

Snowprint: a predictive tool for genetic biosensor discovery

Simon d'Oelsnitz*¹, Sarah Stofel¹, Andrew D. Ellington¹

Affiliations

¹Department of Molecular Biosciences, University of Texas at Austin, Austin, TX, 78712, USA

Supplementary Information:

Supplementary Figures 1 - 7

Supplementary Tables 1 - 4



Processed regulators

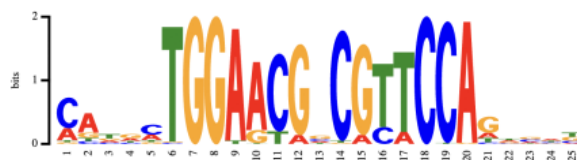
ID	Accession	Score	Sequences Aligned	Organism	Consensus
0	MBW8800226.1	87.754	99	Streptomyces sp.	catgcTGGAACGGCGTTCCAgtagat
1	WP_253371901.1	87.754	99	Streptomyces sp.	catgcTGGAACGGCGTTCCAgtagat
2	QIS11675.1	80.303	98	Nocardia arthritidis	catgcTGGAACGGCGTTCCAgacag
3	WP_148432323.1	93.336	98	Kutzneria sp. CA-103260	cacatTGGAACGACGTTCCAgtagt
4	WP_011031522.1	87.754	99	Streptomyces sp.	catgcTGGAACGGCGTTCCAgtagat

1 row selected

1-5 of 5

MBW8800226.1

Snowprint



Inter-operon region

TGGCGCCATCTTGACATGCCTGGAGCGTCGTTCCAACATGGAACGACGTTCCAGACTGGAATGCCGTTCCACTCTCTTCGGGTGGCGGCC
GGCCACAAGAGCCCGGCCCTGCTCTGCCCTTCTCGGAGGAGTCACCA

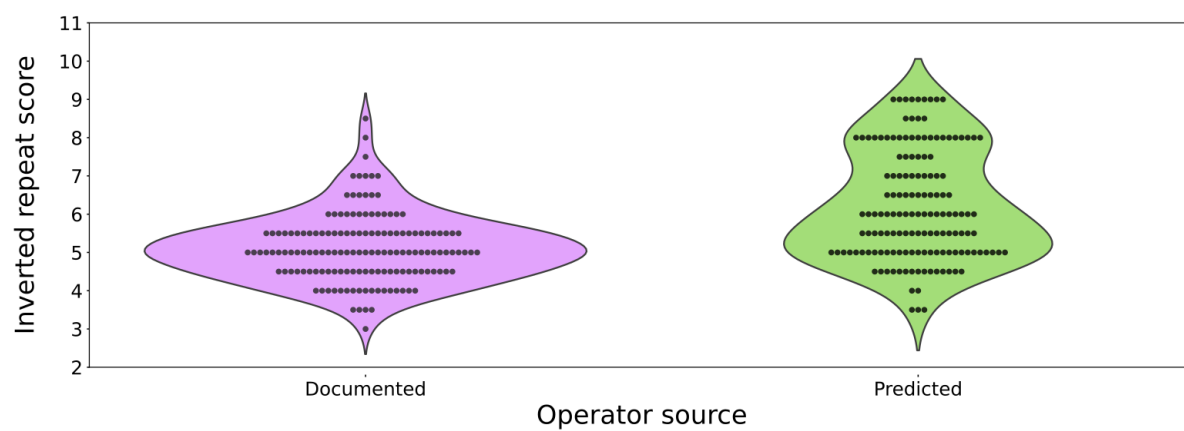
All homologs

ID	Accession	Protein identity	Protein coverage	Predicted operator	Alignment score	Organism
0	MBW8800226.1	100.000	100	caacaTGGAACGACGTTCCAgactg	27.5	Streptomyces
1	WP_217553150.1	83.203	94	atggcTGGAGCGCCGTTCCAcactg	25	Streptomyces
2	WP_234543907.1	80.000	87	cagacTGGAACGTTGTTCCAgaaaa	25	Streptomyces
3	WP_088100280.1	83.794	93	catgcTGGAACGGCGCTCCAcacat	25	Streptomyces
4	WP_189262991.1	79.693	91	caggtTGGAAATGTCGTTCCAgagtg	25	Streptomyces
5	WP_201855583.1	69.466	91	catgrTGGAACGACGCTCCAglicat	25	Streptomyces
6	WP_247051124.1	71.951	90	caagaTGGAAATGACGTTCCAgtagag	25	Amycolatopsis
7	WP_191063196.1	70.896	91	caacaTGGAGCGCGCTTCCAgtaact	25	Streptomyces
8	WP_207050538.1	67.984	93	gtttcTGGAACGACGTTCCAtggcc	27.5	Coralloporum
9	WP_275561155.1	71.698	91	agtaTGGAACGCCGCTCCAtgttg	25	Streptomyces
10	WP_151542274.1	76.113	90	lccatTGGAACGCCGTACCAgaggg	25	Actinomyces
11	WP_220296077.1	72.065	90	cagtaTGGAACGCCGTACCAgagtg	25	Streptomyces
12	WP_198279688.1	71.429	91	cattcTGGAACGTCGCTCCAtgttg	25	Streptomyces
13	MRWR706710.1	72.065	90	cantaTGGAACGCCGTACCAgaggg	25	Streptomyces

