Supplementary Information for

Genomic Mining of Prokaryotic Repressors for Orthogonal Logic Gates

Brynne C. Stanton, Alec A.K. Nielsen, Alvin Tamsir, Kevin Clancy, Todd C. Peterson, and Christopher A. Voigt

Contents

Supplementary Results

The following additional files are uploaded as Supplementary Data Sets:

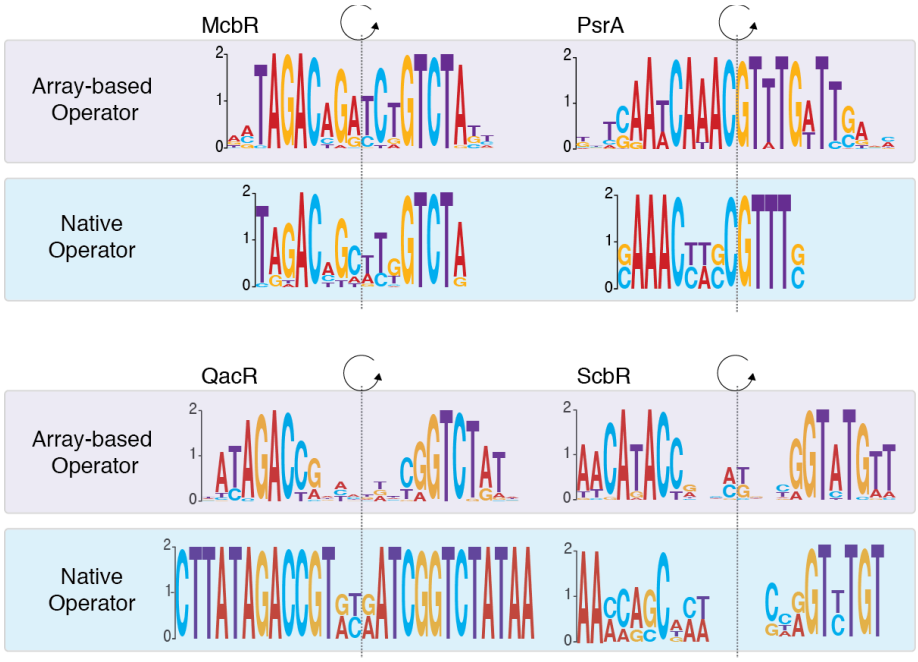
Supplementary Results

Supplementary Figure 1: Native operator consensus sequence. The previously reported operator sequences for the repressor library were obtained from the literature, and used to identify a consensus sequence generated via MEME-motif analysis. This information was used to fix positions within the probe sequence of the custom-designed microarray.

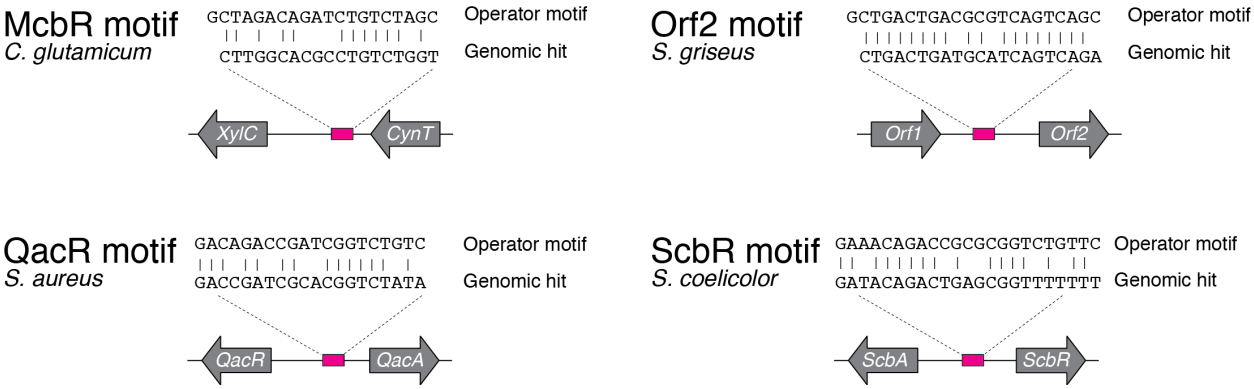
Supplementary Table 1: Native operator sequence alignment

Repressor	Operator Alignment
AprA	CGACATACGGGACGCCCCGTTTAT
FarA	GATACGAACGGGACGGACGGTTTGCAGC
SmeT	ATATACATACATGCTTGTTTGTTTGTAAAC
BarA	AGATACATACCAACCGGTTCTTTTGA
TylP	TTACAAACCGCTGACGCGGTTTGTAT
QacR	CTTATAGACCGATCGCACGGTCTATA
TtgR	GGAATATACTTACATTCATGGTTGTTTGTAAATACTGCTG
MtrR	ATACATACACGATTGCACGGATAAAAA
ScbR	TAAGATACAGACTAGAGCGGTTTTTTTTC
AcrR	TACATACATTTATGAATGTATGTA
MphR	CCTAAATGTAACAGTCACGTCGGTTATATTC
EthR	CACGCTATCAACGTAATGTCGAGGCCGTCAACGAGATGTCGACACTATCG
PhlF	ATGATACGAAACGTACCGTATCGTTAAGGT

Nature Chemical Biology: doi:10.1038/nchembio.1411



Supplementary Figure 2: Array-based and native operator sequence comparison. The array-based (top panel) and native operator (bottom panel) sequences are compared for the McbR[1], PsrA[2], QacR[3], and ScbR[4] repressors. Gray, dashed lines indicate the axis of symmetry. To properly align the ScbR operator, a 4 bp spacer was inserted between the half sites of the native operator sequence.



