SUPPLEMENTARY INFORMATION

Supplementary Methods. Regression model implementation details

Although we assume independence of each enzyme's contribution, we also posit that the relationship between enzyme expression and product titer is not necessarily monotonic. Therefore, one natural framework for building the model is through the use of categorical variables that represent the presence or absence of a particular promoter in front of each gene. Thus, using a log-linear model (the training data were skewed towards zero, and we found a log transform of the data improved performance), the product titer t as a function of the promoter-gene combinations is modeled as

$$t = \exp\left(\beta_{00} + \sum_{i\in\{1,\ldots,\#E\}} \sum_{j\in\{1,\ldots,\#P_i\}} \beta_{ij} x_{ij}\right)$$

where β_{ij} are the unknown coefficients of the model, and $x_{ij} = 1$ if the j-th promoter is driving the i-th gene and 0 otherwise. Because only one promoter can be in front of each gene, the independent variables x_{ij} are constrained such that $\sum_{j\in\{1,\ldots,\#P_i\}} x_{ij} = 1$ for the i-th gene. In the case of five genes and five promoters for each gene, $\#E = 5$ and $\#P_1 = \#P_2 = \#P_3 = \#P_4 = \#P_5 = 5$.

For N experimental measurements, we define the vector of response variable (titer) measurements as

$$T = \begin{bmatrix} t^1 \\ \vdots \\ t^N \end{bmatrix}$$

where the superscript notation t^k denotes the measurement from the k-th experiment. Similarly, we define the matrix of promoter combinations as

$$X = \begin{bmatrix} {x_{11}}^1 & \cdots & {x_{\#E\#P_{\#E}}}^1 \\ \vdots & \ddots & \vdots \\ {x_{11}}^N & \cdots & {x_{\#E\#P_{\#E}}}^N \end{bmatrix}$$

where each row represents the genotype of the k-th sample. The vector of unknown coefficients is

$$B = \begin{bmatrix} \beta_{11} \\ \vdots \\ \beta_{\#E\#P_{\#E}} \end{bmatrix}$$

Thus, the model can be succinctly represented as $\log(T) = \beta_{00} + XB$. Because the logarithm of zero is negative infinity, we set entries of T that are zero to 0.5, because this is the smallest amount that we can experimentally measure. To train this model, we obtained $N = 182$ measurements (*i.e.*, ninety-one clones in duplicate).

Identification of the unknown β_{ij} coefficients in the model is challenging because of the high-dimensional nature of the problem. We used the previously described Exterior Derivative Estimator (EDE) method(26) to identify the coefficients of the model because it can better protect against overfitting than traditional methods (for the violacein pathway, using ordinary least squares regression resulted in a model with almost no correlation in the test set: Pearson R-values of -0.01, 0.06, -0.02, and 0.01 for violacein, deoxyviolacein, proviolacein, and prodeoxyviolacein, respectively). EDE protects against overfitting by learning constraints that the data obeys, and then it uses these constraints to reduce the degrees of freedom in the regression. More specifically, the coefficients estimated by EDE are given by

$$\hat{B} = \arg\min_{B} \|\log(T) - XB - \beta_{00}\|^2 + \lambda\|PB\|^2$$

where $P = UU^T$, and U is a matrix whose columns are the m smallest principal components of the covariance matrix $X^T X$. λ and m are tuning parameters that are chosen in a data-driven manner using cross-validation.

In general, the rows of the matrix X form a manifold, and the projection matrix P enforces that the regression coefficients lie close to the manifold formed by X. This methodology is motivated by differential geometry, which says that the exterior derivative of a function on an embedded submanifold lies in the cotangent space(38).

Supplementary Table S1. List of plasmids used in this study.

