Programming gene expression with combinatorial promoters

Supplementary results

-10 and -35 polymerase box strength

The 288 promoters exhibited five decades of variation in unregulated promoter activity (Supplementary Figure S1). These sequences contained twelve -35 boxes which differed from the consensus TTGACA at up to three positions, and six -10 boxes which differed from the consensus TATAAT at up to two positions. The distributions of unregulated promoter activity for the -35 and -10 boxes were highly variable and overlapping (Supplementary Figure S2). We found that three of the twelve -35 boxes (TTGACA, TTGACT, and TAGACA) and five of the six -10 boxes (TATAAT, TAGATT, TAGAGT, GATACT, and GATAAT) produced sets of relatively strong promoters (~90% of the distributions were higher than 10^3 ALU). All of the strongest promoters in the library (~10^6 ALU) contained two of these 'strong boxes.' We used the median promoter activity of the -35 and -10 box distributions to predict the unregulated promoter activity of each promoter (Supplementary Methods). The predicted promoter activities were weakly correlated (Pearson coefficient = 0.19, Kendall τ = 0.32) with the measured promoter activities, and exhibited the best agreement for the strongest promoters (Fig. S2C). Thus, strong promoters contained strong polymerase boxes; but the presence of strong polymerase boxes did not guarantee high promoter activity.

Activator operators at core and proximal

We examined the effect of activator operators at *core* and *proximal* on maximum promoter activity (fully induced). Supplementary Figure S3 shows cumulative histograms of activity for four classes of promoters: no activator operator, an activator operator at *proximal*, an activator operator at *core*, and an activator operator at *distal*. For LuxR (Fig. S3A) the presence of an operator had no effect on median promoter activity. For AraC (Fig. S3B) we found two notable effects. First, the distribution of maximal promoter activities was higher when AraC acted at *distal*. This revealed that activation increased promoter activity on average, and that the maximal expression in the presence of the activator was uniform (near the 10^5 ALU activation ceiling). This narrow distribution of activated promoter levels is consistent with the LuxR distribution, though many fewer LuxR activated promoters were measured. Second, we found that promoters with an AraC operator only at *proximal* exhibited lower average promoter activity. Half of these promoters had a maximum activity of less than 200 ALU, and all of them exhibited activity less than 10^5 ALU. Conversely, the median strength of promoters without an AraC operator (or with an AraC operator at *core* only) was 20,000 ALU, and their maximal activity was 10^6 ALU. We note that the natural repressor activity of AraC is mediated by looping, not by steric exclusion (Hamilton and Lee, 1988), so this unexpected result is still consistent with previous work. From this analysis we infer that AraC can enact mild (10-100×) arabinose-independent repression at the *proximal* region only, and neither AraC nor LuxR can be transformed into a strong (≥10×) inducible repressor simply by moving its operator.

Spurious regulation by TetR

We found 7 promoters whose activity was induced 2-3× by aTc, without the presence of an operator for TetR. Units containing a λ cI operator (Supporting Methods) have up to 10 out of 14 conserved positions of the TetR consensus operator. Every one of the 7 spurious TetR regulated promoters contained at least one such cryptic site. These results suggest that TetR may repress weakly (3×) by binding to λ cI operators.

Dual-repressor interaction in RR promoters

We used the model of RR promoters (Methods) to analyze the relationship between logical phenotype and the repressor interaction parameter ω. Fixing r, we plotted lines of equal ω, varying a (Supplementary Figure S4). The logic parameter l did not depend strongly on r, though an increase in r was found to increase l at the extremes (near l~0 and l~1; *e.g.*, compare different marker sizes in Fig. S4). We found that the logic parameter l did not depend strongly on a when $a < 0.25$. This means that logic and symmetry are 'decoupled' for near symmetric responses. As a result, the logic parameter l depends only on ω. Asym-OR logic was possible only when r was relatively low ($r \leq 10^3$) and ω was high ($\omega \geq 100$), in agreement with the analytical results (Methods). Conversely, perfect AND logic required r to be high ($r \approx 10^5$) and ω to be low ($\omega \approx 0$).