Supplementary Figure 3: Genomic analysis of array-based operator motifs. The region upstream of each repressor coding sequence (and where available the entire genomic sequence) for which an array-based motif was identified was examined for the sequences similar to the array-based operator motif. Similar motifs were found for the McbR, Orf2, QacR, and ScbR operators. The location of the identified operator is indicated by the pink box, and surrounding arrows signify genes that are labeled accordingly. Alignments correspond to the array-based, operator motif (top) and the genomic hit (bottom) exhibiting similarity. In the case of ScbR and QacR, genome analysis correctly identified operator sequences previously shown to be bound by these repressors *in vivo*[3,4], while the operator identified for McbR did not. The Orf2 repressor has not been previously characterized, and the sites occupied by Orf2 are unknown.

Nature Chemical Biology: doi:10.1038/nchembio.1411

Supplementary Table 2: Native operator sequences

Repressor*	Operator Sequence
AmtR[5]	TTTCTATCGATCTATAGATAAT
BetI[6]	ATTGATTGGACGTTCAATATAA
BM3R1[7]	CGGAATGAACGTTCATTCCG
HapR[8]	TTATTGATTTTTAATCAAATAA
HlyIIR[9]	ATATTTAAAATTCTTGTTTAAA
IcaR (A)[10]	TTCACCTACCTTTCGTTAGGTTA
LmrA[11]	GATAATAGACCAGTCACTATATTT
PhlF[12]	ATGATACGAAACGTACCGTATCGTTAAGGT
SmcR[13]	TTATTGATAAATCTGCGTAAAAT
TetR[14]	TCCCTATCAGTGATAGA

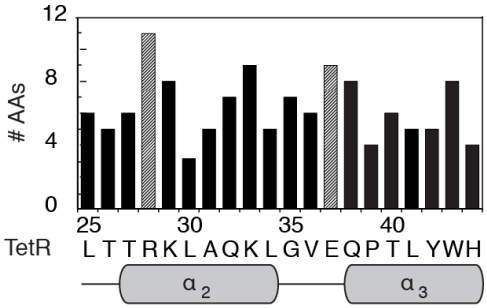
*Superscripts indicate operator sequence references.

Supplementary Figure 4: Orthogonality measurement plasmids maps. Orthogonality measurements were obtained using 2 plasmids: one expresses the repressor and the second contains the promoter reporters. In this way, the two plasmids can be co-transformed to build all of the strains required for the orthogonality screen. For the repressor library, each repressor is placed under the control of a 3OC6HSL inducible system (the pOrtho set of plasmids). For the reporters, the same plasmids are used as were built to measure the response functions (Supplementary Figure 5), but the repressors encoded by these plasmids are not induced.

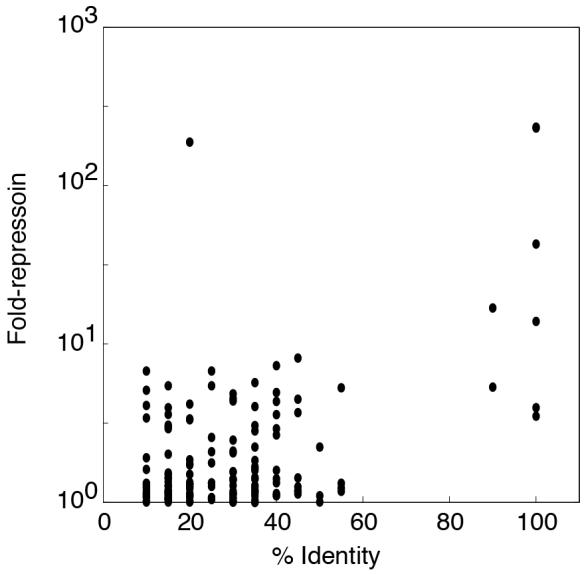
Nature Chemical Biology: doi:10.1038/nchembio.1411

Supplementary Figure 5: NOT gate plasmid maps. These plasmids are used to calculate the response functions shown in Figure 4. The Response Function vectors (pRF-) contain an individual repressor, whose expression is controlled by the P_{Tac} inducible promoter (which corresponds to a version of P_{tac1}[15] that has been modified to contain a perfect inverted repeat sequence for the Lac operator). Each NOT gate also contains the cognate promoter for the repressor, which controls expression of the YFP output. The terminator present after the repressor coding sequence corresponds to BBa_B0015, a double terminator consisting of both BBa_B0010 and BBa_B0012 (partsregistry.org). The wild type promoter of the Lac Repressor (labeled P_{Const})[16] constitutively expresses both LacI and LuxR. These components are maintained on a lower copy number plasmid that was derived from the expression plasmid pEXT20[17]. Activation of repressor expression by IPTG results in repression of the promoter driving YFP (Figure 4).

Nature Chemical Biology: doi:10.1038/nchembio.1411



Supplementary Figure 6: DNA-binding domain recognition region diversity. The recognition regions of the DNA-binding domains for all 20 repressors were aligned, and the number of different residues at each position across the set was counted. The wild-type sequence of TetR is shown below the plot for reference, along with the secondary structure of the protein.