Plasmid	SynBERC Registry ID	Description	Yeast auxotrophic marker
pJED101	SBa_000896	Yeast cloning vector	Met15
pJED102	SBa_000897	Yeast cloning vector	His3
pJED103	SBa_000898	Yeast cloning vector	Leu2
pJED104	SBa_000899	Yeast cloning vector	Ura3
pAH056	SBa_000900	*pTDH3*-RFP-*tADH1*	Leu2
pAH002	SBa_000901	*pTEF1*-RFP-*tADH1*	Leu2
pAH007	SBa_000902	*pRPL18B*-RFP-*tADH1*	Leu2
pSL030	SBa_000903	*pRNR2*-RFP-*tADH1*	Leu2
pAH005	SBa_000904	*pREV1*-RFP-*tADH1*	Leu2
pAH003	SBa_000905	*pRNR1*-RFP-*tADH1*	Leu2
pAH004	SBa_000906	*pCCW12-RFP-tADH1*	Leu2
pAH006	SBa_000907	*pHTA2*-RFP-*tADH1*	Leu2
pAH008	SBa_000908	*pPSP2*-RFP-*tADH1*	Leu2
pAH009	SBa_000909	*pISW2*-RFP-*tADH1*	Leu2
pAH061	SBa_000910	*pARC18*-RFP-*tADH1*	Leu2
pAH065	SBa_000911	*pTEF2*-RFP-*tADH1*	Leu2
pML234	SBa_000912	*pPGK1*-RFP-*tADH1*	Leu2
pML167	SBa_000913	GibA-*pTDH3*-RFP-*tADH1*-GibB	Leu2
pML168	SBa_000914	GibB-*pTDH3*-YFP-*tADH1*-GibC	His
pML159	SBa_000915	GibC-*pTDH3*-CFP-*tADH1*-GibD	Ura3
pML203	SBa_000916	GibA-GibD vector	Met15
pML223	SBa_000917	GibA-GibD vector (KanR)	Ura3
pML242	SBa_000891	GibA-*pTDH3-vioA-tADH1*-GibC	Leu2
pML243	SBa_000892	GibC-*pTDH3-vioC-tADH1*-GibD	Ura3
pML244	SBa_000893	GibA-*pTDH3-vioB-tADH1*-GibB	Leu2
pML245	SBa_000894	GibB-*pTDH3-vioD-tADH1*-GibC	His3
pML246	SBa_000895	GibC-*pTDH3-vioE-tADH1*-GibD	Ura3
pML256	SBa_000918	*vioAC* overexpression plasmid	Met15
pML258	SBa_000919	*vioBDE* overexpression plasmid (KanR)	Ura3

All plasmids contain a ColE1 *E. coli* replication origin, carry an ampicillin resistance gene (unless otherwise indicated), and contain a CEN6/ARS4 *S. cerevisiae* replication origin. Annotated plasmid sequences can be found at the SynBERC Registry (registry.synberc.org). Sequences of plasmids not listed in this table (*e.g.*, the series of YFP plasmids) can be determined simply by replacing the appropriate genes or promoters.

Supplementary Table S2. List of primers used in this study.

Primer	**Sequence**
vioA cloning forward	gcatAGATCTatgaaacattcttccgatat
vioA cloning reverse	atgcCTCGAGttaGGATCCcgcggcgatacgctgcaaca
vioB cloning forward	gcatAGATCTatgagcattctggatttccc
vioB cloning reverse	atgcCTCGAGtcaGGATCCggcctcgcggctcagtttgc
vioC cloning forward	gcatAGATCTatgaaacgtgcgattatcgt
vioC cloning reverse	atgcCTCGAGtcaGGATCCattcacgcgaccaatcttgt
vioD cloning forward	gcatAGATCTatgaagattctggtcattgg
vioD cloning reverse	atgcCTCGAGtcaGGATCCgcgctgcaaagcataacgca
vioE cloning forward	gcatAGATCTatggagaaccgtgagccacc
vioE cloning reverse	atgcCTCGAGtcaGGATCCgcgcttggccgcgaaaaccg
Gibson A amplification forward	ggtacagacactgcgacaac
Gibson A amplification reverse	gtattgcgacgaattgccacgttgtcg
Gibson B amplification forward	gggtcatcacggctcatc
Gibson B amplification reverse	agctgtgttgacatctggc
Gibson C amplification forward	ggtgatccgctgactcct
Gibson C amplification reverse	ggctcacgtcttatttgggc
Gibson D amplification forward	cacaaggtcagggcactcatgcgac
Gibson D amplification reverse	tgcatcgagttgattgtcgc
Gibson A TRAC forward	gccgataattgcagacg
Gibson B TRAC forward	ccagatgtcaacacagctac
Gibson C TRAC forward	acacactggcttaaggagac
vioA TRAC reverse	caatgcagatatcggaagaatg
vioB TRAC reverse	aagtggatacgcgggaaatc
vioC TRAC reverse	gacgtgcacttcgtagcc
vioD TRAC reverse	gtcattcttctccacgatgtca
vioE TRAC reverse	tcgggctccaataagagacata

