Burden-driven feedback control of gene expression

Francesca Ceroni[1,2], Alice Boo[2,3], Simone Furini[4], Thomas E Gorochowski[5], Olivier Borkowski[2,3], Yaseen N Ladak[6], Ali R Awan[2,3], Charlie Gilbert[2,3], Guy-Bart Stan[2,3] & Tom Ellis[2,3]

© 2018 Nature America, Inc., part of Springer Nature. All rights reserved.

Cells use feedback regulation to ensure robust growth despite fluctuating demands for resources and differing environmental conditions. However, the expression of foreign proteins from engineered constructs is an unnatural burden that cells are not adapted for. Here we combined RNA-seq with an *in vivo* assay to identify the major transcriptional changes that occur in *Escherichia coli* when inducible synthetic constructs are expressed. We observed that native promoters related to the heat-shock response activated expression rapidly in response to synthetic expression, regardless of the construct. Using these promoters, we built a dCas9-based feedback-regulation system that automatically adjusts the expression of a synthetic construct in response to burden. Cells equipped with this general-use controller maintained their capacity for native gene expression to ensure robust growth and thus outperformed unregulated cells in terms of protein yield in batch production. This engineered feedback is to our knowledge the first example of a universal, burden-based biomolecular control system and is modular, tunable and portable.

The maintenance and expression of synthetic constructs add an unnatural load to host cells, typically known as burden[1]. Although some burden stems from the specific roles of the proteins encoded by the constructs (for example, enzymes that consume metabolites), most burden placed on cells arises simply from the consumption of finite cellular resources during the expression of construct genes[2–5]. The cellular response to the load of extra expression is typically decreased growth and global physiological changes. These changes are somewhat unpredictable and usually reduce the expected performance of engineered cells[6–9].

Efforts to increase control over and predictability of engineered biological systems have led to tools for measuring and reducing burden, developed through improved understanding of host–construct interactions[10,11]. This has led to the development of orthogonal expression systems[12], a transcriptional resource allocator[13] and libraries of promoters able to tune construct expression to decrease cell stress[14]. These are all strategies that bypass some of the resource sharing required for gene expression but do not fully eliminate it.

In parallel, mathematical modeling has enabled researchers to understand the bottlenecks in expression burden and consider methods to alleviate them[4,15,16]. Previously, we developed a capacity monitor system for *E. coli* that quantifies the extent to which synthetic constructs diminish a cell's capacity for gene expression. We used this system to identify more efficient construct-design strategies that showed robust growth with less chance of mutations arising in populations[17]. In that work, we noted that a period of substantial competition for expression resources causes decreased growth of host cells over the next few hours, even in rich growth media. As sudden overexpression of extra genes is likely to be encountered naturally in bacteria (for example, after phage infection), we reasoned that cells must have native mechanisms to adapt and reallocate resources to continue growth, beyond the well-known stringent response associated with nutrient starvation[18]. Investigating how *E. coli* adapts in the face of synthetic construct expression offers the chance to identify native systems that sense and respond to burden that could be exploited downstream in synthetic gene circuits.

To identify a host response to burden, we here made use of RNA-seq's ability to quantify transcription rates[19] and identify significant gene expression changes between conditions[20]. We paired our previously described *in vivo* capacity monitor assay with a multiplex RNA-seq strategy to explore transcriptional changes in *E. coli* that occur after the onset of expression burden, and used the findings to build a tunable CRISPR–dCas9-based feedback system that can regulate construct expression in response to burden, thereby maintaining cellular capacity during synthetic expression.

RESULTS

Characterizing the burden of different synthetic constructs

We investigated the *in vivo* burden and transcriptome response of *E. coli* strains expressing a chosen set of four synthetic constructs with different genes, codon optimization and regulation (**Fig. 1a**). We chose two strains, DH10B and MG1655, as hosts; both incorporated the previously described capacity monitor, in which a small expression cassette stably integrated into the chromosome produces GFP at a constitutive rate[17]. The induction of

[1]Department of Chemical Engineering, Imperial College London, London, UK. [2]Imperial College Centre for Synthetic Biology, Imperial College London, London, UK. [3]Department of Bioengineering, Imperial College London, London, UK. [4]Department of Medical Biotechnologies, University of Siena, Siena, Italy. [5]BrisSynBio, University of Bristol, Bristol, UK. [6]ITMAT Data Science Group, Department of Surgery and Cancer, Faculty of Medicine, Imperial College London, London, UK. Correspondence should be addressed to T.E. (t.ellis@imperial.ac.uk) or G.-B.S. (g.stan@imperial.ac.uk).

RECEIVED 28 AUGUST 2017; ACCEPTED 1 FEBRUARY 2018; PUBLISHED ONLINE 26 MARCH 2018; DOI:10.1038/NMETH.4635

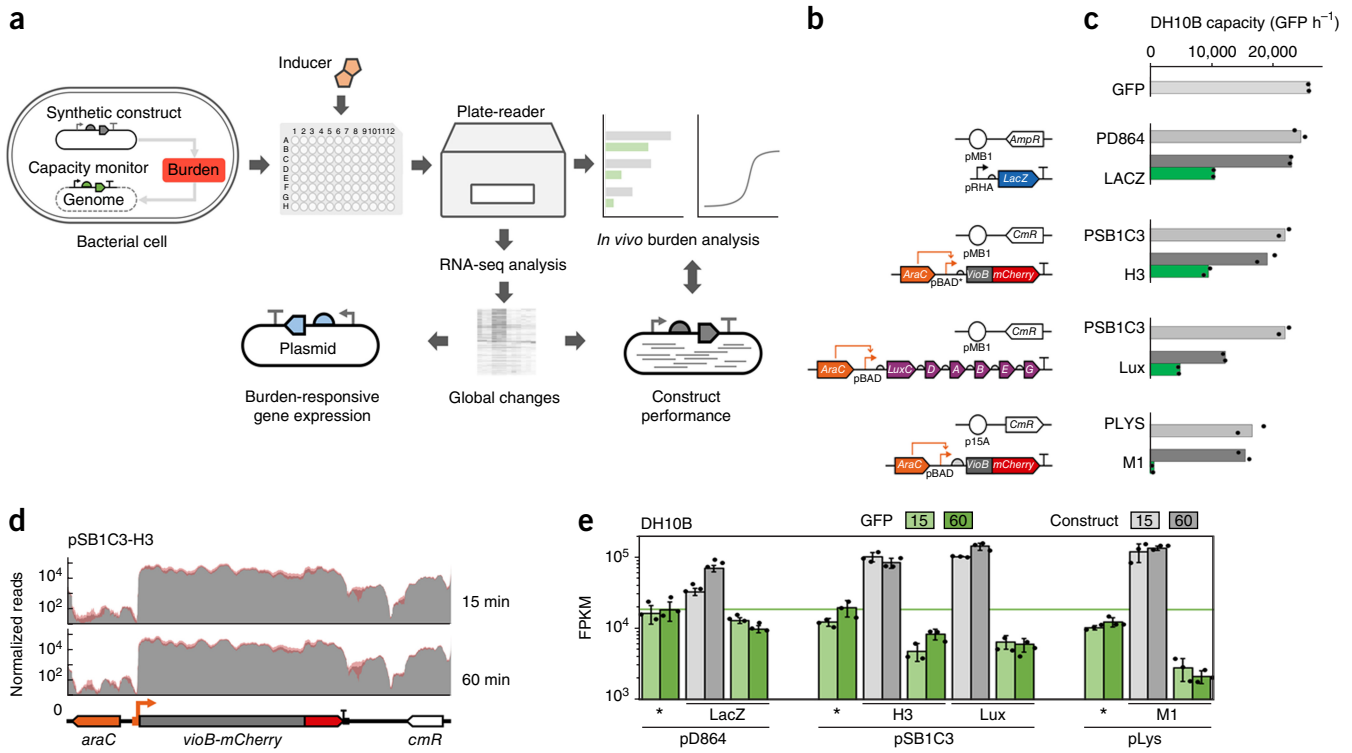

Figure 1 | Characterization of host–construct interactions in *E. coli*. (**a**) The workflow for the present study. Cells were transformed with different synthetic constructs and their empty control plasmids, and *in vivo* burden response and global transcriptional changes were measured with RNA-seq. Burden-responsive promoters were identified and used in the design of burden-based biomolecular feedback systems. (**b**) Diagrams of the four constructs used in this study and their control plasmids. pBAD and pRHA are inducible by arabinose and rhamnose, respectively. (**c**) *In vivo* cell capacity 1 h after induction in DH10B cells, as determined on the basis of the GFP production rate per cell. Green bars represent induced samples. Gray bars represent uninduced samples; light gray indicates the untransformed strain (top) and strains with empty plasmids, and dark gray indicates strains with construct plasmids. Values are shown as means of independent repeats ($n = 2$ per induction state, represented by black dots). (**d**) Transcription profiles of the synthetic construct pSB1C3-H3 in DH10B cells 15 (top) and 60 (bottom) min after induction. Traces show the normalized reads mapping to the plasmid constructs. Solid gray regions represent average values, and transparent red regions span from the minimum to the maximum of the transcription profiles generated from three biological replicates. Genetic designs below the profiles show the precise location of each genetic part. (**e**) Transcriptional induction of the four constructs, represented by mean FPKM values in DH10B cells. Green bars show data for GFP, and gray bars show data for construct genes; color intensity indicates the time point as defined in the keys. Asterisks denote empty plasmid controls. Data are shown as the mean ± s.d. of three independent repeats (specific values are indicated by black dots).

heterologous gene expression burdens the cell by depleting host resources, and this can be measured on the basis of the subsequent decrease in GFP production rate.