Supplementary Figure 7: Fold-repression versus percent pairwise identity of the recognition region. The fold-repression values of all repressor-promoter pairings are the mean repression values from triplicate orthogonality measurements (Figure 3d). These data are plotted versus the corresponding percent pairwise sequence identity of the recognition regions of the repressors' DNA-binding domains.

Nature Chemical Biology: doi:10.1038/nchembio.1411

Supplementary Table 3: Degenerate NOT gate repressor RBS sequences

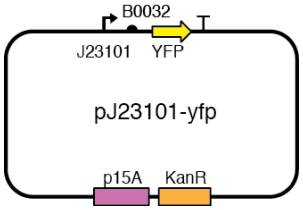
Repressor	RBS Library Sequence*
AmeR	CTATGGACTATGTTTTCACANANGANGNGGATTAG**ATG**
AmtR	CTATGGACTATGTTTGANAGANANAATACTAG**ATG**
BetI	GCTACGACTTGCTCATTTGANAGAGGANAANTACTA**GTG**
BM3R1	CTATGGACTATGTTTNAANTACTAG**ATG**
ButR	CTATGGACTATGTTTTCASASRGGARRTACTASG**ATG**
HapR	CTATGGACTATGTTTAAAGAGGANANNTACTAG**ATG**
HylIIR	CTATGGACTATGTTTGAAAGAGGGANAAANACTAN**ATG**
IcaR	CTATGGACTATGTTTTCACACAGGGSCYSG**ATG**
LitR	CTATGGACTATGTTTTCACACAGGTTTTCACACAGRARARRCCTCG**ATG**
LmrA	CTATGGACTATGTTTTCACACAGGAAAGGNCTCG**ATG**
McbR	CTATGGACTATGNAGGANAANTACTAG**ATG**
Orf2	CTATGGACTATGTTTTGAAAGAGGAGAAANNCTAG**ATG**
PhlF	CTATGGACTATGTTTGANANGGANAANTACTAG**ATG**
PsrA	CTATGGACTATGTTTSAMASAGGATACRAMMTACTAG**ATG**
QacR	GCCATGCCATTGGCTTTTCACASAGGAMAMCKRYTMG**ATG**
ScbR	CTATGGACTATGTTTAMASAGGARAMSTACTAG**ATG**
SmcR	CTATGGACTATGTTTSAMASAGGARRRRWWYTMG**ATG**
SrpR	CTATGGACTATGTTTTSAMASAGGAAMTACMAGS**ATG**
TarA	CTATGGACTATGTTTTTSAMASAGGARAMMTACTAG**ATG**
TetR	CTATGGACTATGTTTTCACACAGGAAAGGCCTCG**ATG**

*Codes are defined as N = A,T,G, or C, S = G or C, R = A or G, Y = T or C, M = A or C, K = G or T, and W = A or T.

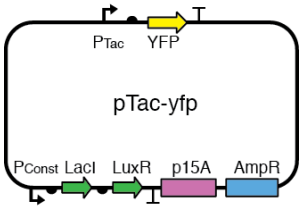
Supplementary Table 4: NOT gate repressor RBS sequences

Repressor	RBS Library Sequence
AmeR	CTATGGACTATGTTTTCACATACGAGGGGGATTAG**ATG**
AmtR	CTATGGACTATGTTTGAAAGAGAGAATACTAG**ATG**
BetI	GCTACGACTTGCTCATTTGACAGAGGATAACTACTA**GTG**
BM3R1	CTATGGACTATGTTTTAACTACTAG**ATG**
ButR	CTATGGACTATGTTTTCACACAGGAAATACTACG**ATG**
HapR	CTATGGACTATGTTTAAAGAGGACACATACTAG**ATG**
HylIIR	CTATGGACTATGTTTGAAAGAGGGACAAACACTAA**ATG**
IcaR (A)	CTATGGACTATGTTTTCACACAGGGGCCGG**ATG**
LitR	CTATGGACTATGTTTTCACACAGGGTTTTCACACAGGAGAAACCTCG**ATG**
LmrA	CTATGGACTATGTTTTCACACAGGAAAGGCCTCG**ATG**
McbR	CTATGGACTATGTAGGAGAAATACTAG**ATG**
Orf2	CTATGGACTATGTTTTGAAAGAGGAGAAACACTAG**ATG**
PhlF	CTATGGACTATGTTTGAAAGGGAGAAATACTAG**ATG**
PsrA	CTATGGACTATGTTTGAAAGAGGATACGAACTACTAG**ATG**
QacR	GCCATGCCATTGGCTTTTCACACAGGACACCGGTTAG**ATG**
ScbR	CTATGGACTATGTTTAAAGAGGAAAAGTACTAG**ATG**
SmcR	CTATGGACTATGTTTGAAAGAGGAGAAATACTAG**ATG**
SrpR	CTATGGACTATGTTTTCACACAGGAAATACCAGG**ATG**
TarA	CTATGGACTATGTTTTCAAAGAGGAGAAATACTAG**ATG**
TetR	CTATGGACTATGTTTTCACACAGGAAAGGCCTCG**ATG**

Nature Chemical Biology: doi:10.1038/nchembio.1411



Supplementary Figure 8: Reference plasmid for converting fluorescence units to REU. The fluorescent measurements are normalized by the fluorescence produced from a constitutive promoter (BBa_J23101)[9]. The corresponding output, defined as a single REU, serves as the unit to which all other fluorescence values are normalized (Online Methods).