Supplementary Table S3. Sequences of TRAC duplex probes.

	Dye strand (5′-3′)	Quencher strand (5′-3′)
pTDH3	[6-FAM]-ACACAAGGCAATTGACCCACG-(P)	TGGGTCAATTGCCTTGTGT-[IABkFQ]
pTEF1	[Cy3]-ACAACAGAAAGCGACCACCCAAC-(P)	GGTGGTCGCTTTCTGTTGT-[IABkFQ]
pRPL18B	[Cy5.5]-TCACGCCCAAGAAATCAGGC-(P)	CTGATTTCTTGGGCGTGA-[IAbRQSp]
pRNR2	[6-ROXN]-AAGCACGGGCAGATAGCACC-(P)	GCTATCTGCCCGTGCTT-[IAbRQSp]
pREV1	[Cy5]-ATGCCGCGTTCACAGATTCC-(P)	CTGTGAACGCGGCAT-[IAbRQSp]

Dye strands are labeled on their 5′ end with a fluorescent dye, indicated in brackets, and on their 3′ end with a phosphate (P). Quencher strands are labeled on their 3′ end with Iowa Black® FQ or RQ quenchers, indicated in brackets.

Supplementary Table S4. Recombination rates of tandem expression plasmids.

A	RFP loss	YFP loss	intact
pTDH3	0	0	48
pTEF1	0	1	47
pRPL18B	0	0	48
pRNR2	1	0	47

B	RFP loss	YFP loss	intact
pTDH3	0	0	48
pTEF1	3	0	45
pRPL18B	0	0	48
pRNR2	0	0	48

C	RFP loss	YFP loss	CFP loss	intact
pTDH3	1	0	0	47
pTEF1	1	0	0	47
pRPL18B	0	0	0	48
pRNR2	0	1	0	47

Plasmids containing tandem expression cassettes of RFP and YFP (**A**), YFP and RFP (**B**), or RFP, YFP, and CFP (**C**) were transformed into yeast, and forty-eight colonies were picked for each construct. *pREV1* was omitted due to its low signal over background making it difficult to discern a loss of fluorescence.

Supplementary Table S5. “TRAC barcode” results.

