Effects of genetic variation on the E. coli host-circuit interface

Cardinale, Stefano; Pawel Joachimiak, Marcin

Published in:
Cell Reports

Link to article, DOI:
10.1016/j.celrep.2013.06.023

Publication date:
2013

Document Version
Publisher's PDF, also known as Version of record

Cell Reports
Report

Effects of Genetic Variation on the *E. coli* Host-Circuit Interface

Stefano Cardinale,[1,2] Marcin Pawel Joachimiak,[2] and Adam Paul Arkin[1,2,*]
[1]Department of Bioengineering, University of California, 955 Energy Biosciences Building, 2151 Berkeley Way, Berkeley, CA 94704, USA
[2]Physical Biosciences Division, Lawrence Berkeley National Laboratory, 1 Cyclotron Road, Berkeley, CA 94720, USA
*Correspondence: aparkin@lbl.gov
http://dx.doi.org/10.1016/j.celrep.2013.06.023

SUMMARY

Predictable operation of engineered biological circuitry requires the knowledge of host factors that compete or interfere with designed function. Here, we perform a detailed analysis of the interaction between constitutive expression from a test circuit and cell-growth properties in a subset of genetic variants of the bacterium *Escherichia coli*. Differences in generic cellular parameters such as ribosome availability and growth rate are the main determinants (89%) of strain-specific differences of circuit performance in laboratory-adapted strains but are responsible for only 35% of expression variation across 88 mutants of *E. coli* BW25113. In the latter strains, we identify specific cell functions, such as nitrogen metabolism, that directly modulate circuit behavior. Finally, we expose aspects of carbon metabolism that act in a strain- and sequence-specific manner. This method of dissecting interactions between host factors and heterologous circuits enables the discovery of mechanisms of interference necessary for the development of design principles for predictable cellular engineering.

INTRODUCTION

The last decade has shown that predictable engineering of cell functions is hampered by ignorance of the host factors that affect and are affected by the engineered pathway and that lead to nonoptimal or undesirable behavior (Cardinale and Arkin, 2012; Lu et al., 2009; Purnick and Weiss, 2009). Toxic interactions or competition for general resources between implanted or modified systems and host components can represent a substantial barrier to predictable engineering of heterologous biological components in a host organism (Cardinale and Arkin, 2012). For example, a sizable fraction of host-circuit interactions revolve around competition for host ribosomes and RNA polymerases (Carrera et al., 2011), and bacterial ribosomes have been evolved in the laboratory to reduce the competitive functional interaction of translational factors (Wang et al., 2007). Operation of a synthetic circuit can also cause saturation of various endogenous functions, such as, for example, protein degradation, which only recently has been reported in detail (Cookson et al., 2011). The type and regulation of these activities can vary significantly among bacterial strains and ultimately disrupt the function of even simple genetic circuits (Wang et al., 2011).

To pursue genome-scale engineering efforts, we need to develop mechanistic understanding that enables prediction of the behavior of genetic circuits engineered in genetically variable hosts. Here, we hypothesize that most genetic variation that impacts circuit function is mediated through variations in global host resources and dynamics such as ribosomes and growth, which we call generic effects. Recent studies suggest that these generic effects might be reduced to few simple mathematical relations between the circuit, cell growth rate, and the cellular ribosomal fraction (Scott et al., 2010). However, these generic relations do not help discriminate cell functions that may act as effectors in a specific strain background or a target gene or protein sequence. Here, we investigate generic and specific effects by measuring growth-related variables and reporter expression in genetic variants of the bacterium *Escherichia coli* (*E. coli*). We analyze these variables first in standard *E. coli* laboratory strains and establish that these generic effects can explain most of expression variation. We then analyze in more detail specific cellular effectors by studying the host-to-circuit interaction in 88 single-gene knockout mutants and identify chemotaxis and nitrogen sensing as two direct modulators of circuit activity. Finally, we describe the strain-specific impact of carbon metabolism and lag phase of growth on circuit performance.

RESULTS

Designing and Modeling the Host-Circuit Interface

We developed a simplified model of the host-circuit interface, in which a change in the host genome is transmitted to the operation of a genetic circuit through either a generic or a specific modality. The generic mode affects all genes of the circuit, and here we analyze cellular growth properties as the main effectors (Klumpp et al., 2009) and specifically the rate of growth (λ), growth lag (κ), and cellular ribosomes (r). However, not all effectors are generic and a genetic modification can affect a specific circuit component or one of its properties (Figure 1Ai). We chose