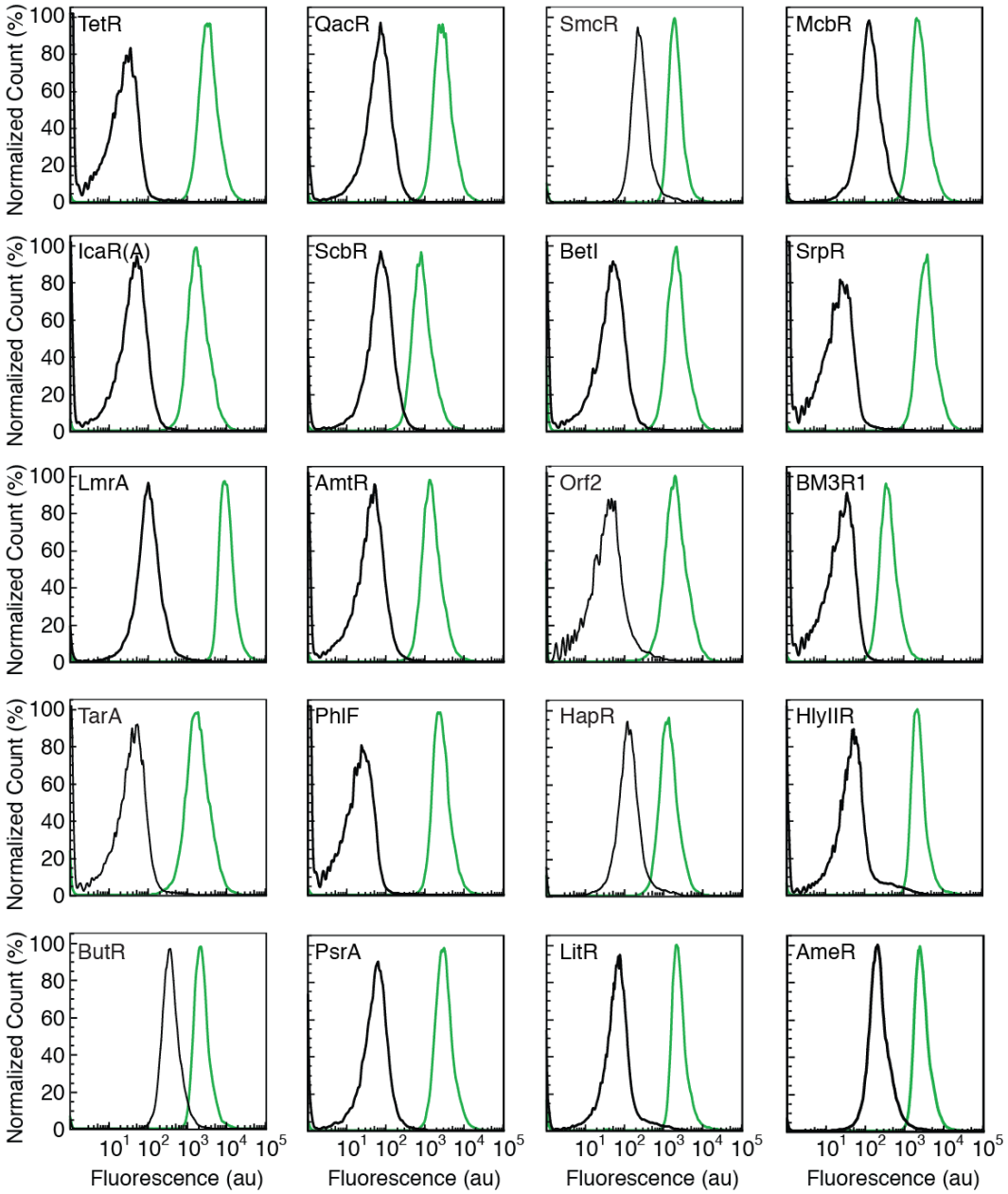
Supplementary Figure 9: Response function input measurement plasmid. To report the response function input as REU, the activity of the input promoter is measured separately.

Supplementary Table 5: NOT gate response function parameters

Name	*K**	*n*	y_{max}*	y_{min}*	Fold-change**
TetR	0.1	2.7	24	0.2	137
QacR	0.5	1.4	21	0.2	32
IcaR(A)	0.4	1.8	13	0.4	34
AmeR	0.5	1.4	17	1.7	10
ScbR	0.2	2.6	5	0.6	8
LmrA	1.2	3.1	70	1.1	61
AmtR	0.2	1.8	9	0.3	28
SmcR	0.1	2	13	2.1	5
McbR	0.4	1.6	16	1.1	14
BetI	0.2	2.4	13	0.4	35
SrpR	0.3	3.2	25	0.1	207
Orf2	0.4	6.1	14	0.2	46
BM3R1	0.6	4.5	3	0.1	28
TarA	0.1	1.8	13	0.2	49
PhlF	0.4	4.5	16	0.1	193
ButR	1.3	2.4	12	1.8	5
PsrA	0.4	2	20	0.5	43
HapR	0.2	1.4	10	0.9	8
HlyIIR	0.5	2.7	17	0.3	48
LitR	0.1	1.9	16	0.5	35

