as the estimate of protein expression level. We calculated these differences for all eight induction states. To do that, differential gene expression analysis was performed using EdgeR, where its ‘exact test’ analysis was performed in R (version 3.4.4) and adjusted $p$-values were calculated using build-in false discovery rate (FDR) correction. EdgeR requires biological replicates for its exact test analysis in order to estimate the dispersion in data (dispersion = BCV) where BCV is the biological coefficient of variation). A fixed BCV = 0.1 was used in all exact test analysis. Genes >2-fold change in expression and with $p$ value and FDR less than 0.05 were selected as differentially expressed genes. These up and downregulated genes were used as the input for Gene Ontology analysis (GO) and identifies a list of regulated pathways based on their co-regulated gene members. Each regulated pathway was evaluated for potential relationship with estimated circuit proteome fraction by cross induction state correlation between RD of each gene member in the pathway and the estimated circuit proteome fraction across induction states. Not all pathways were correlated. Only those pathways/gens with continuous decrease or increases in RD as a function of estimated circuit proteome fraction are reported in Supplementary Tables 2 and 3.

Kinetic simulations of the genetic circuit. A model is developed that tracks the mRNA and protein production over time and includes parameters associated with each genetic part, for which quantitative values can be extracted from the –omics data (Tables 1 and 2). The kinetics of protein dimerization and protein binding to DNA are assumed to be much faster than transcription, translation, and degradation, and thus are treated as being at pseudo-steady-state. In addition, the model does not take either RNAP or ribosome movement into account and assumes that every initiation event results in the production of a mRNA or protein, respectively. Thus, transcriptional attenuation and ribosomal pausing and early truncation are not taken into account. Second, there is no coupling between devices due to ribosomal and RNAF usage and no impact from the circuit on the host cell, including changing the expression rate. Note, however, that since the parameters are empirically extracted from –omics data involving the entire circuit, these effects will be accounted for to some extent in their values. Equations for the number of copies of the mRNA $m_{i}$ of gene $i$ are written below. Consider first the equation for the PhiG gate, which is located at the 5’-end of the circuit DNA sequence,

$$\frac{dm_{\text{PhiG}}}{dt} = \text{I}_{\text{PhiG}} - \gamma_{\text{ySrpR}}(1-b) + b) m_{\text{PhiG}}$$

The RNAP flux from the two input promoters are summed as

$$\text{I}_{\text{RNAP}} = \text{y}_{\text{RNAP}} + \text{y}_{\text{Phi}}$$

The degradation term includes the ribosome cleavage efficiency $\eta$[0.1] of the RiboJ(i) insulator. Uncleaved mRNA degrades more quickly, captured in the parameter $b = 2$, the value of which is obtained from the literature. The mRNA degradation rate $\gamma = 0.0067 s^{-1}$ is also taken from the literature and assumed to be equal for all the circuit mRNA. The next gate on the circuit DNA has a similar form, except that it also has the possibility of RNAP read-through from the upstream gate (PhiG) due to an imperfect terminator. This can be written as

$$\frac{dm_{\text{Phi}}}{dt} = \text{I}_{\text{Phi}} - \gamma_{\text{yTet}}(1-b) + b) m_{\text{Phi}}$$

$$\text{I}_{\text{RNAP}} = \text{y}_{\text{RNAP}} + \text{y}_{\text{Phi}} + \text{I}_{\text{Phi}}$$