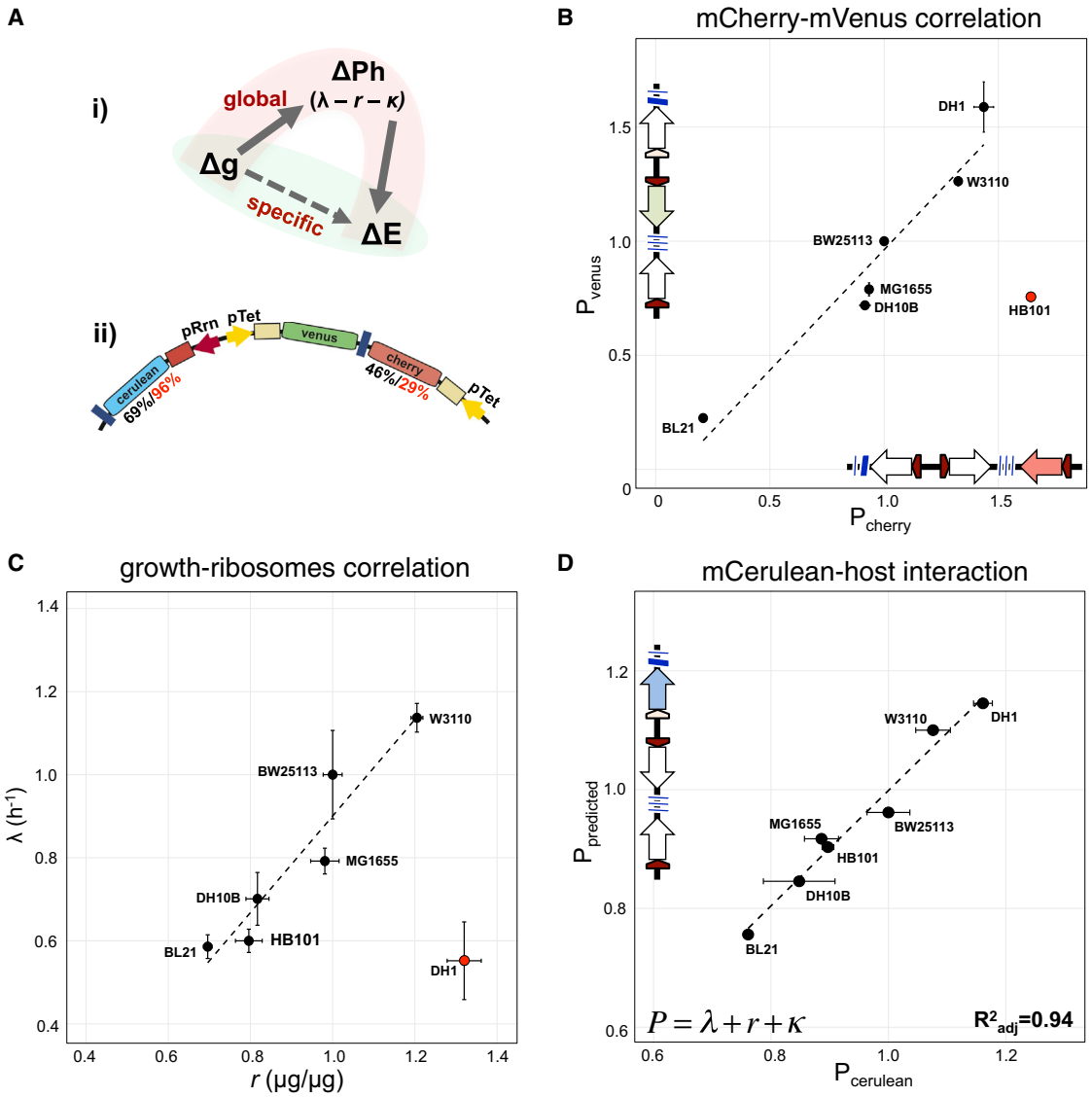

Figure 1. Designing and Modeling the Host-Circuit Interface

(A) (i) Model of the interaction between a host genetic variation (Δg) with heterologous gene expression (ΔE). Global effects induced by variation of cell physiology (ΔPh), mediated by growth variables (λ, r, and κ), impact expression of all heterologous genes (pink), whereas specific effects act on an individual circuit component or property (green). (ii) Schematic of the genetic probe. Promoters (arrows) and ribosomal binding sites (rectangles) of identical sequence have same color. The degree of coding (black) and protein (red) sequence identity of the reporter genes, relative to mVenus, is also indicated.

(B) The expression of reporters mVenus and mCherry (P_{venus} and P_{cherry}) is correlated with the exception of *E. coli* strain HB101 (ρ = 0.97).

(C) Growth is linearly related to ribosomal capacity, with strain DH1 presenting an outlying ribosome level (ρ = 0.95 without DH1).

(D) mCerulean expression predicted from a linear regression with variables λ, *r*, and a measure of lag-phase κ (R^2_{adj} = 0.94, Extended Experimental Procedures; Equation S4). (All values relative to BW25113.)

n = 4–6. See also Figures S1, S2, S3, S4, S5, and S6.

a plasmid-borne circuit composed of three constitutive promoters each driving a different fluorescent protein to probe the interaction, or interface, between circuit and host functions in diverse genetic backgrounds (Figure 1Aii). The probe is a modification of the "Three-Color Scaffold," whose design prevents transcriptional readthrough and encodes three fluorescence reporters, mVenus, mCherry, and mCerulean, with similar chromophore maturation times, established both in a recent study (Cox et al., 2010) and here by fitting reporter fluorescence time series (see below, Tables S1, S2, and S3, and Extended Experimental Procedures section "Determination of Reporter Production"). Reporters mVenus and mCherry, derived from different natural fluorescent proteins (46% nucleic acid and 29% amino acid identity), are expressed from copies of the same constitutive (in BW25113 background) $P_{LTet\text{-}O1}$ promoter and 5′ UTR so that differences in their mean strain-dependent expression can

be localized to translation efficiency, protein folding, solubility, and degradation of the encoded protein. The third reporter, mCerulean, which like mVenus is derived from GFP (69% nucleic acid and 96% amino acid identity) (Extended Experimental Procedures), was expressed from the ribosomal RNA promoter *rrnB*_P2 and a unique 5′ UTR (Figure 1Aii). Any differences between this reporter and its homolog then would largely be expected to arise due to specific differences in interaction of its transcriptional control, and possibly the particular UTRs, with changes in the host. Generic effects of strain background, then, should change the expression of all three reporters, whereas specific mechanisms affect only a subset. This probe design aids in locating the source of any specificity.