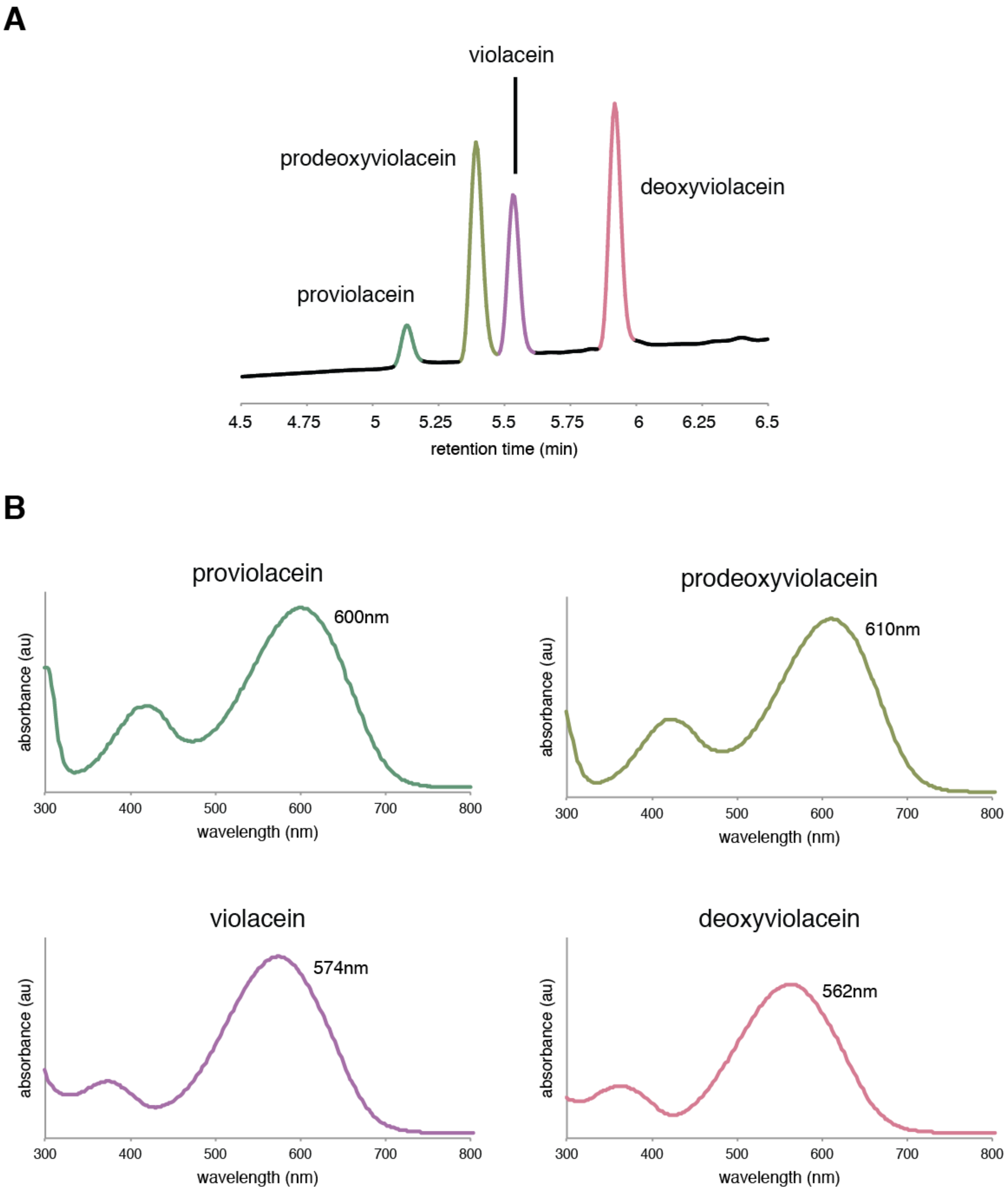
Barcode	*pTDH3*	*pTEF1*	*pRPL18B*	*pRNR2*	*pREV1*
1	327	69	144	97	410
2	368	78	174	113	176
3	350	69	189	39	441
4	346	78	183	38	145
5	307	69	47	114	456
6	332	71	51	116	158
7	351	74	54	37	467
8	357	77	39	36	144
9	358	27	188	113	535
10	358	26	170	113	195
11	343	24	156	39	416
12	392	23	151	47	195
13	330	25	43	98	457
14	444	21	42	140	188
15	377	19	43	38	514
16	302	21	66	34	137
17	127	64	109	93	464
18	129	79	186	111	177
19	122	73	186	50	460
20	120	86	189	41	190
21	131	79	31	123	607
22	128	74	58	124	167
23	121	71	72	34	488
24	122	70	49	41	175
25	123	22	155	95	429
26	123	21	219	123	147
27	118	23	150	48	577
28	124	21	177	42	159
29	122	24	67	136	565
30	117	21	45	95	150
31	127	19	58	36	395
32	123	18	49	39	149

Thirty-two possible barcode sequences were cloned and used as templates for a TRAC reaction. Boxes outlined in red indicate expected positive probe targets; boxes shaded in blue indicate positive signals identified by TRAC. For example, Barcode #4 has sequences complementary to the *pTDH3*, *pTEF1*, and *pRPL18B* probes, but non-complementary to *pRNR2* and *pREV1*, and a corresponding TRAC reaction only had fluorescence for the three complementary probes. TRAC successfully identified all thirty-two unique barcode sequences.