where the last term captures the read-through due to the terminator (Eq. 7). The terminator strength is $T_{\text{Phi}}$ (suffix is the last two digits of the terminator gene).
The equations are numerically solved using Python and the forward Euler method ($dt = 10$ s) and typically solved to 12 hours. For each simulation trajectory, the equations are initiated by solving them until they reach steady-state under the initial states of the sensor inputs, noting that two sensors control two identical promoters in the circuit (\(y_{\text{Tet}}(y_{\text{Tet}1},y_{\text{Tet}2}), y_{\text{BAD}1}(y_{\text{BAD}1}), y_{\text{BAD}2}(y_{\text{BAD}2})\)). At time $t = 0$, a step change is applied to change the sensor promoters to their new activities. The on/off values used for the promoters are: $0.045/0.0000018$ RNAps (\(y_{\text{Tet}}\)), $1.537/0.0000266$ RNAps (\(y_{\text{Tet}1}\)), $0.982/0.0000018$ RNAps (\(y_{\text{Tet}2}\)), $0.043/0.0015$ RNAps (\(y_{\text{BAD}1}\)), $0.054/0.000037$ RNAps (\(y_{\text{BAD}2}\)), also measured from RNA-seq data (Table 1). For the sensitivity analysis, one input variable was moved keeping all others at their nominal values. For each of 8 induction states, a full simulation trajectory (including reinitialization) was then run to 12 hours with the new input variable. The YFP protein level at steady-state (end of the simulation trajectory) was obtained for all eight induction states. The circuit by design has two on states (+/+−, +/+/+), with IPTG/iTF/ara induction, and six off states (the remaining combinations of inducers). The circuit score is calculated by taking the minimum $R_{\text{YFP}}$ of the two states that should be on, and dividing it by the maximum $R_{\text{YFP}}$ of the six states that should be off. This process was repeated with 1000 different values for each input variable, the variable was then returned to its nominal value before the sensitivity analysis was conducted for the next variable. Changes in circuit score as a function of each input variable are plotted in Fig. 5c; a circuit score less than 1 renders the circuit nonfunctional. A higher margin of circuit score $= 3$ was used as a cutoff for the plots. Parameters for the promoters are plotted for $y/N$, where $N = 9$ is the plasmid copy number for p15A to present in units of RNAps per DNA. For $\text{P}_{\text{BAD}1}R_{\text{BAD}1}$, $N = 4$ is used due to the output promoter being on a different plasmid than the circuit with a higher copy number$^{35}$.

**Statistics and reproducibility**

For each statistical comparison, $R^2$ and $p$ value were calculated using corcoef function in MATLAB (R2020a). When $p$ value <0.05, the corresponding $R^2$ is considered significant. Cell doubling times and fluorescence measurements were each repeated three times with similar results. For each induction state of the circuit, RNA-seq and ribosome profiling were performed once.

**Reporting summary**

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

**Data availability**

Source data are provided with this paper. FPKM, RD, and proteome proportions of all circuit and genomic proteins are listed in Supplementary Data 1. RNA-seq and ribosome profiling data collected in this study (both raw and processed) as well as the E. coli MG1655 genome, and DNA sequence of E. coli DH10B genome (NC_010473.1) were obtained from NCBI (https://www.ncbi.nlm.nih.gov/nuccore/NC_010473). Any other relevant data are available from the authors upon reasonable request.

**Code availability**

Python scripts that implement the complete characterization of genetic circuits, sensitivity analysis, and dynamic model of genetic circuits are released as open-source software under the MIT license (GitHub repository: https://github.com/VoigtLab/Comprehensive_Genetic_Circuit_Analysis).

Received: 22 April 2020; Accepted: 2 September 2020; Published online: 05 October 2020

**References**

1. Brophy, J. A. & Voigt, C. A. Principles of genetic circuit design. Nat. Methods **11**, 508–520 (2014).

2. Guet, C. C., Elowitz, M. B., Hsing, W. & Leibler, S. Combinatorial synthesis of genetic networks. Science **296**, 1466–1470 (2002).