To model the interaction between global host resources and circuit function, we measured growth rate (λ), lag-phase duration (κ), and ribosomal content (r) of different strains upon transformation with the three-color construct above. Growth dynamics are measured by fitting optical density (OD) time series with a model (Figure S1; Extended Experimental Procedures; Equation S1) that identifies lag times and exponential growth rates of bacterial cultures. To measure circuit performance, we derive protein production rate (P) by fitting a model of protein expression to fluorescence time series. In particular, a system of two ordinary differential equations describes the change over time of unfolded (UF) and folded (F) protein. UF is produced, at midexponential phase, at a constant rate P: $dUF(t) = P - (\lambda(t) + ft) * UF(t)$, and the rate of conversion of UF to mature fluorescent protein F depends on the rate of maturation ft: $dF(t) = ft * UF(t) - \lambda(t) * F(t)$ (Figure S2 detailed description in Extended Experimental Procedures).

Cellular ribosomal content r was calculated at midexponential phase of growth from total RNA and protein amount (Scott et al., 2010) (Extended Experimental Procedures). The degree to which generic cellular properties are explanatory of variation in each marker expression can be inferred by regression of P against these variables. The unexplained variation can then be traced to specific interactions with circuit components or experimental measurement error.

Host-Circuit Interaction in Genetic Variants of *E. coli*

We first tested six common K12-derivative strains *W3110*, *MG1655*, *HB101*, *DH10B*, *DH*1, *BW25113*, and the B-type strain *BL21 (DE3)*. These strains are evolutionary divergent to different degrees. Among K strains, *W3110* and *MG1655*, are the closest to the ancestral isolate and ultimately they differ in just eight positions (Hayashi et al., 2006), whereas the genomes of *E. coli* K-12 and B are more divergent, sharing ~92% of their genes with >99% nucleotide identity (Jeong et al., 2009). Reporter output levels (P) showed a 3-fold change with *BL21* having the lowest and *DH1* the highest output. Expression of mCherry (P_{cherry}) and mVenus (P_{venus}) markers, which have identical upstream sequence, was correlated across strains ($\rho_{ve\text{-}che}$ = 0.72), although we found much higher production of mCherry than mVenus in strain HB101 ($\rho_{ve\text{-}che}$ = 0.97 with HB101 removed, Figures 1B and S3A). Expression of mCerulean ($P_{cerulean}$) was also correlated with that of mVenus ($\rho_{ve\text{-}ce}$ = 0.93) and mCherry ($\rho_{che\text{-}ce}$ = 0.73 with HB101 and $\rho_{che\text{-}ce}$ = 0.91 without) (Figure S3B). Variables λ and r were positively related, although strain *DH1* had an anomalously high ribosomal content and a growth curve that differed substantially from other strains (ρ = 0.42, ρ = 0.95 without inclusion of DH1; Figures 1C and S4).

We fitted protein production rates P with a linear model of physiological variables λ, r, and κ to estimate the degree to which these cellular variables could explain strain-specific variation in expression. Excluding strain DH1, variables λ and r could individually explain, respectively, ~70% ($p < 0.05$) and ~80% ($p < 0.01$) of P_{venus} and $P_{cerulean}$, All three explanatory variables contributed significantly ($p < 0.05$) to P_{cherry} regression and together recapitulated ~79% of P_{cherry} or ~94% of $P_{cerulean}$ and P_{venus} variation across all strains (Figures 1D, S5A, and S5B). The studentized residuals of the regressions were randomly distributed across the predictions implying no systematic biases associated with the model (Figure S6).

We then asked whether the observed cell-circuit relationship is conserved in bacterial strains whose genomes differ only by the absence of single genes. We selected 88 strains from the *E. coli* KEIO collection of single-gene deletions broadly covering: (1) membrane transport (n = 21), (2) two-component signal transduction (n = 21), (3) metabolism (n = 26), (4) chemotaxis (n = 6), (5) transcription factors (n = 9), and (6) others (n = 5) (COG, (Tatusov et al., 2000). First, we verified that changes in growth and expression profiles were directly linked to the strain-specific genotype through a simple complementation assay, which demonstrated that both growth and probe output could be restored toward wild-type upon expression of the missing gene (Figure S7). On average, the test circuit decreased strain growth by ~10% (load; relative to promoterless control plasmid; Extended Experimental Procedures; Equation S3). The load could not explain more than ~10% of variation in growth or marker expression, which suggests that the circuit, encoded on a low-copy plasmid, did not impair cell fitness considerably (Figure S8A and S8B). P_{venus} and P_{cherry} were well correlated ($\rho_{ve\text{-}che}$ = 0.73) in all strains except Δ*fis*, which showed much higher P_{cherry} (Figure 2A) ($\rho_{ve\text{-}ce}$ = 0.72, $\rho_{ch\text{-}ce}$ = 0.62). This result suggests that reporter upstream sequences were not themselves selectively targeted across the host backgrounds tested. Several strains affected all three markers significantly (>2 SD): Δ*qseB*, Δ*cheY*, Δ*glpX*, Δ*glnL*, and Δ*narL*; however, strains Δ*gnd*, Δ*tktA*, Δ*cysA*, and Δ*pfkA*, in addition to Δ*fis*, showed higher expression of just mCherry (>2 SD, brown in Figures 2A and S9).