Rows per page: 100 1-99 of 99

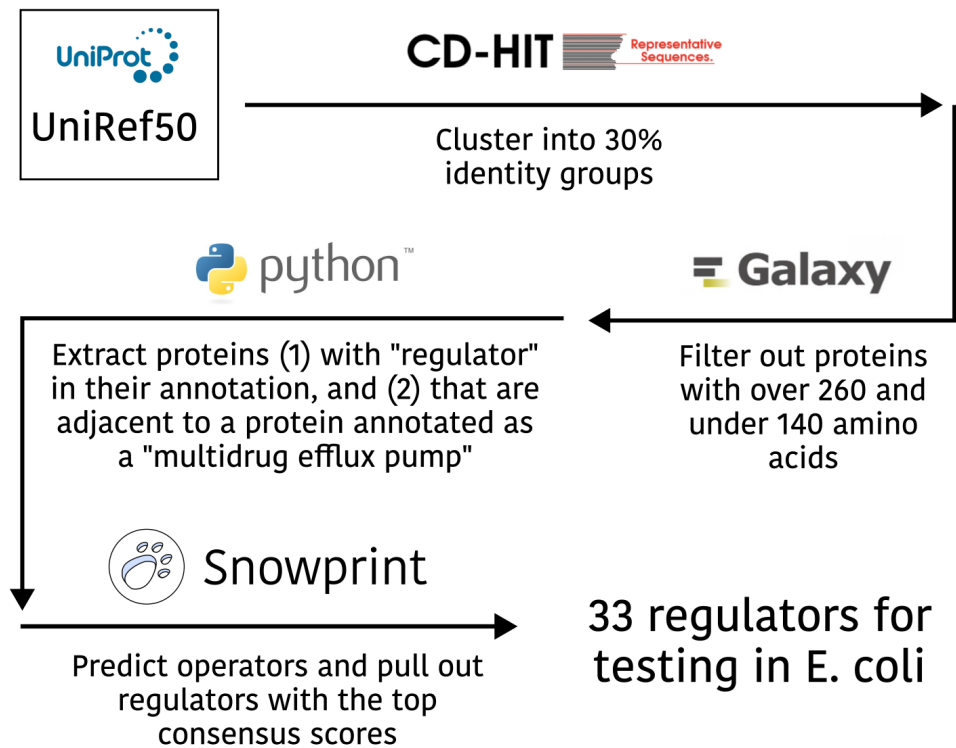
Supplementary Figure 1. React-based local web page displaying results from a Snowprint query.

Users can browse among predictions made for regulators run locally and view the (1) consensus prediction logo, (2) Inter-operon region containing the predicted operator, and (3) homologous predicted operators used to create the consensus prediction displayed in (1).



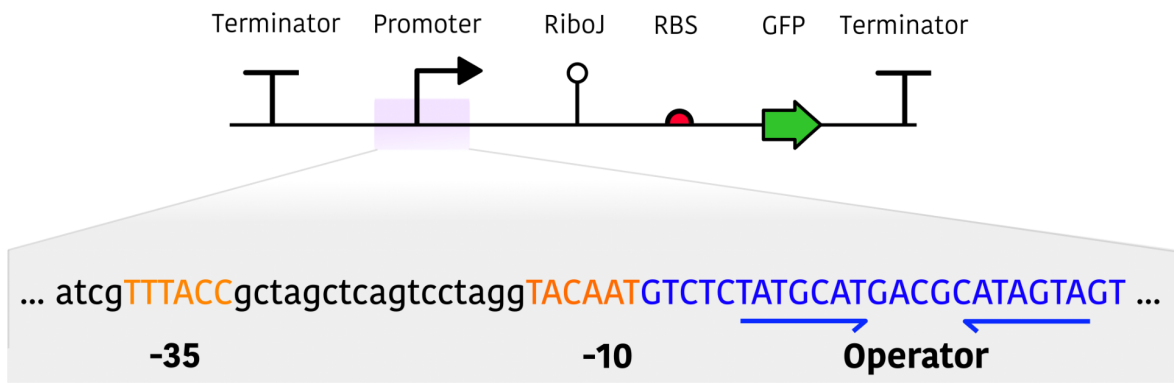
Supplementary Figure 2. Inverted repeat scores of operators within the benchmarking dataset.

Documented operators, shown in purple, represent the operator sequences extracted from literature (see **Supplementary Data 1**), while predicted operators, shown in green, represent operator sequences predicted by Snowprint using the associated regulator RefSeq accession ID.