\[
\frac{dm_{\text{mamR}}}{dt} = I_{\text{mamR}} - \gamma (\eta_{\text{mamR}}(1 - b) + b) m_{\text{mamR}} \\
I_{\text{mamR}} = y_{\text{mamR}} + \frac{I_{\text{mamR}}}{T_{\text{mamR}}} 
\]

\[
\frac{dm_{\text{betI}}}{dt} = I_{\text{betI}} - \gamma (\eta_{\text{betI}}(1 - b) + b) m_{\text{betI}} \\
I_{\text{betI}} = y_{\text{betI}} + \frac{I_{\text{betI}}}{T_{\text{betI}}} 
\]

\[
\frac{dm_{\text{mamR}}}{dt} = I_{\text{mamR}} - \gamma (\eta_{\text{mamR}}(1 - b) + b) m_{\text{mamR}} \\
I_{\text{mamR}} = y_{\text{mamR}} + \frac{I_{\text{mamR}}}{T_{\text{mamR}}} 
\]

\[
\frac{dm_{\text{betI}}}{dt} = I_{\text{betI}} - \gamma (\eta_{\text{betI}}(1 - b) + b) m_{\text{betI}} \\
I_{\text{betI}} = y_{\text{betI}} + \frac{I_{\text{betI}}}{T_{\text{betI}}} 
\]

where the flux terms $I_j$ capture all the transcription coming from upstream of the gate because they are all oriented in the same direction. Finally, the transcription of $y_{\text{fcp}}$ from the circuit output promoter is written as

\[
\frac{dm_{\text{yfcp}}}{dt} = y_{\text{fcp}} - \gamma (\eta_{\text{fcp}}(1 - b) + b) m_{\text{yfcp}} 
\]

The induction of the sensor and gate output promoters is assumed to be fast with respect to the repressor production or degradation\(^{140}\). The pseudo-steady-state approximation allows the response function of the gates to be used to calculate the activity of the output promoter as a function of the number of repressor proteins in the cell,

\[
y_i = y_{\text{max}i} + \left( y_{\text{max}i} - y_{\text{max}i} \right) \frac{k_i}{K_i + R_i} 
\]

where $y_{\text{max}i}$, $y_{\text{max}i}$, and $k_i$ for gate $i$ are measured from the RNA-seq data (Tables 1 and 2, and Supplementary Fig. 19). The term $K_i$ is the binding constant of repressor $i$ to the output promoter of gate $i$ and is in units of protein number. This is related to the term $K_i$ used in the response function of gate $i$ (Supplementary Fig. 19; in units of RNAps), through

\[
k_i = \left( \frac{a_i}{\gamma(\eta_i(1 - b) + b)} \right) K_i 
\]

where $a_i$ is the translation efficiency of the RBS controlling the translation of the repressor mRNA, $\mu = 0.00026 s^{-1}$ is assumed to be dominated by cell division (the magnitude of which corresponds to a 45 min doubling time). Fits for the calculation of the $k_i$ for each gate are shown in Supplementary Fig. 19. Note that the parameters in Eq. 22 are not varied as part of the sensitivity analysis; rather, the simulation of the mutation of the output promoter is done by varying $k_i$. Finally, the number $R_i$ of protein $i$ in the cell can be calculated using simple mass action kinetics and the mRNA number:

\[
\frac{dR_i}{dt} = a_i m_{\text{yfcp}} - \mu R_i \\
\frac{dR_{\text{fcp}}}{dt} = a_\text{fcp} m_{\text{fcp}} \mu R_{\text{fcp}} \\
\frac{dR_{\text{mamR}}}{dt} = a_\text{mamR} m_{\text{mamR}} - \mu R_{\text{mamR}} \\
\frac{dR_{\text{betI}}}{dt} = a_\text{betI} m_{\text{betI}} - \mu R_{\text{betI}} \\
\frac{dR_{\text{yfcp}}}{dt} = a_\text{yfcp} m_{\text{yfcp}} - \mu R_{\text{yfcp}} 
\]
3. Burgard, A. P., Pharkya, P. & Maranas, C. D. Optknock: a bilevel programming framework for identifying gene knockout strategies for metabolic strain optimization. Biotechnol. Bioeng. 84, 647–657 (2003).

4. Cho, B. K. et al. The transcription unit architecture of the Escherichia coli genome. Nat. Biotechnol. 27, 1043–1049 (2009).