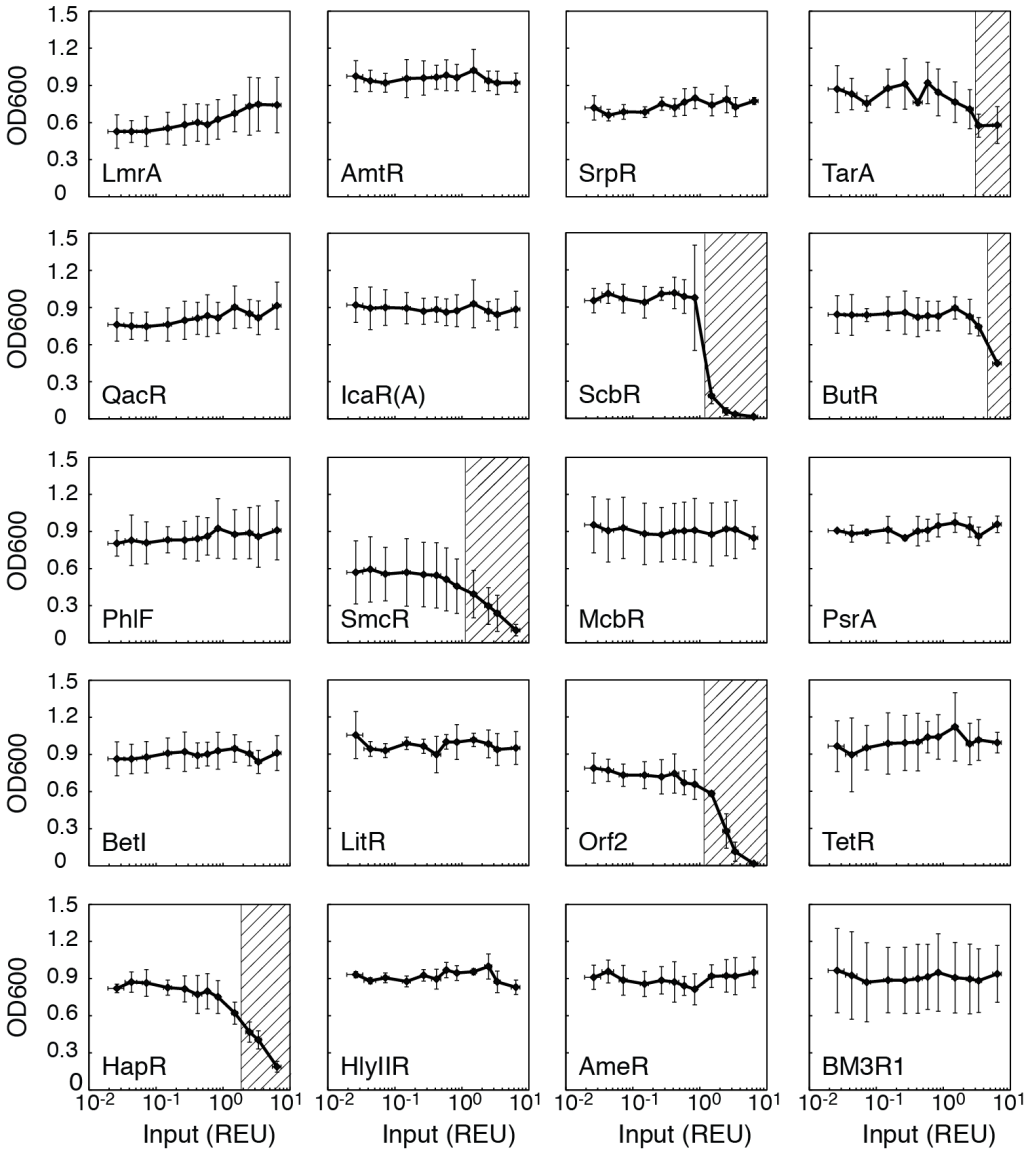
*Parameter values are listed in REUs. **Fold-change was calculated by dividing the highest ON state by the lowest non-toxic OFF state.

Nature Chemical Biology: doi:10.1038/nchembio.1411



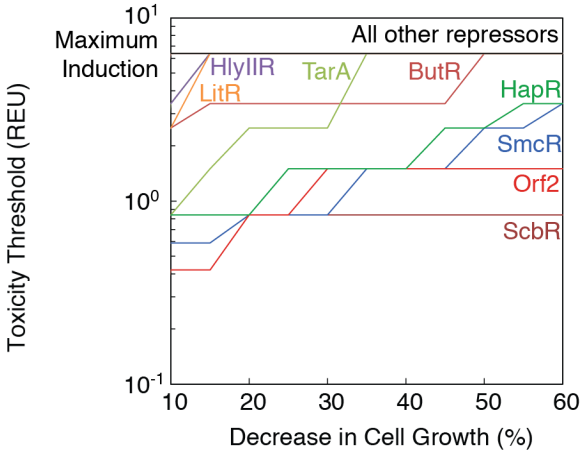
Supplementary Figure 10: Flow cytometry data for each NOT gate. Fluorescence histograms correspond to representative single cytometry replicates for induced (black) and uninduced (green) states. The induced state corresponds to the highest IPTG concentration before toxicity was observed (200 μM for ButR, 150 μM for TarA, 100 μM for HapR, 70 μM for ScbR, 70 μM for SmcR, 70 μM for Orf2, and 1 mM IPTG for all other repressors). Each histogram comprises >10000 cells.