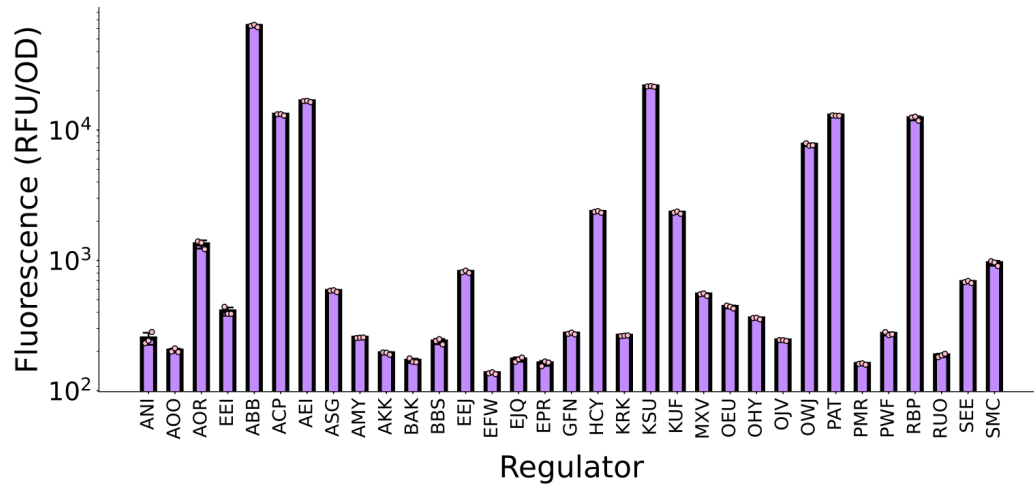
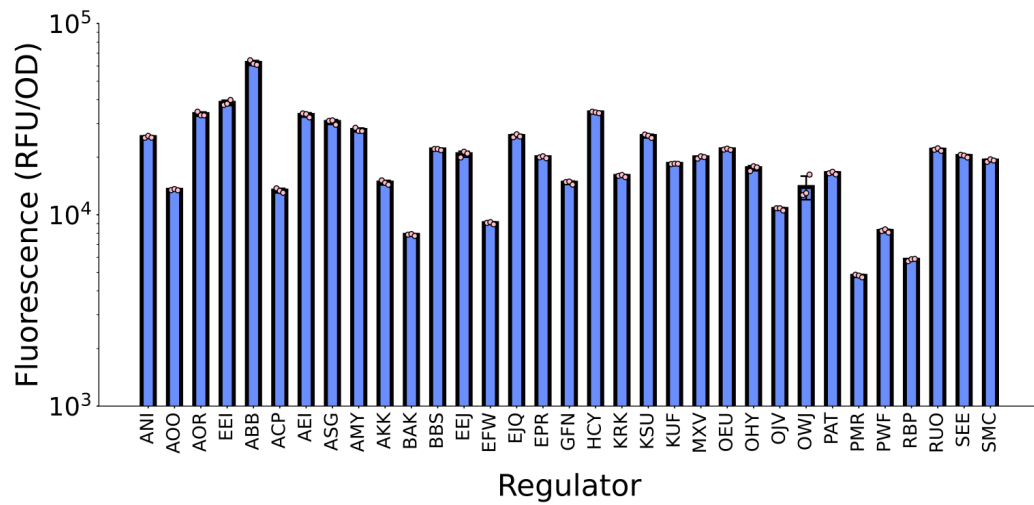
Supplementary Figure 3. Generalist regulator curation workflow.

TetR-family regulators were fetched from the UniRef50 sequence database, which were then clustered into 30% similarity groups using the CD-HIT software and filtered to contain only protein sequences shorter than 260 amino acids and longer than 140 amino acids using the Galaxy Suite software. Custom python scripts were then used with the Entrez API to filter proteins that contain “regulator” in their annotation and that are adjacent to proteins annotated as “multidrug efflux pumps”. The resulting regulators were then used as inputs to generate operator predictions with Snowprint, and the 33 regulators with the top consensus scores were selected for functional testing within *E. coli* cells.



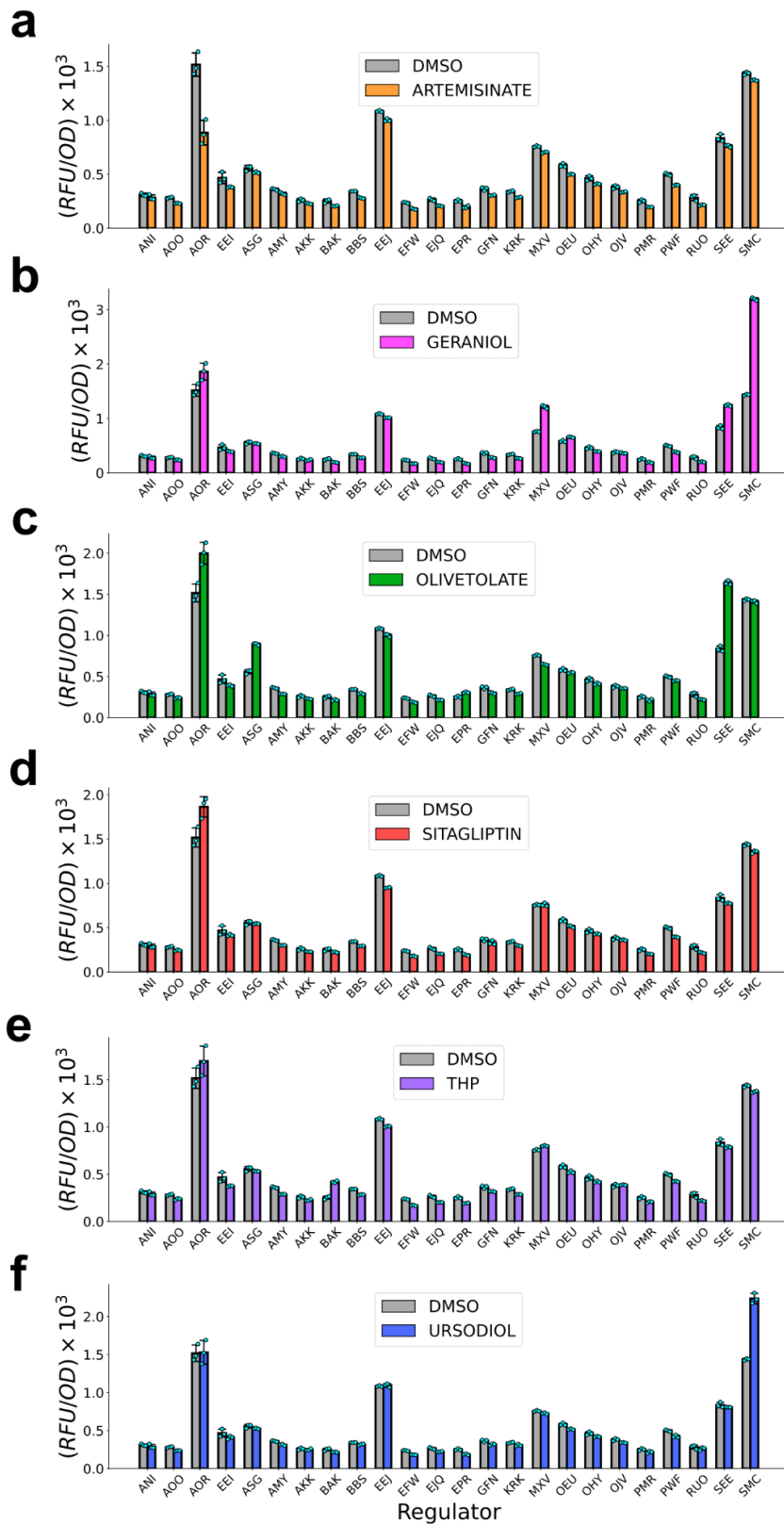
Supplementary Figure 4. Promoter design for domesticated regulators