Logic robustness to inducer concentrations

We examined the logical phenotypes of promoters with intermediate inducer concentrations. We chose three RR promoters from Fig. 5A, and measured their response to 16 combinations of inducer concentrations (Supplementary Methods). These three

promoters exhibited diverse logic: AND (clone A3), asym-AND (clone D8), and SLOPE (clone D9). We found that all three promoters increased their activity monotonically with increasing concentrations of each inducer, both singly and in combination.

For 16 different combinations of inducer inputs, we calculated the logic parameters (*r, a, l*) corresponding to the fully induced and 8 partially induced states (Supplementary Figure S5). As expected, the parameters *r* and *a* were highly sensitive to inducer concentrations. The range *r* of each promoter decreased when either of the inducer concentrations was lowered. Lowering the concentration of only one inducer significantly below its threshold predictably resulted in asymmetric behavior ($a \sim 1$). Conversely, lowering the concentration of a dominant inducer could make the response more symmetric ($a \sim 0$).

The logic parameter *l* was less dependent on inducer concentrations, and varied differentially for the three promoters. Partial induction reduced *l* for the AND and asym-AND gates. The AND gate A3, with the largest *l*, had the highest variation in *l* ($l = 0.46$ to $l = 0.86$); while the SLOPE gate D9, with the smallest *l*, exhibited the least variation ($l = 0.48$ to $l = 0.50$). These results show that the SLOPE gate logic parameter *l* is extremely robust to different input concentrations, while the AND-like gates are more sensitive.

Supplementary methods

Library fragments with λ cI operators

Each unit sequence was designed either from a consensus sequence (strong) or a sequence known to be responsive to one of five transcription factors (AraC, λ cI, LacI, TetR, LuxR), with variations in consensus signal strength, transcription factor binding site strength, spacing, and orientation (Table S1). We did not assay the response to λ cI (labeled con1-con4 for each unit, Supplementary Data 1, Supplementary Table S1, and Table I), although 68% of the sequenced promoters contained at least one λ cI operator.

Library construction and handling

The crude randomized assembly ligation mix (Methods) was diluted 20× and combined with the bacterial luciferase reporter plasmid pCS26 (Bjarnason et al, 2003). This vector was cut with XhoI and BamHI, to match the 5' terminal overhangs on the *distal* and *proximal* ends. The vector-insert mixture was again ligated for one week, and transformed by electroporation (2.48kV, 0.2cm gap, 200uF) into Electromax DH10B cells (Invitrogen). A fraction of the recovered transformation mix was plated onto selective plates, grown overnight, and counted. These colony counts provided an estimate of 22,000 independent assembly events.

The remaining transformants were directly inoculated into LB containing antibiotics and grown for 8 hours at 37°C. Harvested cells were used to prepare liquid libraries of Midi prep DNA (Qiagen) which were re-transformed into *E. coli* K12 strain MG1655 (Blattner

et al, 1997; Riley et al, 2006) containing the native *ara* operon, the LacI- and TetR-overexpressing Z1 cassette (Lutz and Bujard, 1997), and the medium-copy plasmid pCD136 which constitutively expresses LuxR).

Approximately 10,000 transformants were plated on selective media and picked into 35 384-well plates with a colony-picking robot (Norgren Systems). Each clone of the first 384-well plate was re-streaked on selective media and inoculated from a single colony into 96-well plates. 288 clones were selected randomly for commercial sequencing (Laragen Inc., Los Angeles, CA), amplified with primers pZE05 (CCAGCTGGCAATTCCGA) and pZE06 (AATCATCACTTTCGGGAA) using the Accuprime PCR System (Invitrogen), and sequenced from the purified PCR products with primer pZE05. Sequence traces were analyzed by hand for quality (4Peaks by A. Griekspoor and Tom Groothuis, mekentosj.com).