For the synthetic constructs, we investigated three exemplary cases of plasmid-based inducible expression: (i) inducible reporter expression (pD864-LacZ), (ii) overexpression of a large heterologous protein (pSB1C3-H3 and pLys-M1) and (iii) expression of an operon encoding a metabolic pathway (pSB1C3-Lux). **Supplementary Note 1** presents a description of these constructs and justification for their selection.

We characterized the four constructs and plasmid-only controls (**Fig. 1b**, **Supplementary Table 1**) in the two host strains, taking capacity measurements via *in vivo* plate-based burden assay[17]. By comparing GFP production rates per cell 1 h after induction (or without induction for controls), we were able to infer the burden imposed by the constructs (**Fig. 1c**). We observed a substantial decrease in GFP production rate for all samples with induced synthetic constructs, which confirmed that they all caused burden due to the use of host expression resources. The assay also determined the cell growth rates during the experiment, which in this case were decreased after 1 h of construct expression (**Supplementary Fig. 1**).

Having determined *in vivo* burden, we next used multiplex RNA-seq analysis to characterize the transcriptional performance of the constructs and their host cells. We grew cells for RNA harvesting as for the *in vivo* assay, but collected RNA 15 and 60 min after induction to investigate rapid changes in the transcriptome. The total amount of RNA harvested per sample closely correlated with the concentration of cells measured in the *in vivo* assay at equivalent time points, which supported the idea that RNA-seq and *in vivo* assay data sets can be directly compared (**Supplementary Fig. 2**). To avoid confusing the host response to burden with the response to the inducers (arabinose and rhamnose), we used strains with empty plasmids also exposed to the inducers as control strains.

We multiplexed 90 samples for RNA-seq (**Supplementary Table 2**) after individual RNA purification, rRNA removal and further sample-preparation steps (Online Methods). This allowed for three biological replicates for constructs and controls, at two time points and in the two strains. After sequencing, raw reads

© 2018 Nature America, Inc., part of Springer Nature. All rights reserved.

were aligned to a reference sequence for each strain, plasmid and GFP cassette, normalized and then used to determine the fragments per kilobase of transcript per million mapped reads (FPKM) for all genes (Online Methods).

Using RNA-seq, we visualized the transcriptional profiles of plasmid constructs at 15 and 60 min after induction (**Fig. 1d**, **Supplementary Fig. 3a**). Mapped reads that aligned to plasmids showed the transcribed regions and the magnitude of their transcription, revealing activated expression from induced promoters and selection markers and some pervasive transcription from backbone regions. Total plasmid-based transcription in construct samples accounted for a substantial percentage of all non-rRNA transcription in *E. coli* (**Supplementary Tables 3** and **4**). Indeed, only 15 min after induction in DH10B cells, between 11.7% (pD864-LacZ) and 47.3% (pLys-M1) of all mapped reads aligned to synthetic constructs. Notably, total transcription from the empty pLys plasmid was unexpectedly higher than that from the other empty plasmids, especially in DH10B cells.

We used FPKM values to assess both transcription of the GFP capacity monitor and the output expression of the synthetic constructs. At both time points, high levels of transcription from constructs corresponded to decreased chromosomal GFP transcription compared with that in controls (**Fig. 1e**, **Supplementary Fig. 3b**). This demonstrates that induced transcription of synthetic constructs rapidly affected genomic transcription with all constructs tested and that burden clearly affected the host transcriptome. Indeed, when we directly compared the decrease in GFP RNA-seq reads per sample with the decrease in GFP production rate per cell (as measured *in vivo*), we found very similar profiles (**Supplementary Fig. 3c**). This correlation between GFP production rate per cell and GFP FPKM held for most of the samples tested, including controls (**Supplementary Fig. 2**).

Notably, the decrease in GFP expression was least pronounced for pD864-LacZ, which expresses its output transcript (*lacZ* mRNA) less than other constructs, as the activator of the pRhaBAD promoter, RhaS, is provided only *in trans* by the host genomic copy. As in the *in vivo* assay, pLys-M1 had the greatest effect on host GFP transcription, and this effect was especially pronounced in DH10B cells.

RNA-seq reveals a common host response to burden

The *in vivo* and RNA-seq data together revealed variation in the degree of burden imparted by different constructs and differences in strain behavior. Next, we investigated the effect of construct expression on the host transcriptome, and specifically examined whether cells had a common response to burden regardless of the genetic content of synthetic constructs or the host strain. For every genomic gene except rRNA genes, we calculated the fold change in expression between empty plasmid sample and construct, and from this we identified genes with significant differential expression. We determined the number of genes with significant up- or downregulation after 60 min for the three main constructs, pSB1C3-H3, pSB1C3-Lux and pD864-LacZ, in both host strains (**Fig. 2a**, **Supplementary Fig. 4a**), and carried out the same analysis for samples at 15 min after induction (**Supplementary Fig. 5**). We also conducted equivalent analyses that included pLys-M1, a more burdensome version of pSB1C3-H3 (**Supplementary Fig. 6**).

The total number of genes with differential expression 60 min after induction was higher in DH10B cells than in MG1655 cells. While many genes were up- or downregulated with only one construct, a considerable number changed expression for multiple constructs, with 61 genes in DH10B and 6 genes in MG1655 differentially expressed in all three cases. We examined the ten genes showing the greatest significant difference with each construct to determine their regulation by identifying their σ-factor family, and made an equivalent list for the top ten differentially expressed genes shared among all three constructs for both strains (**Fig. 2b**, **Supplementary Fig. 4b**). All ten common genes showed upregulation due to burden, and nearly all were expressed from promoters using the heat-shock response σ-factor (σ^{32}). Three promoters in particular showed universal and striking upregulation of transcription: the *ibpAB* and *dnaKJ* operon promoters, and the *htpG* promoter (**Fig. 2c**, **Supplementary Fig. 4c**).

We inspected transcription of these three genes at the genomic level, along with the capacity monitor GFP cassette and *groSL* operon (**Fig. 2d**, **Supplementary Fig. 4d**). *groSL* encodes heat-shock genes that control protein folding and the amount and activity of σ^{32} in cells[21,22]. Transcriptional profiles showed that transcription of these σ^{32}-regulated genes increased significantly 60 min after induction of constructs, and GFP transcription stayed relatively high. Using the RNA-seq data, we then determined the fold difference in transcription from these promoters in response to the construct-induced burden (**Fig. 2e**). We observed increases in promoter activity of up to 50-fold in DH10B cells and 10-fold in MG1655 cells. These therefore act as intrinsic biosensors for synthetic-construct-induced burden in *E. coli*, although the precise mechanisms behind their activation are not fully understood.

To demonstrate the use of the promoters in this regard, we constructed GFP expression plasmids with these promoters (**Supplementary Table 5**) and assessed their ability to produce fluorescence in DH10B cells in response to induction of a burden-causing VioB–mCherry expression construct (**Supplementary Fig. 7**). For the *htpG* promoter, this required separating two overlapping natural promoters (*htpG1* and *htpG2*) and testing them as individual promoters. The complex *dnaK* promoter could not be separated into functional promoters and thus was not explored further (**Supplementary Note 2**).

All four promoters showed basal GFP expression with notable activation after induction of burden. For the *htpG1*, *htpG2* and *groSL* promoters, relative basal and activated expression levels closely matched equivalent RNA-seq data, which confirmed that their response to burden was maintained when they were taken out of their genomic context. This was not the case for the *ibpAB* promoter, which showed the strongest expression but with high basal levels.