Regulator-controlled promoters were designed by placing the predicted operator sequence immediately downstream from the -10 box of an *E. coli* sigma-70 promoter. Arrows indicate inverted repeat sequences, showing that five base pairs outside of the inverted repeat region were included in the promoter. The broader context of the reporter transcriptional unit is shown above the promoter sequence.

a**b**

Supplementary Figure 5. Synthetic promoter output with and without domesticated regulator expression.

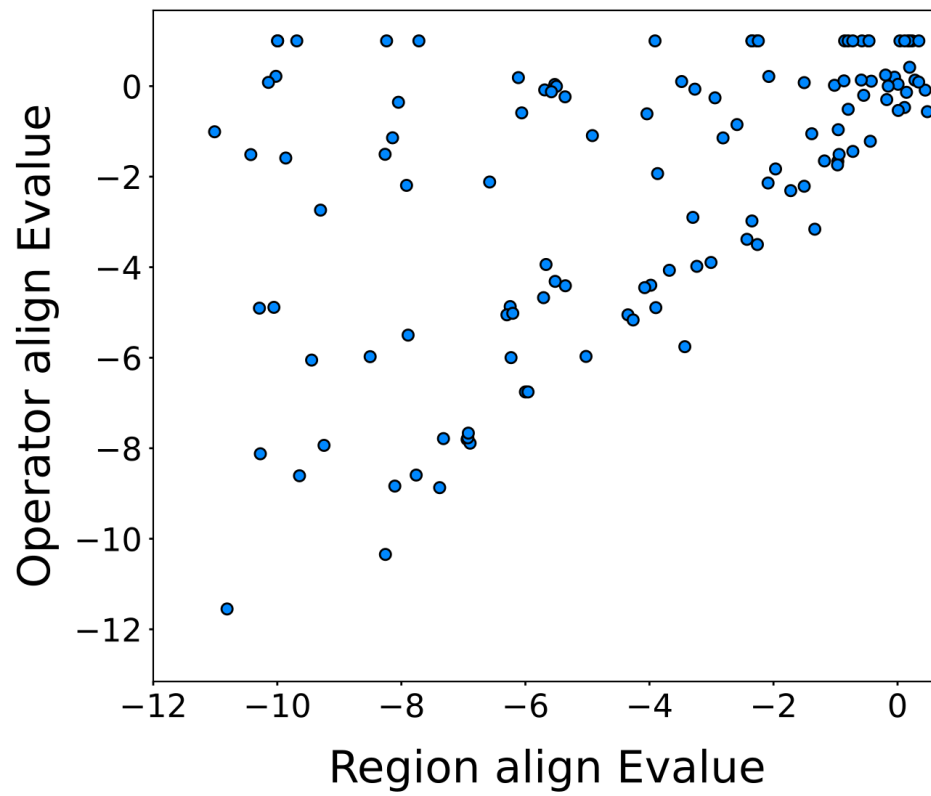
Fluorescence of *E. coli* cells containing a reporter plasmid expressing GFP from a promoter bearing a Snowprint-predicted operator (see Figure 3, Supplementary Figure 4) and a regulator plasmid expressing a transcription factor from a constitutive plasmid. In (a) a regulator plasmid expresses the transcription factor predicted to bind to the Snowprint-predicted operator. In (b) the regulator plasmid expresses the CamR transcription factor. Assays were performed in biological triplicate. Individual data points are shown in pink. Error bars represent the standard deviation +/- the mean.



Supplementary Figure 6. Response of 24 generalist regulators to biomanufacturing-relevant ligands.

The final concentration for induction of each ligand was 100 μ M. Assays were performed in biological triplicate.

Individual data points are shown in cyan. Error bars represent the standard deviation $\pm$ the mean.



Supplementary Figure 7. Comparison of known operator alignment to the predicted operator vs the inter-operon region. Single data points represent the E-value similarity score of a known operator for its corresponding Snowprint predicted operator and the inter-operon region within the host organism's genome predicted to contain the operator.