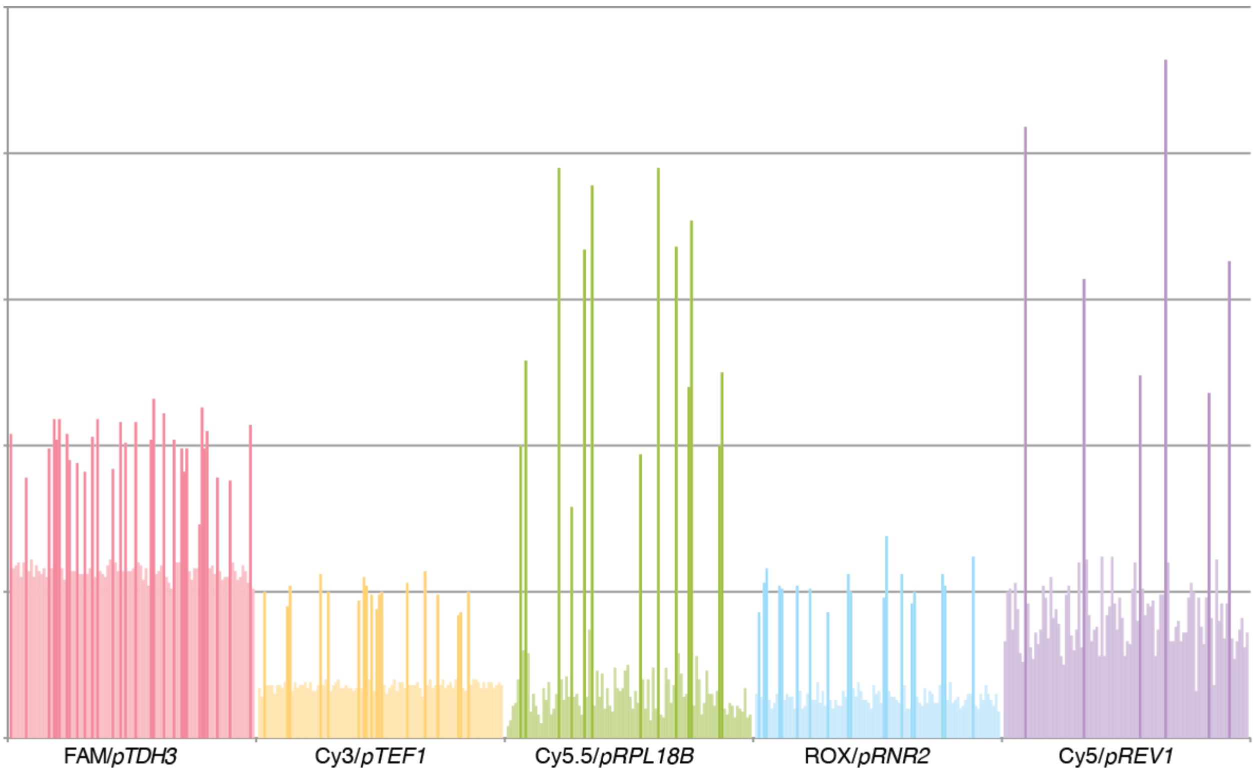
Supplementary Table S6. Strains with directed flux raw data.