5. Heirendt, L. et al. Creation and analysis of biochemical constraint-based models using the COBRA Toolbox v.3.0. Nat. Protoc. 14, 639–702 (2019).

6. Li, Y. et al. Modular construction of mammalian gene circuits using TALE-SRDX repressor protein. J. Biol. Eng. 3, 387–393 (2009).

7. Gorochowski, T. E. et al. Absolute quantification and burden using combined sequencing approaches. Nat. Biotechnol. 32, 621–628 (2008).

8. Leonard, S. et al. APERO: a genome-wide approach for identifying bacterial small RNAs from RNA-Seq data. Nucleic Acids Res. 47, e88 (2019).

9. Luo, R. et al. SOA2enovo2: an empirically improved memory-efficient short-read de novo assembler. Gigascience 1, 18 (2012).

10. McClure, R. et al. Computational analysis of bacterial RNA-Seq data. Nucleic Acids Res. 41, e140 (2013).

11. Schulz, M. H., Zerbin, D. R., Vingron, M. & Birney, E. Oases: robust de novo RNA-Seq assembly across a dynamic range of expression levels. Bioinformatics 28, 1086–1092 (2012).

12. Lalanne, J. B. et al. Evolutionary convergence of pathway-specific enzyme expression stoichiometry. Cell 173, 749–761 e738 (2018).

13. Ingolia, N. T. Ribosome profiling: new views of translation, from single codons to genome scale. Nat. Rev. Genet. 15, 205–213 (2014).

14. Ingolia, N. T., Ghaemmaghami, S., Newman, J. R. & Weissman, J. S. Genome-wide analysis in vivo of translation with nucleotide resolution using ribosome profiling. Science 324, 218–223 (2009).

15. Luse, J. et al. Ribosomal profiling data. PLoS Comput. Biol. 14, e1006619 (2018).
65. Thattai, M. & van Oudenaarden, A. Intrinsic noise in gene regulatory networks. *Proc. Natl Acad. Sci. USA* 98, 8614–8619 (2001).

66. Aspen, J. L. et al. Extensive translation of small open reading frames revealed by Poly-Ribo-Seq. *elife* 3, e03528 (2014).

67. Brophy, J. A. & Voigt, C. A. Antisense transcription as a tool to tune gene expression. *Mol. Syst. Biol.* 12, 854 (2016).

68. Schleif, R. Regulation of the L-arabinose operon of Escherichia coli. *Trends Genet.* 16, 459–465 (2000).

69. Lou, C., Stanton, B., Chen, Y. J., Munsky, B. & Voigt, C. A. Ribozyme-based insulator parts buffer synthetic circuits from genetic context. *Nat. Biotechnol.* 30, 1137–1142 (2012).

70. Naville, M. & Gautheret, D. Transcription attenuation in bacteria: theme and variations. *Brief. Funct. Genom.* 9, 178–189 (2010).

71. Peters, J. M., Vangeloff, A. D. & Landick, R. Bacterial transcription terminators: the RNA 3′-end chronicles. *J. Mol. Biol.* 412, 793–813 (2011).

72. Suzuki, H., Kunisawa, T. & Otsuka, J. Theoretical evaluation of transcriptional pausing effect on the trp leader sequence. *Biophys. J.* 49, 425–435 (1986).

73. Turner, R. J., Lu, Y. & Switzer, R. L. Regulation of the Bacillus subtilis pyrimidine biosynthetic (pyr) gene cluster by an autogenous transcriptional attenuation mechanism. *J. Bacteriol.* 176, 3708–3722 (1994).

74. Yanofsky, C. Attenuation in the control of expression of bacterial operons. *Nature* 289, 751–758 (1981).

75. Yanofsky, C. Transcription attenuation: once viewed as a novel regulatory mechanism. *J. Bacteriol.* 182, 1–8 (2000).