Protein ID	E-value	Alignment			
WP_001224188.1	3.7*10 ⁻¹³	Documented	1 gttttataATAAACGGAgagttaTCCGTTTGtcaa	35	
		Predicted	1 -----ataatAAACGGATAGTTATCCGTTTgtcaa	30	
AEM66515.1	0.0012	Documented	1 TAGCCACGTCCTGGACAAAGTGAGAGATCGTGTCTAGACAACGCCACGGTT	50	
		Predicted	1 -----gatcgTGTCTAGACAacgcc-----	20	
NP_415533.1	0.0076	Documented	1 TGTAAATTTTGACCATTGGTCCACTTTTTTCT	33	
		Predicted	1 -TTAACTTTTAAAAACTGGC-----	20	
WP_011060270.1	0.0064	Documented	1 ----TCAAACAAGTGTGTGTCAGG---	20	
		Predicted	1 aagccTGAAACGTATGTT--TCAaacia	26	
WP_005058758.1	0.0072	Documented	1 -----CTTAACGCATGGCATGTGCGTTATG	25	
		Predicted	1 ggtagCGTGCCTTACGCACGtcata-----	25	
CAY46636.1	0.0049	Documented	1 ---GCT-----TGATGTACAAGT--	16	
		Predicted	1 acgtcTGTATATGTATATACAaataa	26	
NP_391277.1	0.0061	Documented	1 ---AAATTGTCCGTATACATTTT--	21	
		Predicted	1 atttaATTAGTACGTACAAATataga	26	
NP_227848.1	0.0018	Documented	1 --AATTTCTTTCTGAG-GAAGATAGA	23	
		Predicted	1 agacaTTCTCAAAGTGAGAAgaatg-	25	
WP_003084488.1	1.0	Documented	1 tttcTAG--AGTAaaTACTCTAaagt--	24	
		Predicted	1 --tcacgGCGGGGCGCGCCCGCactgt	26	

Supplementary Table 1. Global alignment of documented and predicted operators.

The first and last alignments display excellent and poor scores for comparison, respectively. All other alignments have E-values between 0.01 and 0.001. The pairwise EMBOSS needle algorithm was used to create alignments with default parameters. Documented operators represent sequences experimentally validated to bind the corresponding protein regulator. Refer to Supplementary Data 1 for more information on sequence metrics and sources.

Genbank ID	Number of aligned sequences	Consensus Score	Predicted operator
ANI80042.1	7	94.99	gttcgAAAATAAACAATCTTGTTTCTTTattca
AOO80349	12	89.099	cctgcTAAACTAAACGGTTCAGTTTAttcgc
AOR61801.1	10	62.821	tatacCCATAATTAATGACCCAGAAGTCATTAATTATGGtggtcc
EEI22781.1	38	95.314	ctcatAGTGACACTGTGTCACTcgatt
ABB11310.1	24	85.375	cgattTGCCTGCCCCGGGTGCCGGGGCACGCAtcctg
ACP22327	15	75.794	cttatTAGATTTCATTTGACATCTatatat
AEI82798.1	38	71.203	cttTGCCTGCCcGGACCGCAtga
ASG24684.1	9	100	ggactGTACTGTATGATACAGTACgtaat
AMY71450	11	82.012	caatcACTGTACCGTCCAGACGGTAAAGTcaacc
AKK09602.1	24	68.741	acataTATTTTCATCAACCGATGAAATAcggag
BAK71752.1	36	89.688	tttatAATACATACGTATGTATTtaata
BBS35964.1	8	71.611	agataAGCTAAACTTGAATAAAGATTCAAGTTTAGCTacttt
EEJ50829.1	21	76.158	ggtgtGGTGACAAGTTGTCACCggttg
EFW06302.1	70	72.686	cctttTTTAAATGAATATTTATTCATTCAAAttata
EJQ60916.1	99	92.352	attgtCTAACAGTGTTAGtttga
EPR37357.1	9	90.361	tccagAACTGACCGGTCAGTTtttga
GFN32285.1	9	85.012	acaatATGAATGAATCATTCAttcga
HCY59852	24	81.386	aaACCTAGTAAGTTACTAGGTga
KRK48889	12	96.543	aacggTTGTAAATTTAAATTACAAagagt
KSU66562.1	34	56.409	aaacaGGTGAGCGCCCGAGCATTATACTCGGGCGCTCACCggttt
KUF06535.1	18	86.464	agaatATACCCATCGGTATctagg
MXV43606.1	21	82.634	ttagcGTGGAACCGATTAGTTTCCACcccct
OEU66612	6	84.907	aacacTTGTTTAAAACAAtcgtt
OHV36883.1	13	87.706	ctactACCGTCCGTCCGGACGGTcggtc
OJV33711	28	84.969	gcaaaaACGTGAGTGAACGATCACTCACGTactag
OWJ67207.1	100	45.685	cccgtCCGCCCCGATTTCTCGGGGTGCGcgctg
PAT04833.1	12	95.26	cgcccCAAGGTATTACCTTGcacac
PMR76500	16	96.12	tggtgTTAACGCCGTTAAtatag
PWF24669	15	62.766	ctctaAAATTAGGTGCGTTAACCAACCTAATTTtatgc
RBP44292.1	3	86.607	tttacAGGTATGCAAGCAAGTACTTACTTGCATACCTggtgc
RUO21779.1	11	71.091	gaatcTCGGACTGACTAGTCCGTCCGAtggtt
SEE04737	41	76.365	attgaATTGCACCAGATAGTGCAATaggag
SMC09139	8	90.222	ccTTGTGAATgagATTCACAAaa

Supplementary Table 2. Snowprint predictions for mined regulators.

Results and metadata associated with Snowprint predictions of TetR-family transcriptional regulators targeted for testing within *E. coli*. Number of aligned sequences represents the number of homologous operator sequences used to generate the consensus. See **Methods** for consensus score calculation.