A burden-responsive biomolecular feedback controller

The *htpG1* promoter had the best on/off characteristics, and RNA-seq data showed that this promoter was activated by burden in both strains for all constructs despite substantial differences in genetic content, inducers and plasmid backbones. This promoter thus has promise as a biosensor for burden, and also offers a key component for creating a negative feedback system to control gene expression burden in *E. coli*: burden is sensed via the *htpG1* promoter, and in response an effector represses the expression of burdensome genes. A feedback system with such a design should

© 2018 Nature America, Inc., part of Springer Nature. All rights reserved.

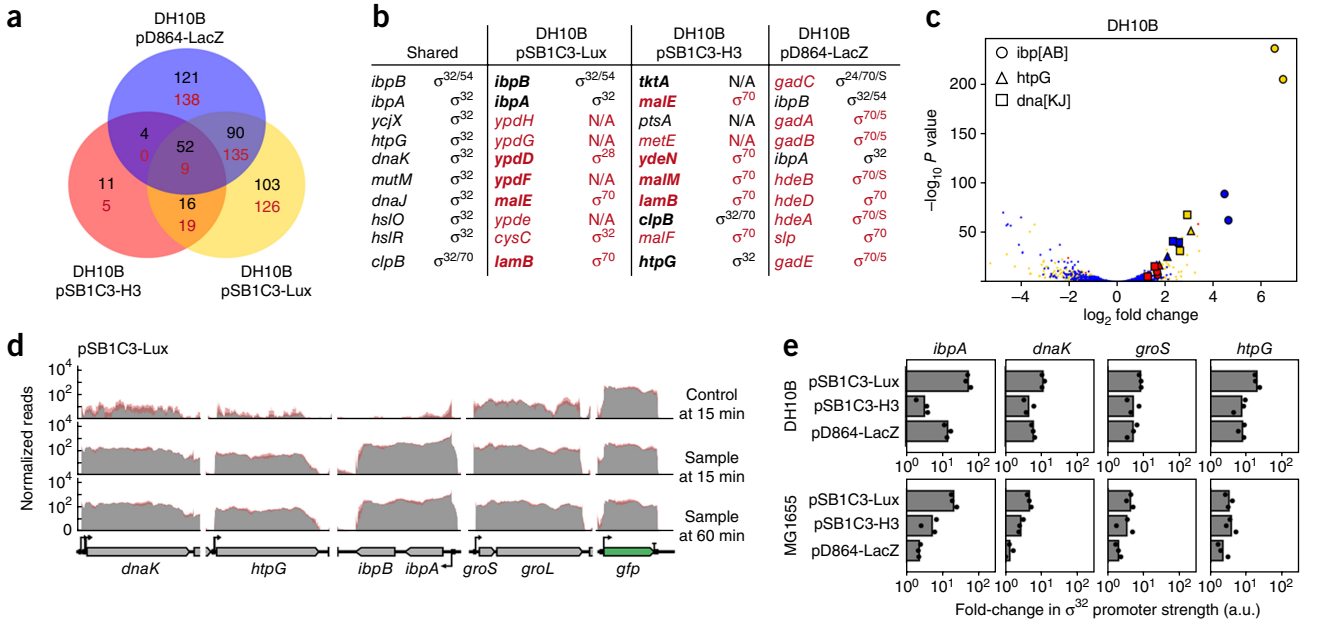

Figure 2 | Global transcriptional changes and activation of genomic σ^{32}-regulated promoters in response to synthetic construct expression. (**a**) A Venn diagram of differentially expressed genes. The set of significantly differentially expressed native genes ($\alpha < 0.05$) was identified in DH10B cells separately for each of three synthetic gene circuits, comparing cells transformed with the synthetic circuits to cells transformed with the corresponding empty plasmids. Upregulated and downregulated genes for each sample in DH10B cells are displayed in black and red, respectively. (**b**) The top ten differentially expressed genes (black, upregulated; red, downregulated) for each construct and the associated regulatory σ-factors. Bold gene symbols indicate genes that were also differentially expressed at 15 min post-induction. The leftmost column shows the differentially expressed genes shared among all three constructs. (**c**) A volcano plot showing genes with statistically significant changes in DH10B cells. Color-coding corresponds to that in **a**. (**d**) Transcription profiles of the *dnaKJ*, *htpG*, *ibpAB* and *groSL* regulons and the GFP monitor cassette from DH10B cells with induced synthetic construct pSB1C3-Lux. The profiles in the top row are for strains containing only the empty plasmid 15 min post-induction; the middle and bottom rows represent strains containing synthetic constructs 15 and 60 min post-induction, respectively. Solid gray regions show average values, and transparent red regions span the minimum to maximum values of transcription profiles generated from three biological replicates. (**e**) Transcriptional response of chromosomal *ibpAB*, *dnaKJ*, *groSL* and *htpG* promoters. The bar graphs show the fold increase in transcription strength in response to construct expression from chromosomal *ibpAB*, *dnaKJ*, *groSL* and *htpG* promoters in DH10B cells after 60 min, as calculated from RNA-seq data. Dots and bars represent single repeats ($n = 3$) and mean values, respectively.

enable robust control of host capacity, regardless of growth conditions and construct design.

To make this as general as possible, we designed a biomolecular feedback controller (**Fig. 3a**) in which the *htpG1* promoter drives the expression of CRISPR single guide RNA (sgRNA), which in turn directs binding of dCas9 to target sequences in specific promoter regions to repress their expression (**Supplementary Fig. 8a**). To enable a rapid response, we designed the construct to express dCas9 constitutively (**Supplementary Table 6**), so that repression occurs as soon as sgRNA is transcribed from the *htpG1* promoter. Because constitutive dCas9 expression is an extra cost for the host, we tuned its expression to a suitably low level (**Supplementary Fig. 8b**). Negative feedback is encoded onto a medium-copy plasmid with a design conceived to be modular and thus easily adapted to act on any synthetic construct. Modularity allows easy replacement of the *htpG1* promoter or sgRNA via rare restriction sites flanking these parts.

We tested the feedback controller in DH10B cells expressing pSB1C3-H2, a high-burden version of pSB1C3-H3 with a stronger ribosome-binding site[17]. The sgRNA of the controller was designed to bind the core pBAD promoter to inhibit transcription of VioB–mCherry. In *E. coli* without the feedback controller (pSB1C3-H2 alone), as the level of arabinose in the media was increased, the VioB–mCherry production rate per cell increased and the host capacity (GFP production rate per cell) decreased. With a feedback plasmid present, VioB–mCherry production remained low even at high arabinose concentrations, whereas the host capacity remained high (**Fig. 3b**). This demonstrates that the controller can limit output (VioB–mCherry) over a range of induction levels to maintain high host expression capacity.

Further testing in a different environmental condition (30 °C) and in two alternative genetic backgrounds (MG1655 and BL21-DE3 *E. coli*) confirmed that, with no changes required, the controller could maintain its ability to confer robustness of host capacity and growth while expressing VioB–mCherry (**Supplementary Figs. 9** and **10**). We also further demonstrated portability by changing the sgRNA targeting sequence to exert control over a firefly luciferase operon expressed from a different promoter (**Supplementary Fig. 11**, **Supplementary Table 7**).

At 37 °C with pSB1C3-H2, the feedback controller limited VioB–mCherry expression to the equivalent of a low induction in cells without feedback (0.002% arabinose). However, when conditions changed (for example, with a shift to 30 °C) the effect of a low induction was typically perturbed. In contrast, with a feedback controller present the performance characteristics were maintained (**Supplementary Fig. 12**). This is

© 2018 Nature America, Inc., part of Springer Nature. All rights reserved.