We tested how much of the variation in marker expression could be explained by growth-related measures λ, r and κ, and their interaction across the selected 88 knockout strains. For this large number of genetic host backgrounds, the model recapitulated approximately one-third of the expression of P_{cherry} (R^2_{adj} = 0.36, Figure 2B), P_{venus} (R^2_{adj} = 0.35, Figure 2C), and $P_{cerulean}$ (R^2_{adj} = 0.33, Figure 2D), with cross-validated R^2 between 0.20 and 0.41 (Extended Experimental Procedures; Table S4). Again, an analysis of the studentized residuals of the regressions did not show systematic biases (Figure S10). Significant differences in marker expression among the five strains above and wild-type (gold marks, Figure 2A) was largely explained by the strain-dependent physiological (generic) variables (Figures 2B–2D). Strains Δ*cysA*, Δ*tktA*, Δ*fruR*, Δ*gnd*, and Δ*pfkA*, whose fluorescence measures diverged from predicted (Cook's

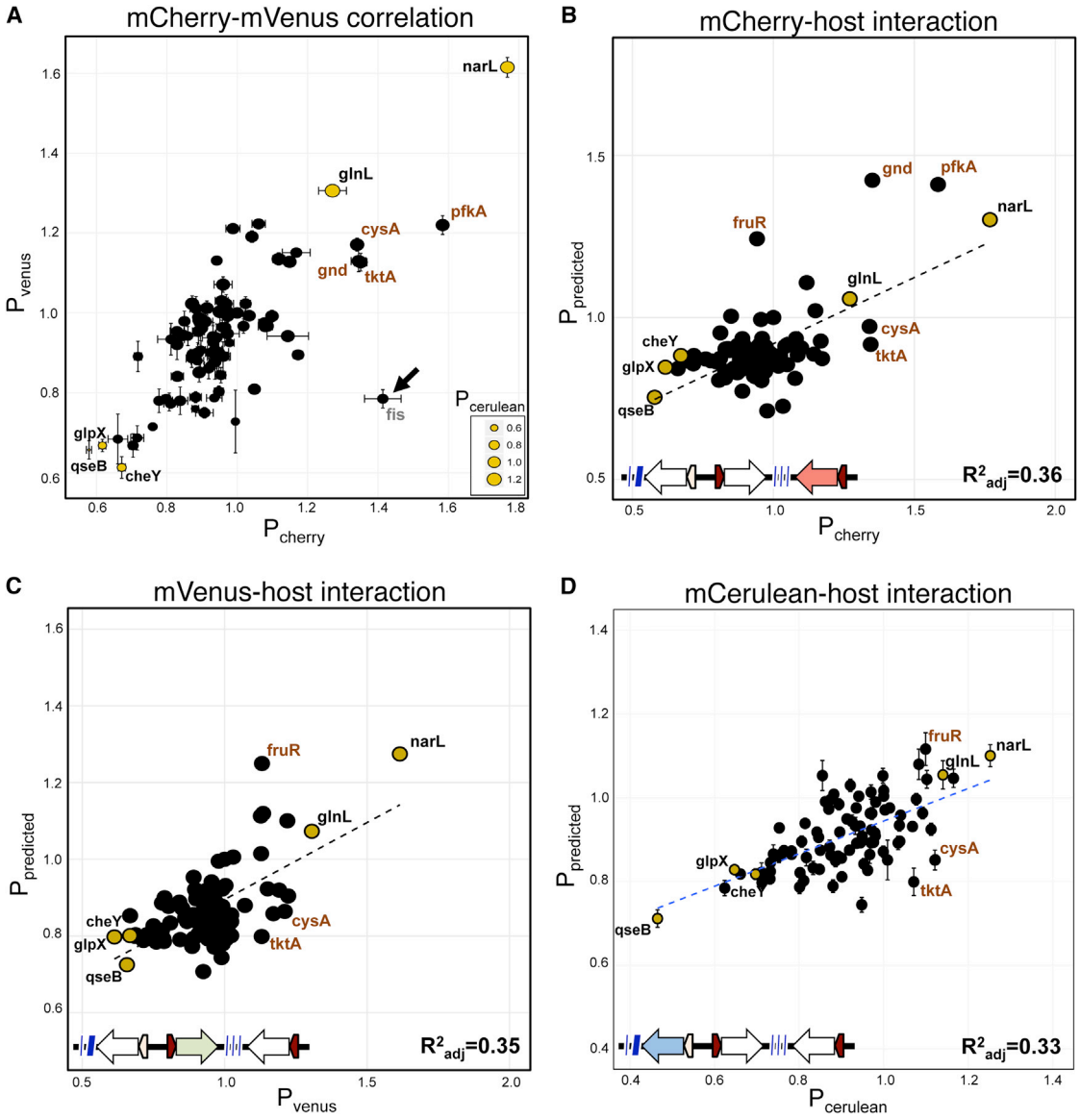

Figure 2. Host-Circuit Interaction in Single-Gene Knockouts

(A) P_{venus}, P_{cherry}, and $P_{cerulean}$ (marker size) are well correlated across KEIO knockout strains, with Δ*fis* as an obvious outlier (black arrow) (ρ = 0.72–0.86). In some strains, expression of all three markers (gold) or only cherry (brown text) is significantly affected (>2 SD).