TGCTTCATTCCTCGGTACCAAATTCAGAAAAGAGGGGAGCGGGAAACCGCTCCCCTTTTTTCGTTTTGGTCCCTCGC
 TGCAGTGTCTGTAACCAACAGTTGGGGAGACGCTGTGAATCCCGGGCAAGCAAAATCGCTGGAAGTTCTTATACTTTC
 TAGAGAATAGGAACCTCTTTCTAAATACATTCAAATATGTATCCGCTCATGAGACAATAACCGTGATAAATGCTTCAA
 TAATATTGAAAAAGGAAGAGTATGAGCCATATTCAACGGGAAACGCTTTGCTCCAGGCCCGGATTAAATTCCAACAT
 GGATGCTGATTTATATGGGTATAAATGGGCTCGCGATAATGTGGGCAATCAGGTGCGACAATCTATCGATTGTATGG
 GAAGCCCGATGCGCCAGAGTTGTTTCTGAAACATGGCAAAGGTAGCGTTGCCAATGATGTTACAGATGAGATGGTCA
 GACTAAACTGGCTGACGGAATTTATGCCTCTTCCGACCATCAAGCATTATCCGTACTCCTGATGATGCATGGTTAC
 TCACCACTGCGATCCCCGGGAAAAACAGCATTCCAGGTATTAGAAGAATATCCTGATTACAGGTAAAAATATTGTTGATG
 CGCTGGCAGTGTTCCTGCGCCGGTTGCATTTCGATTCTGTTTGTAAATGTCCTTTTAAACAGCGATCGCGTATTTCGCC
 TCGCTCAGGCGCAATCACGAATGAATAACGGTTTGGTTGATGCGAGTGATTTTGATGACGAGCGTAATGGCTGGCCT
 GTTGAACAAGTCTGGAAGAAATGCATAAGCTTTTGCCATTCTCACCGGATTGATCGTCACTCATGGTGATTTCTC
 ACTTGATAACCTTATTTTTGACGAGGGGAAATTAATAGGTTGTATTGATGTTGGACGAGTCCGAATCGCAGACCGATA
 CCAGGATCTTGCCATCCTATGGAACCTGCCTCGGTGAGTTTTCTCCTTCATTACAGAAACGGCTTTTTCAAAAATATGG
 TATTGATAATCCTGATATGAATAAATTCAGTTTCATTTGATGCTCGATGAGTTTTTCTAAGTTGTGATGGCGGTAGG
 AATGTAATCGTTAATCCGCAAATAACGTAAAAACCGCTTCGGCGGGTTTTTTTTATGGGGGGAGTTTAGGGAAAGAG
 CATTTGTCATCCCGTTGAATATGGCTCCCTTAACGTGAGGAAGTTCTTATACTTTCTAGAGAATAGGAACCTTACAG
 ATGGACTTGGGTGGCGGTTTCAGGAGTAGGTGCTTCCTCGCTCACTGACTCGCTGCACGAGGCAGACCTCAGCGCT
 AGCGGAGTGTATACTGGCTTACTATGTTGGCACTGATGAGGGTGTGAGTGAAGTGCTTCATGTGGCAGGAGAAAAAA
 GGCTGCACCGGTGCGTCAGCAGAATATGTGATACAGGATATATCCGCTTCCTCGCTCACTGACTCGCTACGCTCGGTC
 GTTCGACTGCGGCGAGCGGAAATGGCTTACGAACGGGGCGGAGATTTCTGGAAGATGCCAGGAAGATACTTAACAG
 GGAAGTGAGAGGGCGCGGCAAAGCCGTTTTTCCATAGGCTCCGCCCCCTGACAAGCATCACGAAATCTGACGCTCA
 AATCAGTGGTGGCGAAACCCGACAGGACTATAAAGATACCAGGCGTTTTCCCCCTGGCGGCTCCCTCGTGCCTCTCCT
 GTTCCTGCCTTTTCGGTTTACCGGTGTCTTCCGCTGTTATGGCCGCGTTTGTCTCATTCCACGCCTGACACTCAGTTC
 CGGGTAGGCAGTTTCGCTCCAAGCTGGACTGTATGCACGAACCCCCCGTTGAGTCCGACCGCTGCGCCTTATCCGGTAA
 CTATCGCTTGTAGTCCAACCCGGAAGACATGCAAAAGCACCCTGGCAGCAGCCACTGGTAATTGATTTAGAGGAG
 TTAGTCTTGAAGTCATGCGCCGGTTAAGGCTAAACTGAAAGGACAAGTTTTTGGTGACTGCGCTCCTCCAAGCCAGTT
 ACCTCGGTTCAAAGAGTTGGTAGCTCAGAGAACCCTCGAAAAACCGCCCTGCAAGGCGGTTTTTTTCGTTTTTCAGAGC
 AAGAGATTACGCGCAGACCAAAACGATCTCAAGAAGATCATCTTATTAAGGGGTCTGACGCTCAGTGGAACGAAACC
 CTGTTGGTCAGTTTTTCGGGAGGTGTGCGTCTCGCATCCGGAAGGTGTGATAGGTAGCGCAGCAAAATAAACGAAA
 GGCTCAGTCGAAAGACTGGGCCTTTCGTTTTATCTGTTGTTTGTGCGGTGAACGCTCTCCTCAACGAAAAATATTTTTT
 CAAAAGTATCGTTTACCGCTAGCTCAGTCCTAGGTACAATNNNNNNNNNNNNNNNNNTGGCAGCTGTCACCGGAT
 GTGCTTTCCGGTCTGATGAGTCCGTGAGGACGAAACAGCCTCTACAAATAATTTTGTTTAAGGGCCCAAGTTCACTTA
 AAAAGGAGATCAACAATGAAAGCAATTTTCGTACTGAAACATCTTAATCATGCGGTAAGGAGTTAAATATGTCGAAG
 GGAGAGGAATTTGTTTACTGGAGTCGTCCCAATTCTTGTTGAGTTGGACGGAGATGTCAATGGTCATAAGTTTAGCGT
 ATCCGGTGAGGGTGAGGGAGATGCTACTTACGGAAAAATTAACATTAAAATTCATTTGTACCACAGGCAAACTTCCTGT
 TCCTTGGCCAACTCTTGTTGACGACTTTTGGCTACGGTGTGAGTGTTCGCTCGTTATCCTGACCACATGAAGCAACA
 TGATTTCTTCAAGTCTGCGATGCCTGAAGGATATGTTCAAGAACGTACCATCTTCTTCAAAGATGATGGTAACTATAA
 GACTCGTGCAGAGGTAATAATTCGAGGGAGATACGTTGGTAAATCGCATTGAGCTGAAGGGAATCGATTTCAAGGAAG
 ATGGAACATTTCTGGGACACAAGCTGGAGTACAATTACAATAGCCATAACGTCTATATCATGGCAGATAAGCAAAAAA
 ACGGAATCAAGTTAACTTCAAGATTCGCCATAATATTGAGGACGGCTCTGTGCAATTGGCGGATCATTATCAACAGA
 ACACCCCGATTGGAGATGGTCCCGTGCTGCTGCCAGATAATCACTATTTGTCTACACAATCGGCCCTTTCCAAAGATC
 CGAACGAAAAGCGGATCATATGGTACTTTTAGAGTTGCTAACTGCCGTGGAATCACCCACGGCATGGATGAGTTGT
 ATAAGTAATAATCC