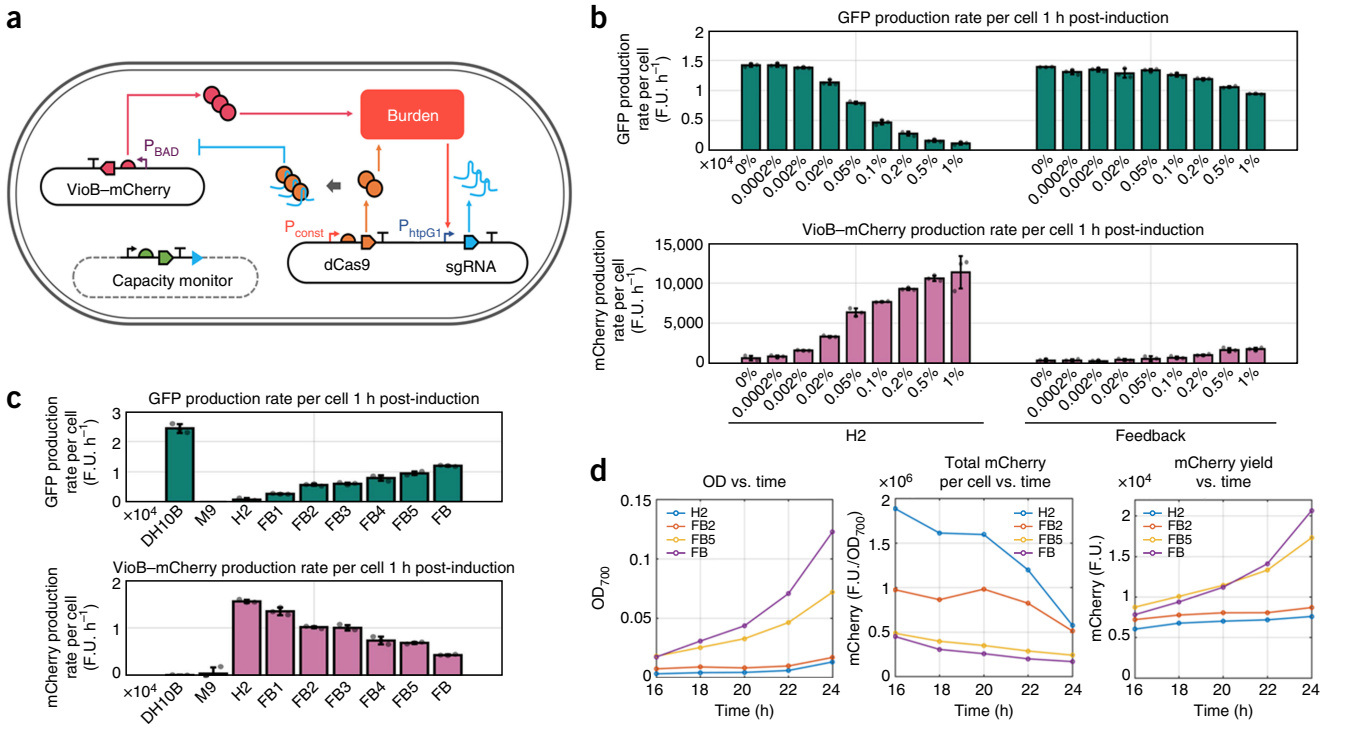

Figure 3 | Burden-driven feedback control system. (**a**) A schematic of the feedback system. P_{BAD}, *E. coli* arabinose-inducible promoter; P_{const}, synthetic low-strength constitutive promoter (see Online Methods); P_{htpG1}, *E. coli htpG* promoter (see **Supplementary Note 2**). (**b**) Functionality of the feedback. Shown are GFP (top) and VioB–mCherry (bottom) production rates with different levels of arabinose used to trigger expression in DH10B cells. Data are shown as the mean ± s.d. of three independent repeats (black dots). (**c**) Tunability of the feedback system. A library of feedback controllers was designed in which single point mutations were inserted in the sgRNA sequence to decrease binding affinity to the pBAD promoter. Shown are GFP (top) and VioB–mCherry (bottom) production rates 1 h after induction. Data are shown as the mean ± s.d. of three independent repeats (black dots). (**d**) 24-h time-course experiment. Constructs FB, FB2 and FB5 were chosen for a 24-h batch growth time course, with 1% arabinose added to the cultures containing the feedback system and pSB1C3-H2 construct. The plots compare the culture density (OD) and total VioB–mCherry yield in growth flasks from 16 to 24 h for one example experiment that was repeated three times, with similar results each time. F.U., fluorescence units.

because small changes in inducer levels at low concentrations have dramatic effects on construct output. With feedback, both expression output and the capacity of the cells were maintained within a narrow window, even over large changes in inducer levels (**Supplementary Fig. 13**).

Because different applications may require more construct output at the cost of host capacity, we next tuned the feedback strength. We designed mismatches into the sgRNA targeting sequence of the feedback controller (**Supplementary Table 6**). Mismatches reduce the affinity of the dCas9–sgRNA complex for its target promoter and thus decrease the strength of repression[23]. Random mutations at single positions within the sgRNA targeting sequence modified the feedback gain, leading to increased pSB1C3-H2 output and decreased host capacity (**Fig. 3c**). This simple approach of creating sgRNA mismatches enables the creation of a library of feedback controllers with different maximum construct outputs and host capacities.

We assessed DH10B cells expressing pSB1C3-H2 along with two mismatch controllers (FB2 and FB5) or with the original controller (FB) at both population and single-cell levels in batch cultures over 24 h and compared them to cells with pSB1C3-H2 only (**Supplementary Figs. 14** and **15**). As a proxy for total production of synthetic gene product, we measured the total red fluorescence of the culture. At up to 16 h of growth, the most productive culture (as determined by the total amount of VioB–mCherry per flask) was the strain with mismatch feedback controller FB5. However, after 24 h, the culture with the strongest feedback gain (FB) yielded the largest amount of total product, presumably because this strain grew the fastest over the experiment and therefore produced more biomass. Thus, although the cells with these feedback controllers had lower construct production rates, they showed greater total production over time, because their increased capacity enabled considerably improved growth (**Fig. 3d**).

DISCUSSION

All constructs tested in this study, regardless of genetic content, reduced host expression capacity and consumed a major fraction of transcriptional resources. RNA-seq gave new insights into how cells respond to gene overexpression, and data obtained here provide new information about resource reallocation and highlight that effects on transcription are a major factor in burden (further discussed in **Supplementary Note 3**). This could be investigated in future work through the combination of transcriptome measurements with mRNA-per-cell quantification[24], mRNA half-life data[25] and Ribo-seq, a method that quantifies the allocation of ribosomes in cells[26]. The recently described RNAtag-seq method[27] also greatly simplifies multiplex RNA-seq and should enable our approach to be extended to other strains and synthetic

© 2018 Nature America, Inc., part of Springer Nature. All rights reserved.

constructs, such as those at lower copy numbers, with different functions (e.g., secreted proteins, RNA-only devices) or with dynamic expression (e.g., oscillators, logic gates).

All tested constructs triggered upregulation of σ^{32}-regulated promoters, and as this was also seen with pD864-LacZ, which expresses only *E. coli* genes, this demonstrates that the heat-shock response, known to be associated with overexpression of recombinant proteins[28], is also triggered when native genes are expressed in non-native contexts (and thus is not necessarily avoided with codon optimization). The σ^{32} response therefore appears to be a rapid and sensitive mechanism for cells to adapt resources after unexpected gene expression[29,30]. This suggests that the burden of expression is more than just consumption of polymerases and ribosomes and may also involve limited pools of translational cofactors such as the chaperones and proteases that are regulated by σ^{32}.

By exploiting the *htpG1* promoter's response to burden, we produced a tunable and modular dCas9-based feedback controller that can be used as a simple regulator plasmid alongside synthetic constructs. Because the only burden-activated component is the guide RNA, it has fast dynamics and can be easily retargeted, which makes it a major improvement on feedback devices that require specific promoter–regulator pairs[31]. It also has no requirement for strain modification, unlike a method that achieves protein overexpression via mutation of σ^{32} (ref. 28), or a recent genome-integrated quorum-based regulator that controls metabolic pathways[32].

Theoretically, the controller can regulate many genes by coexpression of multiple sgRNAs, and as homologs of σ^{32} from different bacteria have been shown to trigger *E. coli* σ^{32} promoters[33], it is possible that it could also be functional in other bacteria with only minor refactoring. As it confers robustness even in changed environmental conditions, it offers an interesting new way to maintain growth and expression in the dynamic and heterogeneous conditions typically seen in scaled-up bioproduction. A more complete understanding of the dynamic behavior of the feedback could be achieved in future work through assessment of the controller's performance in response to multiple fluctuating conditions (for example, in chemostat experiments).

METHODS

Methods, including statements of data availability and any associated accession codes and references, are available in the online version of the paper.

Note: Any Supplementary Information and Source Data files are available in the online version of the paper.

ACKNOWLEDGMENTS

This work was supported by EPSRC (grant EP/J021849/1 to F.C., G.-B.S. and T.E.; grant EP/M002306/1 to A.B., O.B. and C.G.; grant EP/P009352/1 to G.-B.S.; Fellowship EP/M002187/1 to G.-B.S.; Fellowship EP/M002306/1 to T.E.), BBSRC/EPSRC SBRC BrisSynBio (grant BB/L01386X/1 to T.E.G.), the Royal Society (Fellowship UF160357 to T.E.G.), the NIHR Imperial Biomedical Research Centre (Y.N.L.) and BBSRC (grant BB/K006290/1 to A.R.A.). F.C. acknowledges the support of the Imperial College London Junior Research Fellowship Scheme. All authors thank G. Cambray for useful discussions.