(B–D) A linear regression model (Extended Experimental Procedures; Equation S5) of variables λ, *r*, and κ (see main text) recapitulates ~36% of variation in P_{cherry} (B), ~35% of P_{venus} (C), and ~0.33% of $P_{cerulean}$ (D). Text indicates outliers based on Cook's test. (P relative to wild-type.)

n = 3–4. See also Figures S7, S8, S9, and S10.

distance > 0.5 > 4/N, N = number of strains), influenced the estimation of P_{cherry} (brown, Figure 2B). Only strains Δ*cysA*, Δ*tktA*, and Δ*fruR*, but not Δ*gnd* and Δ*pfkA*, were also outliers in the P_{venus} and $P_{cerulean}$ regression (brown, Figures 2C and 2D). These results, and the higher expression of mCherry but not mVenus or mCerulean in strain HB101, suggest a type of interference that is specific to the gene coding or protein sequence.

Characterization of Host-Circuit Interference

The analysis of cellular expression and growth properties for genetic variants of *E. coli* revealed mutant strains Δ*qseB*, ΔcheY, ΔglpX, Δ*glnL*, and Δ*narL*, with all variables significantly shifted in the same direction from wild-type levels. However, in several strains the expression of the three markers was differentially affected and uncorrelated to physiological features (outliers in Figures 2B–2D). We hierarchically clustered all *Z* scores across all strains (Figure S11) and averaged the contribution of each variable within four clusters superimposed on identified expression outliers (Figure 3). Strains Δ*narL* and Δ*glnL* had a monotonic upshift of all variables and clustered together and with strain DH1 (cluster a, Figure 3), Strains Δ*glpX*, Δ*qseB*, and Δ*cheY*, as well as BL21, also formed a distinct cluster with a downshift of

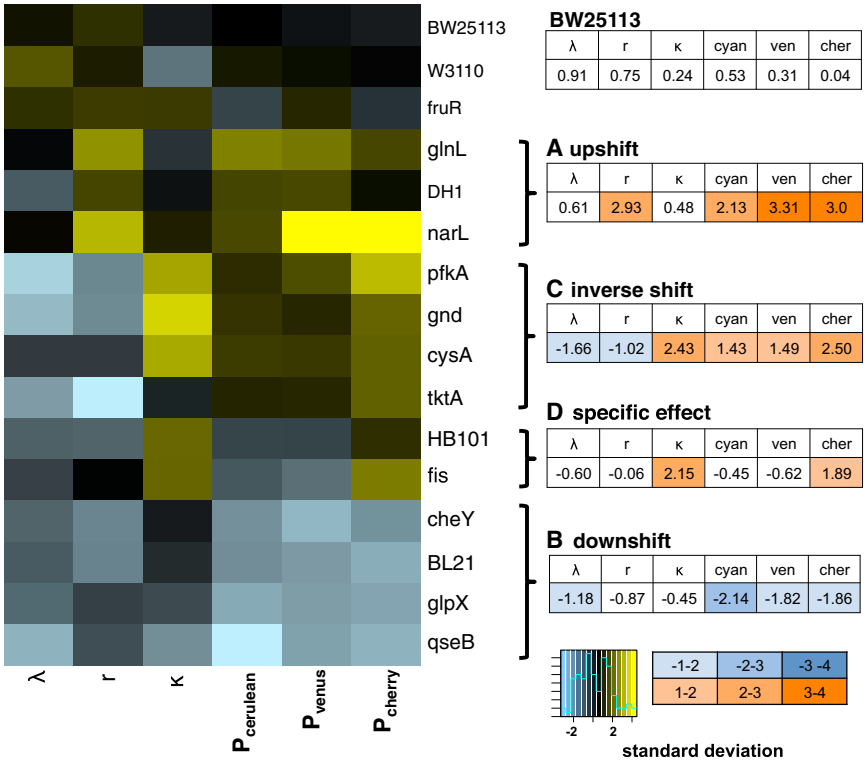

Figure 3. Characterization of Host-Circuit Interference

Heatmap of *Z* scores for *E. coli* knockout and wild-type expression outliers from regression and cluster analysis (Figures 2A–2D and S11). Apart from parental BW25113 and W3110, four clusters of strains are apparent: two groups with a homogenous increase (A) or decrease (B) of the measured properties, and two with an inverse variable pattern (C and D). Tables present averaged scores for each feature within each strain cluster. See also Figures S11 and S12.

generic and expression functionality (cluster b). Four strains Δ*cysA*, Δ*gnd*, Δ*tktA*, and Δ*pfkA* (cluster c) had cellular growth and ribosome levels inversely correlated to reporter expression and particularly of the marker mCherry. Last, in strains Δ*fis* and HB101 the expression of the three reporters was not correlated (cluster d). Outlier strain Δ*fruR* and the remaining parent *E. coli* strains clustered separately (red bracket, Figure 3). To classify the main effects at the host-circuit interface, we averaged feature *Z* scores within each group of strains and compared them with wild-type BW25113 (tables in Figure 3). In clusters a and b, growth and ribosome levels were generally correlated with the expression profile. In contrast, cluster c with three knockouts in central carbon metabolism out of four shows negative correlation with the expression profile and significantly longer lag phase. Finally, in strains HB101 and Δ*fis* (group d), which also have abnormally extended lag phase, individual reporter genes are not homogeneously affected and only mCherry fluorescence is significantly increased. This analysis allows us to identify biological examples of the two modes of interaction in our simple model of the host-circuit interface (Figure 1Ai). Over a total of 88 single-gene knockouts and six laboratory *E. coli* strains (BW25113 is the strain of reference), 60 (or ~64%) had no significant modification of any reporter expression (based on *Z* scores, Figure S11, > 1 SD), of which 16 with a growth variation. In the remaining 34 strains, we identified 11 genetic backgrounds in which the expression of all reporters was affected (global interaction) and correlated with a growth phenotype (Figure 4). In the other 23 strains the expression of a specific reporter gene (specific interaction) varied significantly either in combination (12/94 or ~13%) or not (11/94 or ~12%) with a growth variable (Figure 4).