Library measurements

Each set of 96 clones was assayed in LB Lennox media made from a single 1200mL batch. Cells were grown in 96-well plates to saturation (16-22 hours at 37°C) and inoculated into 3 replicate plates of each of 16 inducer conditions using a steel 96-pin replicator (V & P Scientific). The library was assayed in these 16 inducer conditions corresponding to all combinations of the four inducible factors: VAI (1μM), IPTG (500μM), L(+)-arabinose (0.1%), and aTc (100ng/mL). Plates were prepared by filling 96-well plates with 150μL of media and inducers on a Genetix QFill2 plate-filler (5% precision), triple-washing the apparatus to prevent inducer-carryover.

These concentrations of inducers did not significantly inhibit cell growth in the conditions used (not shown). The 48 plates were grown at 25°C without shaking for 18 hours in the dark. This growth condition minimized evaporation and sample handling time, while providing nearly uniform culture optical densities (not shown). Luciferase activity was assayed by luminescence counts using a Tecan Safire plate reader (default settings, 100ms integration time) after 30s at 30°C. Three reads of each clone were taken to assure temperature equilibration. To insure stringent control, all 16 conditions were read for one replicate before starting the next replicate.

Polymerase box strength prediction

For each -10 and -35 box in the library, we calculated the distributions of unregulated promoter activity (Fig. S2AB). We took the median of each distribution to represent the -10 and -35 box 'strength.' For each of the 288 promoters, we calculated a predicted promoter activity as the geometric mean of its -35 and -10 box strengths and plotted each against the measured unregulated promoter activity (Fig. S2C). Alternative functions of the two box strengths (arithmetic mean, product, etc.) produced similar results.

Partial induction experiment

We measured three RR promoters (A3, D8, and D9) in sixteen inducer conditions. Each clone was grown in selective media to saturation at 37°C, and then diluted 60,000× and inoculated into a 96-well plate. Each well contained 150μL of selective media at 100, 50, 25, or 0 ng/mL aTc and 500, 50, 5, or 0 μM IPTG. We did not explore higher inducer

concentrations, to avoid growth effects. This plate was grown at 25°C for 18 hours without shaking. Luminescence was measured as described above. The minimally induced case (5 μM IPTG and 25 ng/mL aTc) often produced outlying behavior, and was discarded from the phenotype-parameter analysis.

Supplementary figure captions

Figure S1. The 288 characterized clones exhibit diverse regulatory ranges (*r*). The characterized promoters exhibited regulation up to $r = 10^5$. Approximately half of the library promoters are regulated at least 10×.

Figure S2. Many factors contribute to promoter strength. (A) Histograms of unregulated promoter activity for each -35 box reveal large variations in promoter strength. Three strong -35 boxes: TAGACA, TTGACT, and TTGACA (consensus) exhibit higher activities than the other nine. (B) Histograms of unregulated promoter activity for each -10 box reveal highly variable, overlapping distributions for five -10 boxes. The sixth -10 box (TATTTT) requires an activator to achieve high expression. (C) The median strength of each -35 box and -10 box distribution is used to predict the strength of each promoter. For each promoter, the geometric mean of the -10 and -35 box strengths are plotted against the unregulated activity.

Figure S3. Activators have small effects at *core* and *proximal*. The cumulative histograms of maximal promoter activity for LuxR (A) and AraC (B). The maximal activity of promoters with activator operators at the *distal* position (where activation is effective) are shown for comparison.

Figure S4. Operator interactions determine logic in RR promoters. Parametric plots of the logic parameter *l* as a function of the asymmetry *a* and repressor interaction ω. Each point is colored corresponding to ω, from ω = 100 to ω = 0, as shown on the color

bar. For each value of ω, we numerically computed the logic *l* as a function of *a* for both $r = 10^3$ (smaller circles) and $r = 10^5$ (larger circles).

Figure S5. RR promoters respond differentially to partial induction. For each promoter, we measured the response in 16 different inducer conditions (Supplementary Methods). The radius of the circles is proportional to the logarithm of the regulatory range *r*, as in Figure 4B. The minimally induced case (5 μM IPTG and 25 ng/mL aTc) often produced outlying behavior (dashed circles). (A) The logic phenotype space coordinates of 3 RR promoters with respect to fully saturated inducer conditions. (B) The AND gate A3 exhibited differential logic when the inducer concentrations were changed. (C) The asym-AND gate D8 varied in both range *r* asymmetry *a,* and to a lesser extent, the logic parameter *l*. (D) The SLOPE gate D9 varied only in the range *r* and asymmetry *a*, while the logic parameter remained approximately constant ($l = 0.5$).