AUTHOR CONTRIBUTIONS

F.C., G.-B.S. and T.E. designed the research; F.C., A.B., C.G. and A.R.A. performed the experiments; F.C., S.F., T.E.G., Y.N.L., G.-B.S. and T.E. analyzed the data; F.C., S.F., T.E.G., O.B., G.-B.S. and T.E. wrote the paper.

COMPETING INTERESTS

The authors declare no competing interests.

Reprints and permissions information is available online at http://www.nature.com/reprints/index.html. Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

1. Lynch, M. & Marinov, G.K. The bioenergetic costs of a gene. *Proc. Natl. Acad. Sci. USA* **112**, 15690–15695 (2015).
2. Gyorgy, A. *et al.* Isocost lines describe the cellular economy of genetic circuits. *Biophys. J.* **109**, 639–646 (2015).
3. Qian, Y., Huang, H.-H., Jiménez, J.I. & Del Vecchio, D. Resource competition shapes the response of genetic circuits. *ACS Synth. Biol.* **6**, 1263–1272 (2017).
4. Weiße, A.Y., Oyarzún, D.A., Danos, V. & Swain, P.S. Mechanistic links between cellular trade-offs, gene expression, and growth. *Proc. Natl. Acad. Sci. USA* **112**, E1038–E1047 (2015).
5. Kurland, C.G. & Dong, H. Bacterial growth inhibition by overproduction of protein. *Mol. Microbiol.* **21**, 1–4 (1996).
6. Sleight, S.C. & Sauro, H.M. Visualization of evolutionary stability dynamics and competitive fitness of *Escherichia coli* engineered with randomized multigene circuits. *ACS Synth. Biol.* **2**, 519–528 (2013).
7. Cardinale, S. & Arkin, A.P. Contextualizing context for synthetic biology—identifying causes of failure of synthetic biological systems. *Biotechnol. J.* **7**, 856–866 (2012).
8. Moser, F. *et al.* Genetic circuit performance under conditions relevant for industrial bioreactors. *ACS Synth. Biol.* **1**, 555–564 (2012).
9. Tan, C., Marguet, P. & You, L. Emergent bistability by a growth-modulating positive feedback circuit. *Nat. Chem. Biol.* **5**, 842–848 (2009).
10. Borkowski, O., Ceroni, F., Stan, G.-B. & Ellis, T. Overloaded and stressed: whole-cell considerations for bacterial synthetic biology. *Curr. Opin. Microbiol.* **33**, 123–130 (2016).
11. Nielsen, A.A.K. *et al.* Genetic circuit design automation. *Science* **352**, aac7341 (2016).
12. Cameron, D.E. & Collins, J.J. Tunable protein degradation in bacteria. *Nat. Biotechnol.* **32**, 1276–1281 (2014).
13. Segall-Shapiro, T.H., Meyer, A.J., Ellington, A.D., Sontag, E.D. & Voigt, C.A.A. A 'resource allocator' for transcription based on a highly fragmented T7 RNA polymerase. *Mol. Syst. Biol.* **10**, 742 (2014).
14. Pasini, M. *et al.* Using promoter libraries to reduce metabolic burden due to plasmid-encoded proteins in recombinant *Escherichia coli*. *N. Biotechnol.* **33**, 78–90 (2016).
15. Gorochowski, T.E., Avcilar-Kucukgoze, I., Bovenberg, R.A.L., Roubos, J.A. & Ignatova, Z. A minimal model of ribosome allocation dynamics captures trade-offs in expression between endogenous and synthetic genes. *ACS Synth. Biol.* **5**, 710–720 (2016).
16. Carrera, J., Rodrigo, G., Singh, V., Kirov, B. & Jaramillo, A. Empirical model and in vivo characterization of the bacterial response to synthetic gene expression show that ribosome allocation limits growth rate. *Biotechnol. J.* **6**, 773–783 (2011).
17. Ceroni, F., Algar, R., Stan, G.-B. & Ellis, T. Quantifying cellular capacity identifies gene expression designs with reduced burden. *Nat. Methods* **12**, 415–418 (2015).
18. Shachrai, I., Zaslaver, A., Alon, U. & Dekel, E. Cost of unneeded proteins in *E. coli* is reduced after several generations in exponential growth. *Mol. Cell* **38**, 758–767 (2010).
19. Wang, Z., Gerstein, M. & Snyder, M. RNA-Seq: a revolutionary tool for transcriptomics. *Nat. Rev. Genet.* **10**, 57–63 (2009).
20. Houser, J.R. *et al.* Controlled measurement and comparative analysis of cellular components in *E. coli* reveals broad regulatory changes in response to glucose starvation. *PLOS Comput. Biol.* **11**, e1004400 (2015).
21. Nonaka, G., Blankschien, M., Herman, C., Gross, C.A. & Rhodius, V.A. Regulon and promoter analysis of the *E. coli* heat-shock factor, σ32, reveals a multifaceted cellular response to heat stress. *Genes Dev.* **20**, 1776–1789 (2006).
22. Guisbert, E., Herman, C., Lu, C.Z. & Gross, C.A. A chaperone network controls the heat shock response in *E. coli*. *Genes Dev.* **18**, 2812–2821 (2004).
23. Farasat, I. & Salis, H.M. A biophysical model of CRISPR/Cas9 activity for rational design of genome editing and gene regulation. *PLOS Comput. Biol.* **12**, e1004724 (2016).

© 2018 Nature America, Inc., part of Springer Nature. All rights reserved.

24. Jiang, L. *et al.* Synthetic spike-in standards for RNA-seq experiments. *Genome Res.* **21**, 1543–1551 (2011).
25. Arkin, A.P. & Cambray, G. Massive phenotypic measurements reveal complex physiological consequences of differential translation efficacies. *bioRxiv* Preprint at https://www.biorxiv.org/content/early/2017/10/25/209098 (2017).
26. 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).
27. Lohman, B.K., Weber, J.N. & Bolnick, D.I. Evaluation of TagSeq, a reliable low-cost alternative for RNAseq. *Mol. Ecol. Resour.* **16**, 1315–1321 (2016).
28. Zhang, X. *et al.* Heat-shock response transcriptional program enables high-yield and high-quality recombinant protein production in *Escherichia coli*. *ACS Chem. Biol.* **9**, 1945–1949 (2014).
29. El-Samad, H., Kurata, H., Doyle, J.C., Gross, C.A. & Khammash, M. Surviving heat shock: control strategies for robustness and performance. *Proc. Natl. Acad. Sci. USA* **102**, 2736–2741 (2005).
30. Kurata, H. *et al.* Module-based analysis of robustness tradeoffs in the heat shock response system. *PLOS Comput. Biol.* **2**, e59 (2006).
31. Dragosits, M., Nicklas, D. & Tagkopoulos, I. A synthetic biology approach to self-regulatory recombinant protein production in *Escherichia coli*. *J. Biol. Eng.* **6**, 2 (2012).
32. Gupta, A., Reizman, I.M., Reisch, C.R. & Prather, K.L.J. Dynamic regulation of metabolic flux in engineered bacteria using a pathway-independent quorum-sensing circuit. *Nat. Biotechnol.* **35**, 273–279 (2017).
33. Nakahigashi, K., Yanagi, H. & Yura, T. Regulatory conservation and divergence of 32 homologs from *Pseudomonas aeruginosa*, and *Agrobacterium tumefaciens*. *Microbiology* **180**, 2402–2408 (1998).

© 2018 Nature America, Inc., part of Springer Nature. All rights reserved.

ONLINE METHODS

Bacterial strains. Strains MG1655 (K-12 F- λ- *rph*-1) and DH10B (K-12 F- λ- *araD139* Δ*(araA-leu)7697* Δ*(lac)X74 galE15 galK16 galU hsdR2 relA rpsL150*(StrR) *spoT1 deoR* φ*80dlacZ*ΔM15 *endA1 nupG recA1 e14- mcrA* Δ*(mrr hsdRMS mcrBC)*) were obtained from the National BioResource Project Japan. Strain BL21(DE3) (F^{-}*ompT gal dcm lon hsdS*$_B$($r_B^{-}m_B^{-}$) λ(DE3 [*lacI lacUV5-T7p07 ind1 sam7 nin5*]) [*malB*$^{+}$]$_{K\text{-}12}$(λ^S)) was obtained from New England Biolabs (NEB). The capacity monitor cassette consisted of a synthetic strong constitutive promoter, a synthetic ribosome-binding site, a codon-optimized superfolding GFP coding sequence and a synthetic unnatural bidirectional terminator. Details of the construction and validation of the capacity monitor used in this study were published previously[17].