Nature Chemical Biology: doi:10.1038/nchembio.1411

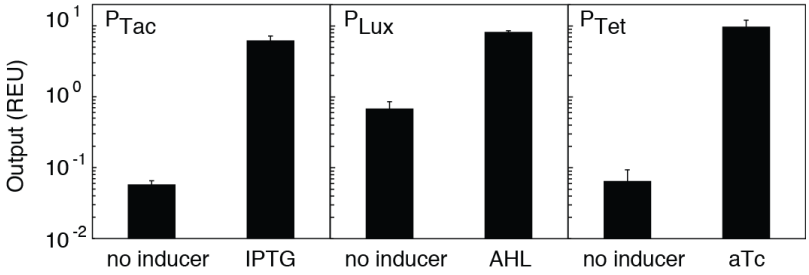


Supplementary Figure 11: Growth measurements for NOT gate response functions. The optical density at 600 nanometers was measured for all NOT gates at each of the twelve inducer concentrations: 0, 5, 10, 20, 30, 40, 50, 70, 100, 150, 200, 1000 μM IPTG in an analogous manner to the response functions (Figure 4). The x-axis values are converted to the REU values measured for the response function assay. Toxicity is indicated by the hash-marked region, and begins when the cell growth falls below 75 percent of the uninduced cell growth. Each data point was measured in triplicate on three separate days, and the data represent mean values ± 1 standard deviation.

Nature Chemical Biology: doi:10.1038/nchembio.1411

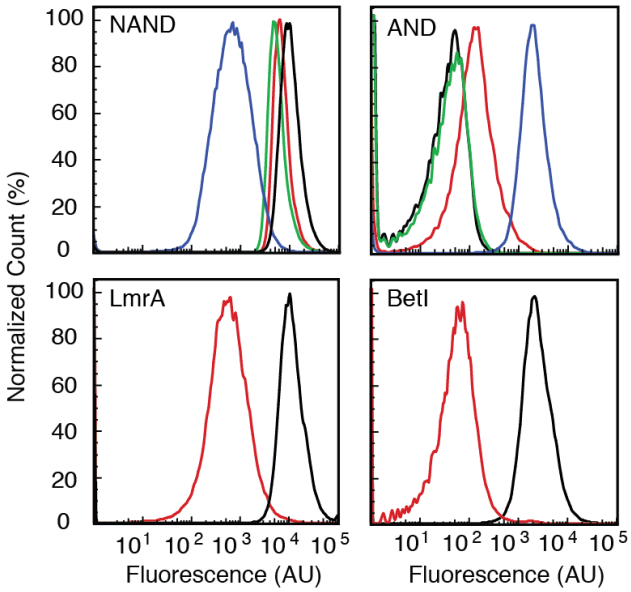


Supplementary Figure 12: Toxic induction threshold versus decrease in cell growth. The highest input level before toxicity is observed is plotted versus the percent decrease in cell growth. For most repressors, toxicity is not observed, and is indicated by the horizontal black line at the top of the graph. HlyIIR and LitR exhibit a 10 percent decrease in growth at high induction levels. The cross-section of the toxicity trajectories at 25% decrease in cell growth for TarA, ButR, HapR, SmcR, Orf2, and ScbR is reflected in the toxic regions of Figure 4 and Supplementary Figure 11. Threshold data (y-axis) represents mean maximum induction levels before the growth decreases beyond a mean percentage (x-axis) from three separate experiments. Each data point was measured in triplicate on three separate days.



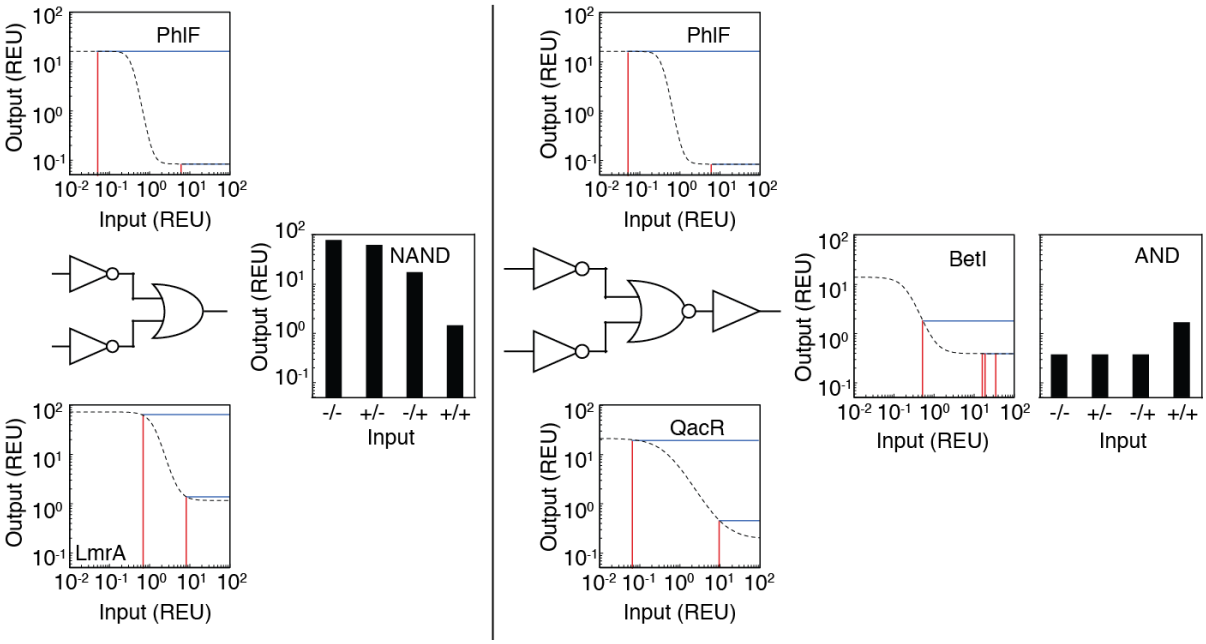
Supplementary Figure 13: Characterization of inducible promoters. Promoters P_{Tac}, P_{Lux}, and P_{Tet} drive yellow fluorescent protein expression and were induced with 1 mM IPTG, 20 μM 3OC6-HSL, and 100 ng/mL aTc, respectively. Cells grown under maximum inducing and non-inducing conditions were measured via cytometry; fluorescence values were normalized by an in vivo reference standard to obtain the promoters' outputs in REU (Supplementary Figure 8). Data was collected in triplicate on three different days and points represent mean values ± 1 standard deviation.