DISCUSSION

Here, we used genetic variants of the bacterium *Escherichia coli* to characterize how the genetic context of the cell, and specific cellular functions, affect a heterologous expression system. We measured protein output rate P of three fluorescent markers and growth-associated variables (e.g., circuit and host profiles) in seven wild-type and 88 single-gene *E. coli* mutants and estimated how generic growth variables explained the up to 6-fold variation of P across the strains. Indeed, a linear model of growth, ribosome, and lag-phase variables recapitulated ~89% of P across standard laboratory strains and ~35%, or cross-validated R^2 ~0.30 on average (Extended Experimental Procedures; Table S4), across single-gene mutant strains, with ribosome content explaining the most variance. We can think of several reasons why growth is not as explanatory for *E. coli* gene mutants as for wild-type strains. One possibility is that these strains may have had time to mutationally adapt to growth conditions in the laboratory and thus optimize the linkage of host physiology to expression, whereas the knockout strains often display unusual growth curves indicative of poor adaptation to laboratory growth (e.g., *pfkA*, Figure S12). Host mutations that specifically affect plasmid metabolism, and thereby affect the expression of plasmid-borne genes, would likely leave the host cell unaffected. Also, there might be specific processes to which our particular choice of genes is more sensitive to, such as, for example, overuse of specific transfer RNAs or the binding of transcriptional cofactors, which would not generally affect *E. coli* behavior.

Among wild-type strains, W3110 had the most significant increase of all measured variables. Although the W3110 genome differs in only a few nucleotides from that of MG1655 (Hayashi et al., 2006), mutations in critical genes encoding the global regulators rpoS and crp, and the 23S ribosomal RNA rrlE, could be the source of the small but significant phenotypic differences observed in this study. Single-gene knockout strains Δ*glnL* and Δ*narL*, which encode regulators of nitrogen metabolism, also showed a similar increase, whereas Δ*qseB* and Δ*cheY*, which transmit and mediate the response to chemotactic and quorum

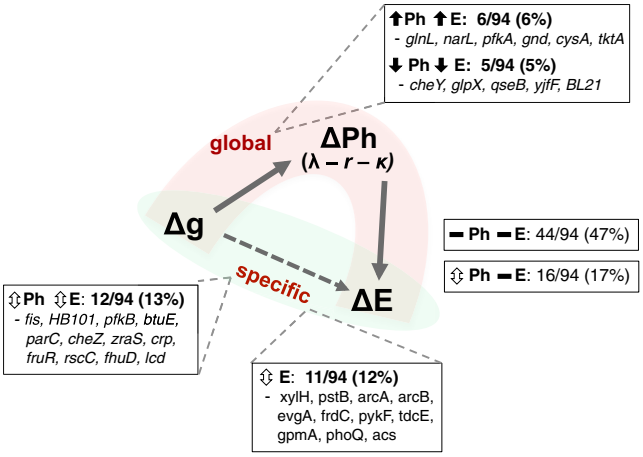

Figure 4. Classification of Host-Circuit Interactions

Among 94 different strains of *E. coli*, 34 genetic backgrounds (Δg) (~36%) influenced a genetic probe by affecting heterologous expression (ΔE) of all (pink) or individual (green) reporter genes. In particular, in 6/94 and 5/94 strains the expression of all reporter genes either increased or decreased and correlated with a growth phenotype (ΔPh) (respectively, full arrows). In 12 strains (~13%), a growth variable affected individual genes specifically and in the same direction (two-sided white arrows), whereas for 11 (~12%) of the strains one or two reporter genes varied without a correlated change in physiology. Finally, 16 strains (~17%) showed variation in growth that did not lead to significant changes in reporter expression (constructed from *Z* scores in Figure S11 with 1 SD cutoff).

signals, showed an opposite behavior. Single-gene variant analysis was important for a more precise understanding of sources of context effects in relation to wild-type *E. coli*. For instance, strains Δ*qseB* and Δ*cheY* clustered with BL21(DE3), for which a recent study revealed much less expression of chemotaxis proteins, such as cheY, compared to MG1655 (Deutschbauer et al., 2011; Price et al., 2013; Yoon et al., 2012).