76. Kohler, R., Mooney, R. A., Mills, D. J., Landick, R. & Cramer, P. Architecture of a transcribing-translating expression. *Science* 356, 194–197 (2017).

77. McGary, K. & Nudler, E. RNA polymerase and the ribosome: the close relationship. *Curr. Opin. Microbiol.* 16, 112–117 (2013).

78. Proshkin, S., Rahmouni, A. R., Mironov, A. & Nudler, E. Cooperation between translating ribosomes and RNA polymerase in transcription elongation. *Science* 328, 504–508 (2010).

79. Darnell, A. M., Subramaniam, A. R. & O’Shea, E. K. Translational control through differential ribosome pausing during amino acid limitation in mammalian cells. *Mol. Cell* 71, 229–243 e211 (2018).

80. Mohammad, F., Woolstenthuule, C. J., Green, R. & Buskirk, A. R. Clarifying coupling of RNA polymerase and ribosome translation elongation speed based on LacZalpha complementation. *Mol. Syst. Biol.* 6, 356 (2010).

81. Bratton, C. A., Arkin, P. A. & Dubendorff, J. W. Effect of tRNAs on ribosome translation elongation. *Nature* 435, 565 (2005).

82. Collins, J. A., Irom, I., Baker, S. & Winkler, W. C. Mechanism of histone mRNP destabilization by the glmS ribozyme. *Genes Dev.* 21, 3356–3368 (2007).

83. Brar, G. A. & Weissman, J. S. Ribosome profiling reveals the what, when, where and how of protein synthesis. *Nat. Rev. Mol. Cell Biol.* 16, 651–664 (2015).

84. Kossi, S. et al. Composability of regulatory sequences controlling transcription and translation in Escherichia coli. *Proc. Natl Acad. Sci. USA* 110, 14024–14029 (2013).

85. Gutell, R. et al. Precise and reliable gene expression via standard transcription and translation initiation elements. *Nat. Methods* 10, 354–360 (2013).

86. Goodman, D. B., Church, G. M. & Kossi, S. Causes and effects of N-terminal codon bias in bacterial genes. *Science* 342, 475–479 (2013).

87. Levin-Karp, A. et al. Quantifying translational coupling in E. coli synthetic operons using RBS modulation and fluorescent reporters. *ACS Synth. Biol.* 2, 327–336 (2013).

88. Tian, T. & Salis, H. M. A predictive biophysical model of translational coupling to coordinate and control protein expression in bacterial operons. *Nucleic Acids Res.* 43, 7137–7151 (2015).

89. Esparzhorojian, A., Channarapapa, A. S. & Salis, H. M. Translation rate is controlled by coupled trade-offs between site accessibility, selective RNA unfoiling and sliding at upstream standby sites. *Nucleic Acids Res.* 42, 2646–2659 (2014).

90. Salis, H. M., Mirsky, E. A. & Voigt, C. A. Automated design of synthetic ribosome binding sites to control protein expression. *Nat. Biotechnol.* 27, 946–950 (2009).

91. Qian, Y., Huang, H. B., Jimenez, I. J. & Del Vecchio, D. Resource competition shapes the response of genetic circuits. *ACS Synth. Biol.* 6, 1263–1272 (2017).

92. Liao, C., Blanchard, A. E. & Lu, T. An integrative circuit-host modelling framework for predicting synthetic gene network behaviours. *Nat. Microbiol.* 2, 1658–1666 (2017).

93. Fontanarrosa, P. F. et al. Genetic circuit dynamics: Hazard and Glitch analysis. *ACS Synth. Biol.* https://doi.org/10.1021/acssynbio.0c00055 (2020).
126. Liu, Y. et al. Bacterial single cell whole transcriptome amplification in microfluidic platform shows putative gene expression heterogeneity. *Anal. Chem.* **91**, 8036–8044 (2019).