Nature Chemical Biology: doi:10.1038/nchembio.1411



Supplementary Figure 14: Flow cytometry data for logic circuits and terminal gates. Upper panel: Representative fluorescence histograms that correspond to the average fluorescence values in Figure 5a, b. For the NAND circuit, the black line corresponds to no inducer, green to 1 mM IPTG, red to 20 μM 3OC6HSL, and blue to the presence of both IPTG and 3OC6HSL. For the AND circuit, the black line corresponds to no inducer, red to 1 mM IPTG, green to 100 ng/mL aTc, and blue to the presence of both IPTG and aTc. Lower panel: Representative fluorescence histograms for repressors connected to circuit outputs. The output distributions for the terminal repressors were taken from response function characterization data, and input levels were chosen such that they approximate the predicted levels seen within the circuits. Fluorescence histograms correspond to representative single cytometry replicates, and each histogram comprises >10000 cells.

Nature Chemical Biology: doi:10.1038/nchembio.1411



Supplementary Figure 15: Modeling of genetic circuits. For the first layer of gates, experimentally characterized input promoter values (red lines) are mapped onto Hill-equation fits of NOT gate response functions (dashed lines), resulting in predicted output values (blue lines) that feed into the next logic layer. For the NAND gate, the individual NOT gate output values from the first layer are summed to yield the final circuit output. For the AND gate, the individual NOT gate outputs from the first layer are summed to yield the BetI inputs (red lines) that drive the final NOR gate output.

Supplementary Table 6: NAND circuit modeling

Input*		Internal*		Output*
P_{Tac}	P_{Lux}	P_{PhlF}	P_{LmrA}	P_{PhlF}-P_{LmrA}
0.06	0.7	16	61	78
6.2	0.7	0.1	61	62
0.06	8.2	16	1.4	17
6.2	8.2	0.1	1.4	1.4

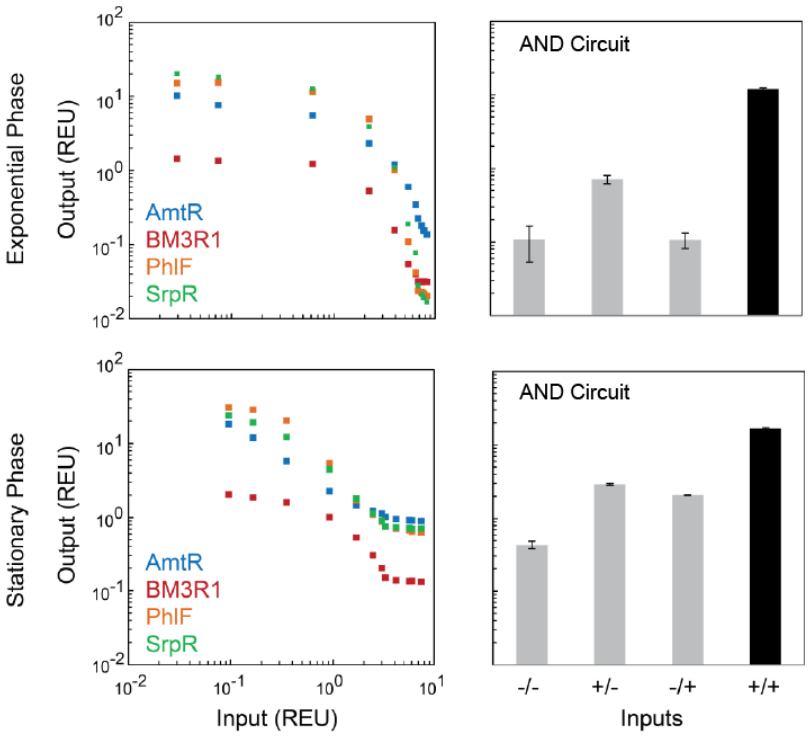
*Parameters used for NAND gate modeling are listed in REUs.

Supplementary Table 7: AND circuit modeling

Input*		Internal*		Output*
P_{Tac}	P_{Tet}	P_{PhlF}	P_{QacR}	P_{BetI}
0.06	0.07	16	20	0.4
6.2	0.07	0.1	20	0.4
0.06	9.8	16	0.4	0.4
6.2	9.8	0.1	0.4	1.7

*Parameters used for NAND gate modeling are listed in REUs.