DNA construct design. The four expression constructs providing inducible heterologous gene expression are illustrated in **Figure 1b** with genetic designs drawn using SBOL Visual notation. Construction of constructs pSB1C3-H3, pLys-M1 and pSB1C3-Lux was described previously[17]. Construct pD864-LacZ was obtained by PCR amplification of the *lacZ* coding sequence from the *E. coli* genome and restriction digestion into the pD864-SR plasmid (DNA2.0). Control plasmids were the high-copy, chloramphenicol-selectable pSB1C3 plasmid obtained by restriction digest and re-ligation; the medium-copy, chloramphenicol-selectable pLys plasmid and the high-copy, ampicillin-selectable pD864 plasmid, both obtained by PCR amplification and re-ligation to remove the synthetic construct insert.

Construction of plasmid-based *htpG1*, *htpG2*, *groSL* and *ibpAB* promoters. Restriction digestion and re-ligation were used after PCR amplification of plasmid YTK095, which codes for GFP and ampicillin resistance and contains the pMB1 origin of replication. PCR was used to add SfiI and PacI restriction sites at the 5′ and 3′ ends of the plasmid, respectively. gBlocks for each of the promoters were ordered from IDT and contained the SfiI and PacI restriction sites upstream and downstream of the promoter sequence, respectively (**Supplementary Table 5**).

Design of the burden-based feedback system. The dCas9–sgRNA feedback controller was placed on a medium-copy plasmid and was initially designed to control expression of a construct on a high-copy plasmid. In the feedback construct the *htpG1* promoter was placed upstream of an sgRNA targeting the synthetic construct promoter. A constitutively expressed dCas9 cassette was also encoded on the feedback control plasmid. When the high-copy synthetic construct was expressed and burden was consequently triggered in the cell, the *htpG1* promoter activated production of sgRNA. The sgRNA–dCas9 complex bound to and repressed the synthetic contrust promoter.

Construction of the burden-based feedback system. The sequence of the dCas9 protein was PCR amplified from a commercial plasmid and inserted on the medium-copy plasmid pZA16, carrying ampicillin resistance, downstream of the constitutive BBa_J23113 promoter. The reverse primer for dCas9 amplification was designed so that it could bear SfiI/PacI/AscI restriction sites. A GeneString containing the *htpG1* promoter sequence was ordered from GeneArt with the SfiI and PacI restriction sites at the 5′ and 3′ ends, respectively. The sgRNA targeting the pBAD promoter was ordered from GeneArt with PacI and AscI restriction sites at the 5′ and 3′ ends, respectively. Both the promoter and the sgRNA were inserted downstream of dCas9 using restriction digestion and re-ligation. All enzymes used for cloning were obtained from NEB.

We used inverse PCR to create a library of randomly mutated versions of the BBa_J23113 promoter upstream of dCas9, to investigate whether higher expression of dCas9 protein would increase the strength of the feedback without itself imposing significant burden. The promoter sequence selected for final implementation in the feedback system is shown in **Supplementary Table 6**.

Construction of the library of randomly point-mutated sgRNAs was done by inverse PCR with insertion-encoding 5′ phosphorylated primers, followed by DpnI digestion and re-ligation before transformation in DH10B cells. Six PCR reactions were carried out separately with six forward primers, each carrying a mutation at a different position. All PCRs were carried out with NEB Phusion High Fidelity Polymerase. Cotransformation of the library with H2 was then performed in DH10B cells. Seventeen colonies were picked and subjected to DNA sequencing to identify the position of the point mutation in the sgRNA of each of these colonies. The constructs were then cotransformed together with pSB1C3-H2 in DH10B cells. Sequences of the sgRNA library members are shown in **Supplementary Table 6**.

Construction of the open-loop counterpart of the feedback plasmid was done via inverse PCR to add BsaI sites around the *htpG1* promoter and the sgRNA targeting pBAD. A 272-bp gBlock was ordered and synthesized by IDT. The gBlock contained BsaI sites around the *htpG1* promoter, and the random sgRNA (**Supplementary Table 6**) was designed using R2o DNA Designer[34]. The two strings were assembled via Golden Gate assembly, and the resulting plasmid was transformed in DH10B monitor cells. The construct was verified by DNA sequencing before use. The open-loop version was then cotransformed with pSB1C3-H2 in DH10B monitor cells.

Construction of the luciferase-expressing plasmid. For construction of the luciferase plasmid with the constitutive LacP promoter, we used inverse PCR on the BBa_K325219 construct (http://parts.igem.org/Part:BBa_K325219 and ref. 17) to replace the promoter present upstream of the operon with the core sequence of the LacP promoter, a known, strong constitutive promoter based on the native Lac promoter and devoid of downstream operators. The sequence of this promoter is given in **Supplementary Table 7**.

Modification of the feedback controller to target the luciferase construct. Construction of the feedback-controller plasmid with sgRNA targeting the LacP promoter was done by restriction digestion of the AraBad-targeting controller (FB) with AscI and PacI enzymes so as to remove the original sgRNA sequence. This sequence was then replaced by a synthesized double-stranded DNA fragment (IDT gBlock) encoding an sgRNA designed to target the core sequence of the LacP promoter. The corresponding sequence is given in **Supplementary Table 7**.

Burden assay and RNA-seq time course. For the burden assay and time course, *E. coli* cells with construct and control plasmids were grown at 37 °C overnight with aeration in a shaking incubator in 5 ml

© 2018 Nature America, Inc., part of Springer Nature. All rights reserved.

of defined supplemented M9 medium (see below). In the morning, 60 µl of each sample was diluted into 3 ml of fresh media and grown at 37 °C with shaking for another hour (outgrowth). 200 µl of each sample were then transferred into 8 wells of a 96-well plate (Costar) at approximately 0.1 OD_{600}. The samples were placed in a Synergy HT microplate reader (BioTek) and incubated at 37 °C with orbital shaking at 1,000 r.p.m. for 1 h, with measurements of GFP (excitation, 485 nm; emission, 528 nm) and OD_{600} taken every 15 min. 60 min into the incubation, the plate was briefly removed so inducer could be added to wells, and this time point was set as time 0. In the burden assay, cells were allowed to grow in the reader for 4.5 h, with measurements of GFP (excitation, 485 nm; emission, 528 nm) and OD_{600} taken every 15 min. In the RNA-seq analysis, samples were instead removed from wells at 15 and 60 min after induction for processing. Specifically, 170 µl were taken from each of four wells per time point and collected in a fresh tube to which 1.360 ml of RNA protection buffer had previously been added. Samples were left for 5 min at room temperature and then centrifuged at 4 °C at maximum speed. Supernatant was discarded and pellets were frozen at –20 °C. Three replicates were repeated independently on three different days for a total of 90 samples used to produce the final data set (7 constructs × 2 strains × 3 replicates × 2 time points = 84 samples; plus DH10B:GFP cells × 3 replicates × 2 time points) (**Supplementary Table 2**). M9 media used here consisted of M9 salts supplemented with 0.4% casamino acids, 0.25 mg/ml thiamine hydrochloride, 2 mM $MgSO_4$, 0.1 mM $CaCl_2$, 0.4% fructose and the appropriate antibiotic. We used fructose as the main carbon source to avoid the strong catabolite repression of AraBAD and RhaBAD promoters known to occur in glucose media. Inducers were added to appropriate wells at 0 min. Final inducer and antibiotic concentrations used in assays were as follows: L-arabinose, 0.2%; L-rhamnose, 2%; ampicillin, 100 µg/ml; and chloramphenicol, 34 µg/ml.

Data analysis. For the plate-reader characterization assays, the growth and the protein expression rates per hour were calculated according to the procedure in ref. 17. Growth rate at $t2$ = $(\ln(OD(t3)) - \ln(OD(t1)))/(t3 - t1)$, GFP production rate at $t2$ = $((\text{Total GFP}(t3) - \text{Total GFP}(t1))/(t3 - t1))/OD(t2)$, and mCherry production rate at $t2$ = $((\text{Total mCherry}(t3) - \text{Total mCherry}(t1))/(t3 - t1))/OD(t2)$, where $t1$ = present time – 15 min, $t2$ = present time, and $t3$ = present time + 15 min. Mean rates and their s.e. were determined from three biological replicates from the same 96-well plate. 400 mCherry F.U. h^{-1} was added to all mCherry output rates to account for the background red fluorescence of the media, which decreases at a rate of approximately 400 F.U. h^{-1} as it is consumed by cells during growth[17].