127. Ma, Q., Bucking, H., Gonzalez Hernandez, J. L. & Subramanian, S. Single-cell RNA sequencing of plant-associated bacterial communities. *Front. Microbiol.* **10**, 2452 (2019).

128. Kim, B., Du, J., Eriksen, D. T. & Zhao, H. Combinatorial design of a highly efficient xylose-utilizing pathway in *Saccharomyces cerevisiae* for the production of cellulolic biofuels. *Appl. Environ. Microbiol.* **79**, 931–941 (2013).

129. Blomberg, P., Wagner, E. G. & Nordstrom, K. Control of replication of *E. coli* genome. *J. Mol. Biol.* **193**, 1–12 (1987).

130. Martin, M. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBBet.* **J.** **17**, 10–12 (2011).

131. Langmead, B., Trapnell, C., Pop, M. & Salzberg, S. L. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol.* **10**, R25 (2009).

132. Lorenz, R. et al. ViennaRNA package 2.0. *Algorithms Mol. Biol.* **6**, 26 (2011).

133. Keseler, I. M. et al. The EcoCyc database: reflecting new knowledge about *Escherichia coli* K-12. *Nucleic Acids Res.* **45**, D543–D550 (2017).

134. Needleman, S. B. & Wunsch, C. D. A general method applicable to the search for similarities in the amino acid sequence of two proteins. *J. Mol. Biol.* **48**, 443–453 (1970).

135. Clifton, K. P. et al. The genetic insulator Ribof increases expression of insulated genes. *J. Biol. Eng.* **12**, 23 (2018).

136. Ashburner, M. et al. Gene ontology: tool for the uniﬁcation of biology. The Gene Ontology Consortium. *Nat. Genet.* **25**, 29–50 (2000).

137. Needleman, S. B. & Wunsch, C. D. A general method applicable to the search for similarities in the amino acid sequence of two proteins. *J. Mol. Biol.* **48**, 443–453 (1970).

138. Crooks, G. E., Hon, G., Chandonia, J. M. & Brenner, S. E. WebLogo: a sequence logo generator. *Genome Res.* **14**, 1188–1190 (2004).

139. Rydenfelt, M., Cox, R. S. 3rd, Garcia, H. & Phillips, R. Statistical mechanical model of coupled transcription from multiple promoters due to transcription factor titration. *Phys. Rev. E Stat. Nonlin. Soft Matter Phys.* **89**, 012702 (2014).

**Acknowledgements**

This work was supported by US Defense Advanced Research Projects Agency (DARPA) 1KM grant HR0011-15-C-0084 (C.A.V., A.E.B., J.Z., and H.D.); US Department of Commerce—NIST (National Institute of Standards and Technology) award 70NANB16H164 (C.A.V., A.A.K.N., and J.Z.); DARPA SD2 grant (C.A.V., A.E.B., and H.D.); and US Department of Energy—Boulder award DE-SC0018368 (C.A.V. and H.D.). We would like to thank Professor Gene-Wei Li and Ariel Schaefer at MIT Biology Department for their training on performing end-enriching RNA-seq and ribosome profiling experiments. We would also like to thank Alexander Cristofaro and Cassandra Bristol at the MIT-Broad Foundry for assisting in EdgeR and Gene Ontology analysis, and for assisting in the sequencing runs.

**Author contributions**

J.Z., A.A.K.N., and C.A.V. conceived the study and designed the experiments. J.Z., A.E.B., and A.A.K.N. performed the experiments. A.E.B., J.Z., and H.D. performed the data analysis. A.E.B., J.Z., H.D., and C.A.V. wrote the manuscript.

**Competing interests**

The authors declare no competing interests.

**Additional information**

Supplementary information is available for this paper at https://doi.org/10.1038/s41467-020-18630-2.

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

**Peer review information** *Nature Communications* thanks the anonymous reviewer(s) for their contribution to the peer review of this work. Peer reviewer reports are available.

**Reprints and permission information** is available at http://www.nature.com/reprints

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

© The Author(s) 2020