Strain	Violacein	Deoxyviolacein	Proviolacein	Prodeoxyviolacein
V1	141.0 ± 14.5	8.1 ± 1.1	46.5 ± 8.5	8.7 ± 0.4
V2	81.5 ± 12.1	4.9 ± 0.7	41.3 ± 3.0	7.0 ± 0.5
V3	80.1 ± 8.1	4.1 ± 0.2	45.0 ± 6.3	6.9 ± 0.4
V4	25.4 ± 11.4	22.2 ± 10.8	14.1 ± 5.1	25.4 ± 8.6
V5	90.5 ± 18.5	12.3 ± 1.9	34.8 ± 3.3	12.5 ± 1.3
DV1	25.4 ± 11.4	22.2 ± 10.8	14.1 ± 5.1	25.4 ± 8.6
DV2	0 ± 0	32.5 ± 23.2	0 ± 0	38.7 ± 12.8
DV3	28.6 ± 4.4	35.5 ± 4.4	13.7 ± 3.3	31.7 ± 5.3
DV4	0 ± 0	35.6 ± 23.9	0 ± 0	47.5 ± 14.3
DV5	19.4 ± 1.6	11.9 ± 1.6	17.6 ± 2.3	26.2 ± 2.5
PV1	4.4 ± 1.7	0.8 ± 1.5	97.8 ± 5.0	11.8 ± 1.2
PV2	4.1 ± 2.4	1.3 ± 1.5	88.0 ± 10.5	10.2 ± 0.6
PV3	0 ± 0	0.7 ± 1.4	74.6 ± 6.5	11.1 ± 1.5
PV4	0 ± 0	3.6 ± 0.7	106.7 ± 5.6	15.0 ± 2.4
PV5	6.7 ± 1.4	3.1 ± 2.7	77.8 ± 52.2	6.7 ± 7.7
PDV1	0 ± 0	0 ± 0	0.9 ± 1.1	78.4 ± 34.6
PDV2	0 ± 0	0 ± 0	0 ± 0	79.3 ± 11.4
PDV3	0 ± 0	0 ± 0	2.3 ± 0.2	82.7 ± 2.6
PDV4	0 ± 0	0.8 ± 1.7	1.0 ± 1.2	73.2 ± 13.0
PDV5	0 ± 0	0 ± 0	8.7 ± 1.0	94.6 ± 15.9

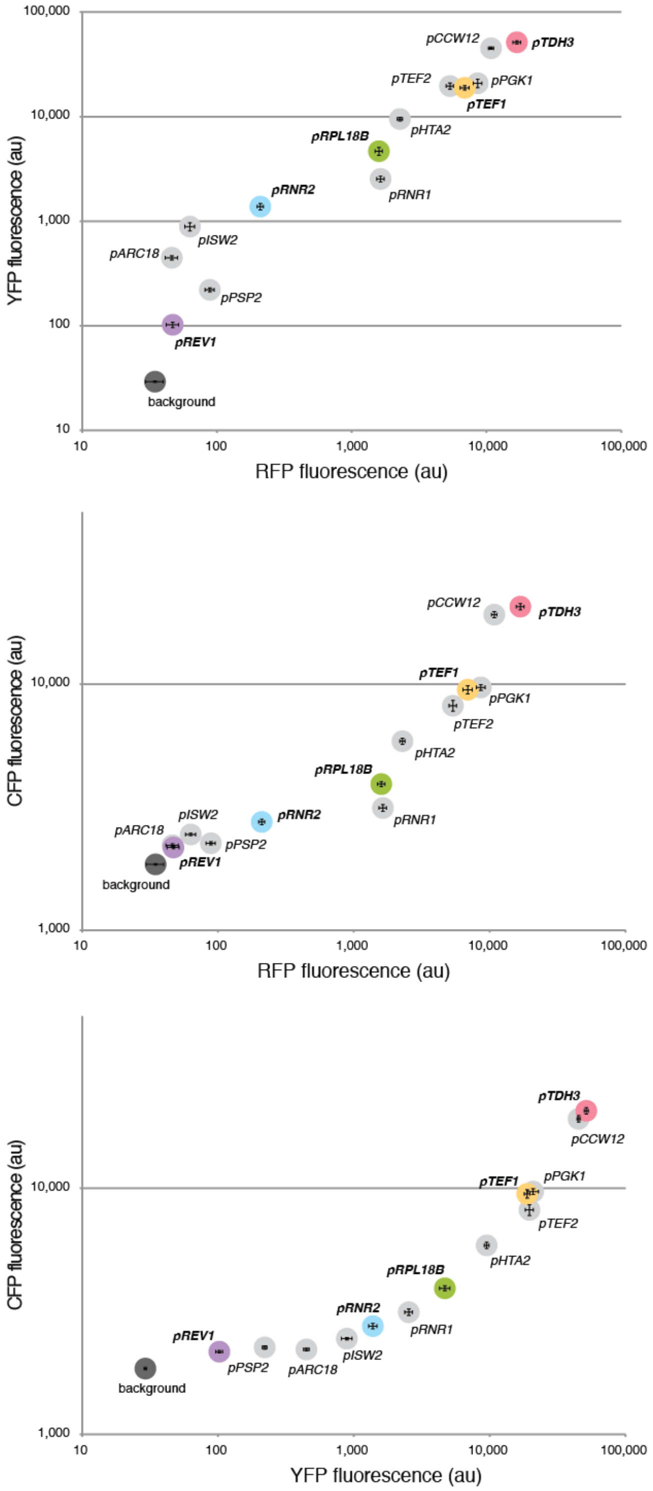
Raw data represented in **Figure 5**. Strains designated V#, DV#, PV#, and PDV# are strains predicted by the model to produce high amounts of violacein, deoxyviolacein, proviolacein, and prodeoxyviolacein, respectively. Values shown are the average titer (as measured by HPLC peak area) of four biological replicates and the standard deviation.