RNA-seq library preparation. For library preparation we used a custom protocol adapted from previous Nextera kit methods[35]. Total RNA was first extracted and assessed (RNA integrity number > 9), and mRNA was enriched by rRNA removal[36] and retro-transcribed to cDNA. The Nextera XT protocol was used for library synthesis with tagmentation to fragment cDNAs and attach adaptor sequences. Limited-cycle PCR targeting the adaptors amplified insert DNA and added barcodes for dual-index sequencing of pooled samples.

RNA extraction was done with the RNeasy mini kit (Qiagen; 74104). To remove possible traces of genomic DNA contamination, we treated 2 µg of each sample for a second time with DNase I (Qiagen; 79254). Total RNA quality and integrity were assessed with an Agilent 2100 Bioanalyzer and Agilent RNA 6000 nano kit (5067-1511). Samples had an average RNA integrity number of 9.5. Enrichment of mRNA was done with the MicrobExpress rRNA removal kit (Thermo Scientific; AM1905)[36]. Successful rRNA depletion was assessed by analysis on the Bioanalyzer. Retrotranscription was then carried out starting from 50 ng of total enriched mRNA with the Tetro cDNA synthesis kit (Bioline; BIO-65043) and 6 µl of random hexamers (Bioline; BIO-38028) per reaction. For the second cDNA synthesis we added to the first-strand synthesis mix 5 µl of NEBNext second-strand synthesis buffer (NEB; B6117S), 3 µl of dNTPs (NEB; N0446S), 2 µl of RNase H (NEB; M0297L), 2 µl of polymerase I (Thermo Scientific; 18010025) and 18 µl of water per reaction. Samples were incubated at 16 °C for 2.5 h. Purification of cDNA was done with the MiniElute PCR purification kit (Qiagen; 28004) with final elution in 10 µl of DEPC-treated free water. cDNA was quantified with a Qubit fluorometer (Invitrogen). Library preparation was carried out with the Nextera XT kit (Illumina; FC-131-1096) starting from 1 ng of total cDNA. The original protocol was modified where 3 min of tagmentation and 13 cycles of step-limited PCR were used. Ampure beads (Beckman Coulter; A63880) were used for library purification. Library quality assessment and quantification were done with an Agilent 2100 Bioanalyzer and Agilent high-sensitivity DNA analysis kit (5067-4626). Finally, all 90 samples were pooled together in the same reaction tube at a final concentration of 1 nM.

RNA-seq library sequencing. Library sequencing was performed at the Imperial College London Genomic Facility. Two lanes from the HiSeq 2500 sequencer were used for paired-end sequencing with read length of 100 bp.

Sequencing quality control and alignments. Raw reads for all sequenced samples were quality-assessed and trimmed using Trim Galore v0.4.1 with default settings. After assessment for potential batch effects, the technical replicates were pooled. *E. coli* strain (DH10B and MG1655) sequences were obtained from Ensembl genomes release 31. A FASTA format sequence file corresponding to the composite of strain, plasmid and integrated GFP was constructed for each sample and used as a reference for read alignment. Trimmed reads were aligned using BWA mem algorithm v0.7.12-r1039 with the default settings. SAMtools v1.3.1 was used on resultant alignments to create a sorted BAM file for each sample. The biological replicates were checked for any batch effects before the raw counts were generated using Bioconductor Rsubread package v1.12.6. All reads identified as unremoved rRNA were discarded, and in the one case where reads could align to either the plasmid or the strain genome (for *lacZ* in pD864-LacZ in MG1655), the raw reads were assigned appropriately to match those of flanking sequence. Biological replicates were checked before the raw counts were generated, and this identified one outlier sample from the 90 sequences (the first replicate of a DH10B empty strain control). The data from this sample were thus discarded. The normalized FPKM counts were generated with Bioconductor edgeR package version 3.4.2, accounting for gene length and library size (by TMM normalization), which were used for downstream analysis.

© 2018 Nature America, Inc., part of Springer Nature. All rights reserved.

doi:10.1038/nmeth.4635

Transcription profiles and promoter characterization. The method described by Gorochowski *et al.*[37] was used to generate the transcription profiles from RNA-seq data. Raw reads from the sequencer in a FASTQ format were mapped to the appropriate *E. coli* host genome reference sequence (NCBI RefSeq NC_000913.3 for MG1655, NCBI RefSeq NC_010473.1 for DH10B), with separate reference sequences included for the GFP monitor and appropriate plasmid constructs using BWA version 0.7.4 with default settings. These BAM files were then separately processed with custom Python scripts to extract the position of the mapped reads, count read depths across the reference sequences, and apply corrections to the profiles at the ends of transcription units. These profiles were then normalized to enable comparisons between samples. Characterization of promoters was done with custom Python scripts as in ref. 37, which took as input a GFF reference of the construct defining the location of all parts. Visualizations of the transcription profiles and associated genetic design information were generated in an SBOL Visual[38] format using DNAplotlib version 1.0 (ref. 39). All analyses were carried out with custom scripts run using Python version 2.7.12, NumPy version 1.11.2, and matplotlib version 1.5.3.

To avoid mislabeling of the reads when the transcription profiles for the pD864-LacZ construct were generated (for example, assigning to the plasmid gene reads to the chromosomal gene, and vice versa), only reads uniquely mapping to the *lacZ* gene in the plasmid construct, and not the chromosomal copy, were displayed. Blue shaded regions in **Supplementary Figure 3** show estimated levels of expression from the *lacZ* gene based on these data.

Differential gene expression. DESeq2 was used for differential expression analyses[40]. Gene expression was compared between cells transformed with synthetic constructs and the analogous cells transformed with the corresponding empty plasmid. Reads mapping to ribosomal genes or to the synthetic constructs were excluded. Differentially expressed genes were annotated with data extracted from the EcoCyc database[41] using custom Python code.

Testing the robustness of the feedback system. For plate-reader assays to test the robustness of the feedback system to arabinose-induction perturbation, the constructs were induced with the following final concentrations of arabinose: 0%, 0.0002%, 0.002%, 0.02%, 0.05%, 0.1%, 0.2%, 0.5% and 1%.

Shake-flask-scale growth. Constructs pSB1C3-H2, FB+H2, FB2+H2 and FB5+H2 were assessed in DH10B cells over 24 h of exponential growth in M9 fructose (0.4%) media with 1% final concentration of arabinose. The experiment was done as previously described[17]. Starter cultures of pSB1C3-H2, FB+H2, FB2+H2 and FB5+H2 were taken on plate from individual colonies and used to inoculate 3 ml of M9 fructose media, supplemented with the appropriate antibiotics, in 15-ml culture tubes. The cultures were then grown in the 37 °C shaking incubator for 5 h before being diluted to 0.015 OD_{600}. 50 µl of the diluted culture (~150,000 cells) was used to inoculate batch cultures of 50 ml of supplemented M9 with 0.4% fructose, 1% L-arabinose, and the appropriate antibiotics (ampicillin, 100 g/ml (50 g/ml when both ampicillin and chloramphenicol were added to the medium), and chloramphenicol, 34 g/ml (17 g/ml when both ampicillin and chloramphenicol were added to the medium)) in 500-ml baffled shake flasks. The cultures were then grown in the 37 °C shaking incubator for 16 h, after which 200 µl of each culture was dispensed into individual wells of a 96-well plate, and 350 µl was diluted in 650 µl of PBS and stored on ice every 60 min from 16 h until 24 h of culture growth. The 96-well plate was placed in a preheated plate reader at 37 °C every 60 min for OD measurements (OD_{600} and OD_{700}), and for GFP measurements (485 nm (excitation)/528 nm (emission)) and mCherry measurements (590 nm/645 nm) every 2 min. Only the first five measurements were taken and averaged to obtain the OD, GFP and mCherry values at specific time points. At the end of the experiment, we ran the PBS-diluted aliquots stored on ice on a BD LSR Fortessa X20 flow cytometer to obtain single-cell data.

To obtain equivalent data for early time points, we started pre-cultures by inoculating 6 ml of M9 fructose, supplemented with the appropriate antibiotics, with a single colony. Pre-cultures were then incubated in the 37 °C shaking incubator for 8.5 h. In the morning, cultures were diluted to OD_{600} = 0.03 in 50 ml and grown in the shaking incubator for 2 h before induction with 1% arabinose. Readings were taken every 2 h in a 96-well plate.

Luminescence assay. To read luminescence from the firefly Luc operon, we added 100 µM of D-luciferin (Sigma-Aldrich) to each well at the time of IPTG induction. Luminescence readings were taken in a plate-reader every 15 min (emission: 645/640 nm). Luminescence per cell was calculated as follows:

$$\text{Luminescence per cell}(t) = (\text{Total luminescence}(t))/(\text{OD}(t))$$

where t = present time post-induction.