We also found strains for which circuit and host variables were not correlated. In knockouts strains Δ*cysA*, Δ*gnd*, Δ*pfkA*, and Δ*tktA* (group c), growth and ribosome profiles were reduced but, in contrast, lag phase was longer and P was higher (Figure S12). These mutations likely reduce flux through both glycolysis and pentose phosphate pathways (Girgis et al., 2012; Hua et al., 2006; Ikeda et al., 2006; Shimizu, 2004). There is a complex interaction between central carbon metabolism and plasmid gene expression as plasmid maintenance can alter glucose uptake, leading to the accumulation of lactic, acetic, and other acids and to the extension of lag phase (Diaz Ricci and Hernández, 2000; Ow et al., 2006). Also, plasmid maintenance is associated with downregulation of glycolytic enzymes and reduced biosynthetic capacity (Ow et al., 2006), which indicates an adaptation to reduce metabolic end product accumulation. We suggest that the higher reporter levels we observe in strains of cluster c may result from relief of metabolic stress. Strains HB101 and Δ*fis* cluster separately (group d) and also have extended lag, but the effect is limited to increased mCherry fluorescence. The global regulator Fis is known to peak in lag phase and regulate many of its processes, and HB101 was found more susceptible to the plasmid-borne metabolic stress described above (Diaz Ricci and Hernández, 2000). We suggest that here the probe circuit is affected by a specific interaction between a cellular mechanism that remains to be elucidated and a protein property such as solubility or folding rate.

The utility of mapping the host-circuit interface is to identify targets that can be engineered to improve circuit performance. In this case, our circuit is exceptionally simple in that it expresses largely "inactive" proteins that are not expected to interfere with specific host functions. Therefore, our genetic probe detects cellular functions that generically affect heterologous expression such as aspects of central carbon metabolism and two-component signaling pathways (Figure 4). When more active proteins are to be expressed, more specific interactions with host processes may be uncovered. Once identified, specific metabolic bottlenecks might be relieved by overexpression of a key metabolic enzyme or cofactor regeneration system. Similarly, other interfering host systems, even those whose effects are hard to explain, also provide a handle to improve expression.

The increasing availability of genome-wide deletion libraries (Deutschbauer et al., 2011) and high-throughput recombineering and genome scale regulation (Cong et al., 2013; Qi et al., 2013; Warner et al., 2010) provides a mean of querying how each host factor affects key measures of heterologous circuit performance. Ultimately, such information and the conceptual framework we present in this study should provide design principles to control for these host context interactions.

EXPERIMENTAL PROCEDURES

Strains and Culture Media

Single-gene knockout strains were obtained from the KEIO collection (Baba et al., 2006) and wild-type laboratory *E. coli* from the Joint Bio-Energy Institute (JBEI). All strains were cultured in Neidhardt's MOPS-based Rich defined medium (TEKnova), supplemented with 0.5% glucose and antibiotics ampicillin or kanamycin (50–70 μg/ml) when required.

Measurement of Bacterial Growth and Fluorescence

For time series, overnight cultures were diluted 1:3,000 in 150 μl of Neidhardt's medium, and OD and fluorescence were acquired every 7.5 min with an Infinite M1000 microplate reader (Tecan). Single-cell fluorescence measures were obtained for cultures grown to midexponential phase, generally for 1 hr and 30 min at 37°C from a 1:80 dilution in warm medium, with a Partec flow cytometer (~80,000 cells).

Total RNA and Protein Isolation

Total RNA and proteins were extracted from bacteria grown to midexponential phase from 1:80 dilutions of overnight cultures as described above. RNA was isolated with the RNAeasy mini kit (QIAGEN) or the Aurum Total RNA 96 kit (Bio-Rad), and total cellular protein content was quantified on 200 μl of the same culture with a Total Protein Lowry Kit (Sigma-Aldrich), according to manufacturer protocols.

Regression and Statistical Analysis

Mathematica (Wolfram Research) was used to fit OD and fluorescence kinetic measurements with Equations S1 and S2 (Extended Experimental Procedures). Linear regression and other statistical validation were performed in R with packages "stats" and "car." Plots were drawn in R with packages "gplots" and "ggplot2."

Please refer to Extended Experimental Procedures for additional experimental procedures and analysis.

SUPPLEMENTAL INFORMATION

Supplemental Information includes Extended Experimental Procedures, 12 figures, and 6 tables and can be found with this article online at http://dx.doi.org/10.1016/j.celrep.2013.06.023.

ACKNOWLEDGMENTS

We thank M. Price and V. Mutalik for thoughtful comments. This work was funded by the National Science Foundation as part of the Synthetic Biology Engineering Research Center grant number 04570/0540879.

S.C. and A.P.A. conceived and designed the study; S.C. performed the experimental work and computational analysis and wrote the manuscript; M.P.J. contributed to the computational analysis and writing the manuscript; A.P.A. supervised the study and manuscript preparation.

Received: March 1, 2013
Revised: May 16, 2013
Accepted: June 18, 2013
Published: July 18, 2013

REFERENCES

Baba, T., Ara, T., Hasegawa, M., Takai, Y., Okumura, Y., Baba, M., Datsenko, K.A., Tomita, M., Wanner, B.L., and Mori, H. (2006). Construction of Escherichia coli K-12 in-frame, single-gene knockout mutants: the Keio collection. Mol. Syst. Biol. *2*, 2006.0008.

Cardinale, S., and Arkin, A.P. (2012). Contextualizing context for synthetic biology—identifying causes of failure of synthetic biological systems. Biotechnol. J. *7*, 856–866.

Carrera, J., Rodrigo, G., Singh, V., Kirov, B., and Jaramillo, A. (2011). 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.

Cong, L., Ran, F.A., Cox, D., Lin, S., Barretto, R., Habib, N., Hsu, P.D., Wu, X., Jiang, W., Marraffini, L.A., and Zhang, F. (2013). Multiplex genome engineering using CRISPR/Cas systems. Science *339*, 819–823.