Supplementary references

Basu S, Mehreja R, Thiberge S, Chen MT, Weiss R (2004) Spatiotemporal control of gene expression with pulse-generating networks. *Proc Natl Acad Sci U S A* 101: 6355-6360.

Becker NA, Kahn JD, Maher LJ, 3rd (2005) Bacterial repression loops require enhanced DNA flexibility. *Journal of molecular biology* 349: 716-730.

Bjarnason J, Southward CM, Surette MG (2003) Genomic profiling of iron-responsive genes in *Salmonella enterica serovar typhimurium* by high-throughput screening of a random promoter library. *J Bacteriol* 185: 4973-4982.

Blattner FR, Plunkett G, 3rd, Bloch CA, Perna NT, Burland V, Riley M, Collado-Vides J, Glasner JD, Rode CK, Mayhew GF, Gregor J, Davis NW, Kirkpatrick HA, Goeden MA, Rose DJ, Mau B, Shao Y (1997) The complete genome sequence of *Escherichia coli* K-12. *Science* 277: 1453-1474.

Crooks GE, Hon G, Chandonia JM, Brenner SE (2004) WebLogo: a sequence logo generator. *Genome research* 14: 1188-1190.

Egland KA, Greenberg EP (1999) Quorum sensing in *Vibrio fischeri*: elements of the luxl promoter. *Mol Microbiol* 31: 1197-1204.

Egland KA, Greenberg EP (2000) Conversion of the *Vibrio fischeri* transcriptional activator, LuxR, to a repressor. *J Bacteriol* 182: 805-811.

Hamilton EP, Lee N (1988) Three binding sites for AraC protein are required for autoregulation of araC in *Escherichia coli. Proc Natl Acad Sci U S A* 85: 1749-1753.

Hochschild A, Ptashne M (1986) Cooperative binding of lambda repressors to sites separated by integral turns of the DNA helix. *Cell* 44: 681-687.

Lanzer M, Bujard H (1988) Promoters largely determine the efficiency of repressor action. *Proc Natl Acad Sci U S A* 85: 8973-8977.

Lutz R, Bujard H (1997) Independent and tight regulation of transcriptional units in *Escherichia coli* via the LacR/O, the TetR/O and AraC/I1-I2 regulatory elements. *Nucleic acids research* 25: 1203-1210.

Michalowski CB, Short MD, Little JW (2004) Sequence tolerance of the phage lambda PRM promoter: implications for evolution of gene regulatory circuitry. *J Bacteriol* 186: 7988-7999.

Ptashne M (2004) *A genetic switch : phage lambda revisited*, 3rd edn. Cold Spring Harbor, N.Y.: Cold Spring Harbor Laboratory Press.

Riley M, Abe T, Arnaud MB, Berlyn MK, Blattner FR, Chaudhuri RR, Glasner JD, Horiuchi T, Keseler IM, Kosuge T, Mori H, Perna NT, Plunkett G, 3rd, Rudd KE, Serres MH, Thomas GH, Thomson NR, Wishart D, Wanner BL (2006) *Escherichia coli* K-12: a cooperatively developed annotation snapshot--2005. *Nucleic acids research* 34: 1-9.

Shadel GS, Baldwin TO (1992) Identification of a distantly located regulatory element in the luxD gene required for negative autoregulation of the *Vibrio fischeri* luxR gene. *The Journal of biological chemistry* 267: 7690-7695.

Sizemore C, Wissmann A, Gulland U, Hillen W (1990) Quantitative analysis of Tn10 Tet repressor binding to a complete set of tet operator mutants. *Nucleic acids research* 18: 2875-2880.

Zhang X, Reeder T, Schleif R (1996) Transcription activation parameters at ara pBAD. *Journal of molecular biology* 258: 14-24.