Mean rates and s.e. were determined from three biological replicates from the same 96-well plate.

Code availability. Custom codes used in this research will be made available by the corresponding author upon reasonable request.

Life Sciences Reporting Summary. Further information on experimental design is available in the **Life Sciences Reporting Summary**.

Data availability. RNA-seq data from this study have been deposited in NCBI's Gene Expression Omnibus[42] (GSE107093). Excel sheet summaries of all the mapped reads and calculated FPKM values for genome and plasmid-derived transcripts in the DH10B and MG1655 experiments are provided as **Supplementary Data Sets 1** and **2**, respectively. Experimental data on bacterial growth and fluorescence can be obtained from the corresponding author upon reasonable request. Source data for **Figures 1** and **3** are available online.

34. Casini, A. *et al.* R2oDNA designer: computational design of biologically neutral synthetic DNA sequences. *ACS Synth. Biol.* **3**, 525–528 (2014).
35. Gertz, J. *et al.* Transposase mediated construction of RNA-seq libraries. *Genome Res.* **22**, 134–141 (2012).
36. He, S. *et al.* Validation of two ribosomal RNA removal methods for microbial metatranscriptomics. *Nat. Methods* **7**, 807–812 (2010).

© 2018 Nature America, Inc., part of Springer Nature. All rights reserved.

37. Gorochowski, T.E. *et al.* Genetic circuit characterization and debugging using RNA-seq. *Mol. Syst. Biol.* **13**, 952 (2017).
38. Myers, C.J. *et al.* A standard-enabled workflow for synthetic biology. *Biochem. Soc. Trans.* **45**, 793–803 (2017).
39. Der, B.S. *et al.* DNAplotlib: programmable visualization of genetic designs and associated data. *ACS Synth. Biol.* **6**, 1115–1119 (2017).
40. Love, M.I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15**, 550 (2014).
41. Keseler, I.M. *et al.* The EcoCyc database: reflecting new knowledge about *Escherichia coli* K-12. *Nucleic Acids Res.* **45**, D543–D550 (2017).
42. Edgar, R., Domrachev, M. & Lash, A.E. Gene Expression Omnibus: NCBI gene expression and hybridization array data repository. *Nucleic Acids Res* **30**, 207–210 (2002).

© 2018 Nature America, Inc., part of Springer Nature. All rights reserved.

doi:10.1038/nmeth.4635

nature research

Corresponding author(s): Tom Ellis and Guy-Bart Stan

☐ Initial submission ☐ Revised version ☒ Final submission

Life Sciences Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form is intended for publication with all accepted life science papers and provides structure for consistency and transparency in reporting. Every life science submission will use this form; some list items might not apply to an individual manuscript, but all fields must be completed for clarity.

For further information on the points included in this form, see Reporting Life Sciences Research. For further information on Nature Research policies, including our data availability policy, see Authors & Referees and the Editorial Policy Checklist.

Experimental design

1. Sample size

Describe how sample size was determined.

In this study we analyzed a total 90 samples with RNAseq. This sample size was sufficient as it was designed to be representative of the common circuits used in Synthetic Biology with triplicate coverage of all samples. The experiment was designed so to have biological triplicates for 4 synthetic constructs transformed in E. coli cells plus control samples corresponding to the test samples. Two strains were selected so that burden could be investigated in two different systems. Finally, two time points were chosen to look at the burden response, 15 and 60 min after induction of gene expression. Thus, DH10B and MG1655 cells transformed with 4 test constructs, their 3 empty plasmid were selected together with empty DH10B as control of untransformed cells.

2. Data exclusions

Describe any data exclusions.

We excluded the RNAseq data for one of the triplicates of the DH10B empty strain as the sample showed lower library abundance and, as a consequence, behaved as an out-lier when compared to the other two replicates.

3. Replication

Describe whether the experimental findings were reliably reproduced.

RNAseq analysis was performed on 3 biological replicates for each sample.
All 90 reaction steps from total RNA to library preparation were successful but the library preparation for the first replicate for DH10 empty cells led to insufficient DNA amount for exhaustive sequencing quality. This sample was then excluded from the analysis. In vivo burden analysis was performed on 2 biological replicates. The biomolecular feedback device characterization and test were performed in 3 biological replicates.

4. Randomization

Describe how samples/organisms/participants were allocated into experimental groups.

Samples were organised into experimental groups based on the aim of the experiment (e.g. for Figure sub-panels). Randomization was not relevant for this study as position of each sample was not likely to influence the outcome.

5. Blinding

Describe whether the investigators were blinded to group allocation during data collection and/or analysis.

No blinding was not use during this study. Blinding was not relevant in the present study as no anticipated result was expected.

Note: all studies involving animals and/or human research participants must disclose whether blinding and randomization were used.

Nature Methods: doi:10.1038/nmeth.4635

6. Statistical parameters

For all figures and tables that use statistical methods, confirm that the following items are present in relevant figure legends (or in the Methods section if additional space is needed).

n/a	Confirmed
☐	☒ The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement (animals, litters, cultures, etc.)
☐	☒ A description of how samples were collected, noting whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☐	☒ A statement indicating how many times each experiment was replicated
☒	☐ The statistical test(s) used and whether they are one- or two-sided (note: only common tests should be described solely by name; more complex techniques should be described in the Methods section)
☐	☒ A description of any assumptions or corrections, such as an adjustment for multiple comparisons
☐	☒ The test results (e.g. *P* values) given as exact values whenever possible and with confidence intervals noted
☐	☒ A clear description of statistics including central tendency (e.g. median, mean) and variation (e.g. standard deviation, interquartile range)
☐	☒ Clearly defined error bars

See the web collection on statistics for biologists for further resources and guidance.

Software

Policy information about availability of computer code

7. Software

Describe the software used to analyze the data in this study.	Excel 2010, Python 2.7 and R (DESeq2_1.10.1) were used for data analysis and simulations

For manuscripts utilizing custom algorithms or software that are central to the paper but not yet described in the published literature, software must be made available to editors and reviewers upon request. We strongly encourage code deposition in a community repository (e.g. GitHub). *Nature Methods* guidance for providing algorithms and software for publication provides further information on this topic.

Materials and reagents

Policy information about availability of materials

8. Materials availability

Indicate whether there are restrictions on availability of unique materials or if these materials are only available for distribution by a for-profit company.	no unique materials were used

9. Antibodies

Describe the antibodies used and how they were validated for use in the system under study (i.e. assay and species).	None used

10. Eukaryotic cell lines

a. State the source of each eukaryotic cell line used.	None used
b. Describe the method of cell line authentication used.	None used
c. Report whether the cell lines were tested for mycoplasma contamination.	None used
d. If any of the cell lines used are listed in the database of commonly misidentified cell lines maintained by ICLAC, provide a scientific rationale for their use.	None used

Animals and human research participants

Policy information about studies involving animals; when reporting animal research, follow the ARRIVE guidelines

11. Description of research animals

Provide details on animals and/or animal-derived materials used in the study.	None used

Nature Methods: doi:10.1038/nmeth.4635

Policy information about studies involving human research participants

12. Description of human research participants

Describe the covariate-relevant population characteristics of the human research participants.	None used

Nature Methods: doi:10.1038/nmeth.4635

nature research

Corresponding author(s): Tom Ellis

☐ Initial submission ☐ Revised version ☒ Final submission

Flow Cytometry Reporting Summary

Form fields will expand as needed. Please do not leave fields blank.

Data presentation

For all flow cytometry data, confirm that:

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

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

☒ 3. All plots are contour plots with outliers or pseudocolor plots.

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

Methodological details

5. Describe the sample preparation.	Bacteria were grown in 50 mL of media in 500-mL baffled shake flasks at 37 degrees C for 24 hours. Every hour from 16 to 24 hr, 350 µl of each culture was diluted in 650 µl of PBS and stored on ice in a 96 well microtitre plate. . At the end of the experiment, the PBS-diluted aliquots stored on ice were run on a BD LSR Fortessa X20 flow cytometer to obtain single-cell data.
6. Identify the instrument used for data collection.	Experiments were performed on an BD LSR Fortessa X20 flow cytometer with 96-well plate autosampler attached.
7. Describe the software used to collect and analyze the flow cytometry data.	Data was collected using the BD LSR Fortessa X20 software and analyzed using FlowJo software.
8. Describe the abundance of the relevant cell populations within post-sort fractions.	Cells were gated for shape and size to ensure only bacteria events are captured.
9. Describe the gating strategy used.	Cells were gated during acquisition to ensure that only bacterial cells were acquired based on past experience using the BD LSR Fortessa X20 cytometer.

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

Nature Methods: doi:10.1038/nmeth.4635