p15A origin	Kanamycin resistance gene
GFP	Promoter
Predicted operator	RiboJ Terminator

Supplementary Table 3. Full plasmid sequence of the reporter vector

AAAACTCACGTTAAGGCCCTCTCCAAGACCGAGCCATCAACAAAGCGTCTCGCTGAGGTTTCATGAGCCTCTGGTTTCATCTCCGG
 CAATTAAAAAAGCGCTAACACAGCCGCTTTTTTACGTCTGCAGGAACGGGCTGTCGACCTTTGAAAAGTTGTTTTACCGCTAGC
 TCAGTCCTAGGTACAATNNNNNNNNNNNNNNNNNTGGCAGCTGTCACCGGATGTGCTTTCCGGTCTGATGAGTCCGTGAGGAC
 GAAACAGCCTCTACAAATAATTTTGTTTAAGGGCCCAAGTTCACTTAAAAAGGAGATCAACAATGAAAGCAATTTTCGTACTGAAA
 CATCTTAATCATGCGGAGGCAGCTTAAATATGTCTACTAAGAACAAGATTGACAAGAATCTGATCAACATCATCGAGGAGATCC
 TGCTGAACGATGAATTAGTGGCTTGAGCATTTCGTAAAGTAGCCACTAAAGCTAATATCTCAATCGGTGGTGTTCAGTATATCTTCG
 GCAATAAGGAAGGGATGATCAAAGCGGTCTTGAAAAAACGAAGAGGATTACAACCGCCAGATTAAAATTCCTCTGAAGGATGA