Nature Chemical Biology: doi:10.1038/nchembio.1411



Supplementary Figure 16: Growth phase robustness of repressors and AND circuit. Left panel: Transfer functions for AmtR (blue squares), BM3R1 (red squares), PhlF (orange squares), and SrpR (green squares) NOT gates measured in exponential phase (top) and stationary phase (bottom) grown in LB media are illustrated. Right panel: The output values for the AND circuit measured in exponential phase and stationary phase in LB media are illustrated. The measured data are grown under conditions of no inducer (-/-), 1 mM IPTG (+/-), 100 ng/mL aTc (-/+), and 1 mM IPTG and 100 ng/mL aTc (+/+). Bars corresponding to the ON and OFF states are colored black and gray, respectively, and all measurements correspond to the average of three technical replicates.

References

1 Rey, D. A. *et al.* The McbR repressor modulated by the effector substance S-adenosylhomocysteine controls directly the transcription of a regulon involved in sulphur metabolism of Corynebacterium glutamicum ATCC 13032. *Mol Microbiol* **56**, 871-887, doi:10.1111/j.1365-2958.2005.04586.x (2005).

2 Kojic, M., Aguilar, C. & Venturi, V. TetR family member psrA directly binds the Pseudomonas rpoS and psrA promoters. *J Bacteriol* **184**, 2324-2330 (2002).

3 Grkovic, S., Brown, M. H., Schumacher, M. A., Brennan, R. G. & Skurray, R. A. The staphylococcal QacR multidrug regulator binds a correctly spaced operator as a pair of dimers. *J Bacteriol* **183**, 7102-7109, doi:10.1128/JB.183.24.7102-7109.2001 (2001).

4 Takano, E. *et al.* A bacterial hormone (the SCB1) directly controls the expression of a pathway-specific regulatory gene in the cryptic type I polyketide biosynthetic gene cluster of Streptomyces coelicolor. *Mol Microbiol* **56**, 465-479, doi:10.1111/j.1365-2958.2005.04543.x (2005).

5 Hasselt, K., Rankl, S., Worsch, S. & Burkovski, A. Adaptation of AmtR-controlled gene expression by modulation of AmtR binding activity in Corynebacterium glutamicum. *Journal of biotechnology* **154**, 156-162, doi:10.1016/j.jbiotec.2010.09.930 (2011).

6 Rkenes, T. P., Lamark, T. & Strom, A. R. DNA-binding properties of the BetI repressor protein of Escherichia coli: the inducer choline stimulates BetI-DNA complex formation. *J Bacteriol* **178**, 1663-1670 (1996).

Nature Chemical Biology: doi:10.1038/nchembio.1411

7 Shaw, G. C. & Fulco, A. J. Inhibition by barbiturates of the binding of Bm3R1 repressor to its operator site on the barbiturate-inducible cytochrome P450BM-3 gene of Bacillus megaterium. *The Journal of biological chemistry* **268**, 2997-3004 (1993).

8 Tsou, A. M., Cai, T., Liu, Z., Zhu, J. & Kulkarni, R. V. Regulatory targets of quorum sensing in Vibrio cholerae: evidence for two distinct HapR-binding motifs. *Nucleic Acids Res* **37**, 2747-2756, doi:10.1093/nar/gkp121 (2009).

9 Rodikova, E. A. *et al.* Two HlyIIR dimers bind to a long perfect inverted repeat in the operator of the hemolysin II gene from Bacillus cereus. *FEBS letters* **581**, 1190-1196, doi:10.1016/j.febslet.2007.02.035 (2007).

10 Jefferson, K. K., Cramton, S. E., Gotz, F. & Pier, G. B. Identification of a 5-nucleotide sequence that controls expression of the ica locus in Staphylococcus aureus and characterization of the DNA-binding properties of IcaR. *Mol Microbiol* **48**, 889-899 (2003).

11 Yoshida, K. *et al.* Bacillus subtilis LmrA is a repressor of the lmrAB and yxaGH operons: identification of its binding site and functional analysis of lmrB and yxaGH. *J Bacteriol* **186**, 5640-5648, doi:10.1128/JB.186.17.5640-5648.2004 (2004).

12 Abbas, A. *et al.* Characterization of interactions between the transcriptional repressor PhlF and its binding site at the phlA promoter in Pseudomonas fluorescens F113. *J Bacteriol* **184**, 3008-3016 (2002).

13 Lee, D. H. *et al.* A consensus sequence for binding of SmcR, a Vibrio vulnificus LuxR homologue, and genome-wide identification of the SmcR regulon. *The Journal of biological chemistry* **283**, 23610-23618, doi:10.1074/jbc.M801480200 (2008).

14 Klock, G. *et al.* Heterologous repressor-operator recognition among four classes of tetracycline resistance determinants. *J Bacteriol* **161**, 326-332 (1985).

15 de Boer, H. A., Comstock, L. J. & Vasser, M. The tac promoter: a functional hybrid derived from the trp and lac promoters. *Proc Natl Acad Sci U S A* **80**, 21-25 (1983).

16 Calos, M. P. DNA sequence for a low-level promoter of the lac repressor gene and an 'up' promoter mutation. *Nature* **274**, 762-765 (1978).

17 Dykxhoorn, D. M., St Pierre, R. & Linn, T. A set of compatible tac promoter expression vectors. *Gene* **177**, 133-136 (1996).

Nature Chemical Biology: doi:10.1038/nchembio.1411