Cookson, N.A., Mather, W.H., Danino, T., Mondragón-Palomino, O., Williams, R.J., Tsimring, L.S., and Hasty, J. (2011). Queueing up for enzymatic processing: correlated signaling through coupled degradation. Mol. Syst. Biol. *7*, 561.

Cox, R.S., 3rd, Dunlop, M.J., and Elowitz, M.B. (2010). A synthetic three-color scaffold for monitoring genetic regulation and noise. J. Biol. Eng. *4*, 10.

Deutschbauer, A., Price, M.N., Wetmore, K.M., Shao, W., Baumohl, J.K., Xu, Z., Nguyen, M., Tamse, R., Davis, R.W., and Arkin, A.P. (2011). Evidence-based annotation of gene function in Shewanella oneidensis MR-1 using genome-wide fitness profiling across 121 conditions. PLoS Genet. *7*, e1002385.

Diaz Ricci, J.C., and Hernández, M.E. (2000). Plasmid effects on Escherichia coli metabolism. Crit. Rev. Biotechnol. *20*, 79–108.

Girgis, H.S., Harris, K., and Tavazoie, S. (2012). Large mutational target size for rapid emergence of bacterial persistence. Proc. Natl. Acad. Sci. USA *109*, 12740–12745.

Hayashi, K., Morooka, N., Yamamoto, Y., Fujita, K., Isono, K., Choi, S., Ohtsubo, E., Baba, T., Wanner, B.L., Mori, H., et al. (2006). Highly accurate genome sequences of Escherichia coli K-12 strains MG1655 and W3110. Mol. Syst. Biol. *2*, 2006.0007.

Hua, Q., Joyce, A.R., Fong, S.S., and Palsson, B.O. (2006). Metabolic analysis of adaptive evolution for in silico-designed lactate-producing strains. Biotechnol. Bioeng. *95*, 992–1002.

Ikeda, M., Ohnishi, J., Hayashi, M., and Mitsuhashi, S. (2006). A genome-based approach to create a minimally mutated Corynebacterium glutamicum strain for efficient L-lysine production. J. Ind. Microbiol. Biotechnol. *33*, 610–615.

Jeong, H., Barbe, V., Lee, C.H., Vallenet, D., Yu, D.S., Choi, S.H., Couloux, A., Lee, S.W., Yoon, S.H., Cattolico, L., et al. (2009). Genome sequences of Escherichia coli B strains REL606 and BL21(DE3). J. Mol. Biol. *394*, 644–652.

Klumpp, S., Zhang, Z., and Hwa, T. (2009). Growth rate-dependent global effects on gene expression in bacteria. Cell *139*, 1366–1375.

Lu, T.K., Khalil, A.S., and Collins, J.J. (2009). Next-generation synthetic gene networks. Nat. Biotechnol. *27*, 1139–1150.

Ow, D.S.W., Nissom, P.M., Philp, R., Oh, S.K.W., and Yap, M.G.S. (2006). Global transcriptional analysis of metabolic burden due to plasmid maintenance in Escherichia coli DH5 alpha during batch fermentation. Enzyme Microb. Technol. *39*, 391–398.

Price, M.N., Deutschbauer, A.M., Skerker, J.M., Wetmore, K.M., Ruths, T., Mar, J.S., Kuehl, J.V., Shao, W., and Arkin, A.P. (2013). Indirect and suboptimal control of gene expression is widespread in bacteria. Mol. Syst. Biol. *9*, 660.

Purnick, P.E., and Weiss, R. (2009). The second wave of synthetic biology: from modules to systems. Nat. Rev. Mol. Cell Biol. *10*, 410–422.

Qi, L.S., Larson, M.H., Gilbert, L.A., Doudna, J.A., Weissman, J.S., Arkin, A.P., and Lim, W.A. (2013). Repurposing CRISPR as an RNA-guided platform for sequence-specific control of gene expression. Cell *152*, 1173–1183.

Scott, M., Gunderson, C.W., Mateescu, E.M., Zhang, Z., and Hwa, T. (2010). Interdependence of cell growth and gene expression: origins and consequences. Science *330*, 1099–1102.

Shimizu, K. (2004). Metabolic flux analysis based on 13C-labeling experiments and integration of the information with gene and protein expression patterns. Adv. Biochem. Eng. Biotechnol. *91*, 1–49.

Tatusov, R.L., Galperin, M.Y., Natale, D.A., and Koonin, E.V. (2000). The COG database: a tool for genome-scale analysis of protein functions and evolution. Nucleic Acids Res. *28*, 33–36.

Wang, K., Neumann, H., Peak-Chew, S.Y., and Chin, J.W. (2007). Evolved orthogonal ribosomes enhance the efficiency of synthetic genetic code expansion. Nat. Biotechnol. *25*, 770–777.

Wang, B., Kitney, R.I., Joly, N., and Buck, M. (2011). Engineering modular and orthogonal genetic logic gates for robust digital-like synthetic biology. Nat. Commun. *2*, 508.

Warner, J.R., Reeder, P.J., Karimpour-Fard, A., Woodruff, L.B., and Gill, R.T. (2010). Rapid profiling of a microbial genome using mixtures of barcoded oligonucleotides. Nat. Biotechnol. *28*, 856–862.

Yoon, S.H., Han, M.J., Jeong, H., Lee, C.H., Xia, X.X., Lee, D.H., Shim, J.H., Lee, S.Y., Oh, T.K., and Kim, J.F. (2012). Comparative multi-omics systems analysis of Escherichia coli strains B and K-12. Genome Biol. *13*, R37.