CAATAGCAAATACTCCCAGTTGAAGGCGCATATCGAGTATATTTTGAAGCACAACGATAATGAGGAGTTCGACAAAACTCTGAAGA
TTATTACGATTCTGCTCCAAGAGAAAATTCGTGTTTGAAGGTCTGCAAGATTGGTACAGTACGTCTTTAAATAGCATTGACACCAATA
CCGATGAGGGCAAAAAATTGCGGCTGGCATTCTGTCTCCGAAGCCATGTTTACACTGATGACCTTAAAGTATATCAACATTAGCC
CGAAGGAACAGAAGGAGATTTTGAAGATCTGAAAACCTTTTACTTTAATAATCTATCGCCACTTTACGCCAAAAAACTTAAGA
CCGCGGTCTTGTCCACTACCTTGCAAGTAATGCGGTGGACAGGATCGGCGGTTTCTTTTCTCTTCTCAACACCTTTCGCGTCAACA
CTTTTCCGCCAAGGAGACGGTTGGTCAGGTTTTCGGGAGGTGTGGCTGGAAGTTCCATATCTTAGAGAAATAGGAACCTCTTTT
TAAATACATTCAAATATGTATCCGCTCATGAGACAATAACCCTGATAAATGCTTCAATAATATTGAAAAAGGAAGAGTATGAGTATT
CAACATTTCCGTGTCGCCCTTATTCCTTTTTCGCGCATTTTGCCTTCTCTGTTTTTGTCTACCCAGAAACGCTGGTGAAAGTAAA
AGATGCTGAAGATCAGTTGGGTGCACGAGTGGGTACATCGAAGTGGATCTCAACAGCGGTAAGATCCTTGAGAGTTTTTCGCCCCG
AAGAACGTTTTTCAATGATGAGCACTTTTAAAGTTCTGCTATGTGGCGCGGTATTATCCCGTGTGACGCCGGGCAAGAGCAACTC
GGTCGCCGCATACACTATTCTCAGAATGACTTGGTTGAGTACTCACCAGTCAACAGAAAAGCATCTTACGGATGGCATGACAGTAAG
AGAATTATGCACTGCTGCCATAACCATGAGTGATAACACTGCGGCCAACTTACTTCTGACAACGATCGGAGGACCGAAGGAGCTAA
CCGCTTTTTTGCACAACATGGGGGATCATGTAACCTGCCTTGATCGTTGGGAACCGGAGCTGAATGAAGCCATACCAAACGACGAG
CGTGACACCACGATGCCTGCAGCAATGGCAACAACGTTGCGCAAACTATTAAGTGGCGAACTACTTACTCTAGCTTCCCGGCAACA
ATTAATAGACTGGATGGAGGCGGATAAAGTTGCAGGACCACTTCTGCGCTCGGCCCTTCCGGCTGGCTGGTTTATTGCTGATAAATC
TGGAGCCGGTGAGCGTGGATCGCGCGGTATCATTTGCAGCACTGGGGCCAGATGGTAAGCCCTCCCGTATCGTAGTTATCTACACGAC
GGGGAGTCAGGCAACTATGGATGAACGAAATAGACAGATCGCTGAGATAGGTGCCTCACTGATTAAGCATTGGTAAGTTGTGATGG
CGGTAGGAATGTAATCGTTAATCCGCAAATAACGTAAAAACCCGCTTCGGCGGGTTTTTTTATGGGGGGAGTTTAGGGAAAGAGCA
TTTGTTCATCCCGTTGAATATGGCTCCCTTAACGTGAGGAAGTTCCCTATACTTTCTAGAGAATAGGAACCTTCTACAGATGGACTTGGG
TTGGCGGTTTTCAGGAGTCTGCAAAACGTCTGCGACCTGAGCAACAACATGAATGGTCATCGGTTTCCGTGTTTCGTAAAGTCTGGA
AACGCGGAAGTCAGCGCCCTGCACCATATGTTCCGGATCTGCATCGCAGGATGCTGCTGGCTACCCTGTGGAACACCTACATCTGT
ATTAACGAAGCGCTGGCATTGACCCTGAGTGATTTTTCTCTGGTCCCGCCGATCCATACCGCCAGTTGTTTACCCTCACAACGTTT
CAGTAACCGGGGATGTTTCATCATCAGTAACCCGTATCGTAGCATCTCTCTCGTTTCATCGGTATCATTACCCCATGAACAGAAA
TCCCCCTTACACGGAGGCATCAGTGACCAACAGGAAAAAACCGCCCTTAACATGGCCCGCTTTATCAGAAGCCAGACATTAACGC
TTCTGGAGAACTCAACGAGCTGGACGCGGTGAACAGGCAGACATCTGTGAATCGCTTACGACACGCTGATGAGCTTTACCGC
AGCTGCCTCGCGCTTTCCGTGATGACGGTGAACACCTCTGACACATGCAGTCCCGCAGACGGTCACAGCTTGTCTGTAAGCGGA
TGCCGGGAGCAGACAAGCCCGTCAGGGCGGTCAGCGGGTGTGGCGGGTGTGCGGGCGCAGCCATGACCCAGTCACGTAGCGATA
GCGGAGTGTATACTGGCTTAACTATGCGGCATCAGAGCAGATTGTACTGAGAGTGCACCGGTGTGAAATACCGCACAGATGCGTAAG
GAGAAAATACCGCATCAGGCGCTCTTCCGCTTCTCGCTCACTGACTCGCTGCGCTCGGTGCTGCGCTGCGGCGAGCGGTATCAGC
TCACTCAAAGGCGGTAATACGGTTATCCACAGAATCAGGGGATAACGCAGGAAAGAACATGTGAGCAAAAGGCCAGCAAAAGGCC
AGGAACCGTAAAAAGGCCGCGTTGCTGGCGTTTTTCCATAGGCTCCGCCCCCTGACGAGCATCACAAAAATCGACGCTCAAGTCA
GAGGTGGCGAAACCCGACAGGACTATAAAGATACCAGCGTTTTCCCTGGAAGCTCCCTCGTGCGCTCTCTGTTCGACCCCTGCC
GCTTACCGGATACCTGTCCGCTTTCTCCCTTCGGGAAGCGTGGCGCTTTCTCATAGCTCACGCTGTAGGTATCTCAGTTCCGGTGT
GGTCGTTTCGCTCCAAGCTGGGCTGTGTGCAGCAACCCCGCTTCAGCCCGACCGCTGCGCTTATCCGGTAACATCGTCTTGAGTC
CAACCCGGTAAGACACGACTTATCGCCACTGGCAGCAGCCACTGGTAACAGGATTAGCAGAGCGAGGTATGTAGGCGGTGCTACAG
AGTTCTTGAAGTGGTGGCTAACTACGGCTACACTAGAAGGACAGTATTTGGTATCTGCGCTCTGCTGAAGCCAGTTACCTTCGGA
AAAAGAGTTGGTAGCTCTTGATCCGGCAACAAACCACCGCTGGTAGCGGTGGTTTTTTTGTGCAAGCAGCAGATTACGCGCAG
AAAAAAGGATCTCAAGAAGATCCTTTGATCTTTTCTACGGGTCTGACGCTCAGTGGAACG

pBR322 + ROP origin

Ampicillin resistance gene

Regulator (BAK as example)

Promoter

Predicted operator

RiboJ

Terminator

Supplementary Table 4. Full plasmid sequence of the regulator expression vector