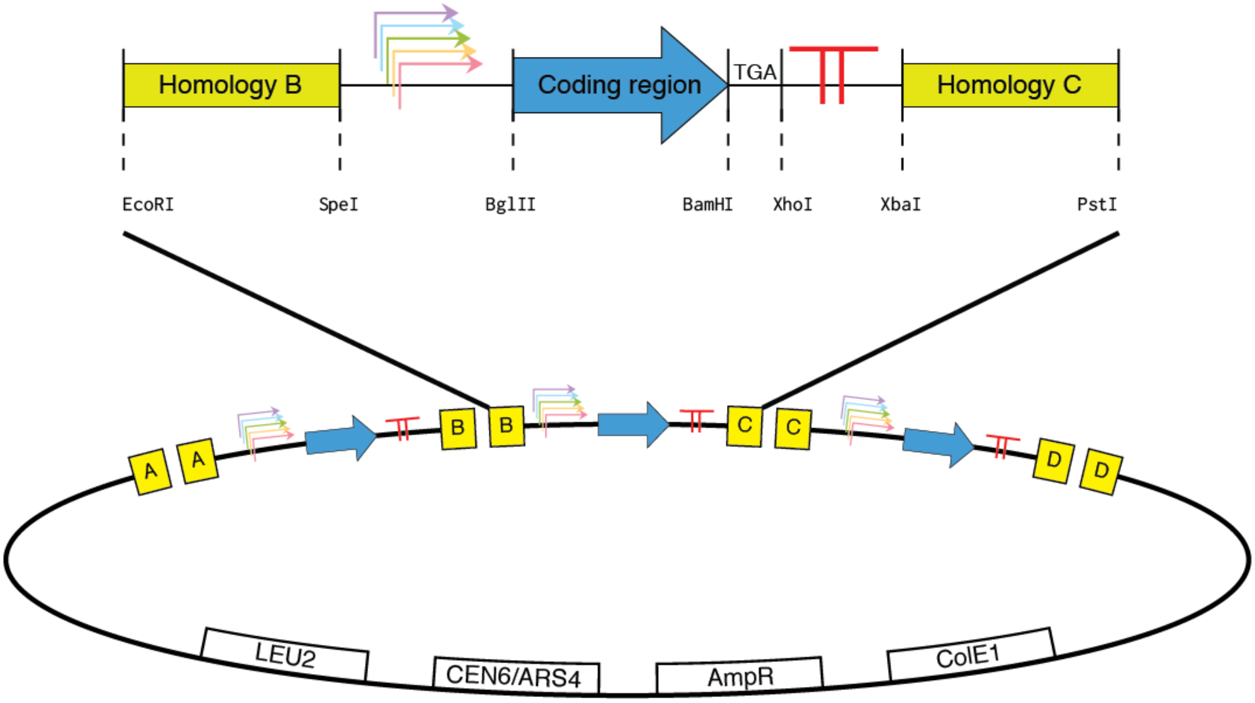
Supplementary Figure S1. Chromatogram and absorbance spectra of violacein extractions. A. Chromatogram for absorbance at 565nm. **B.** Absorbance spectra for the four main products. Maximum absorbance wavelength is indicated.



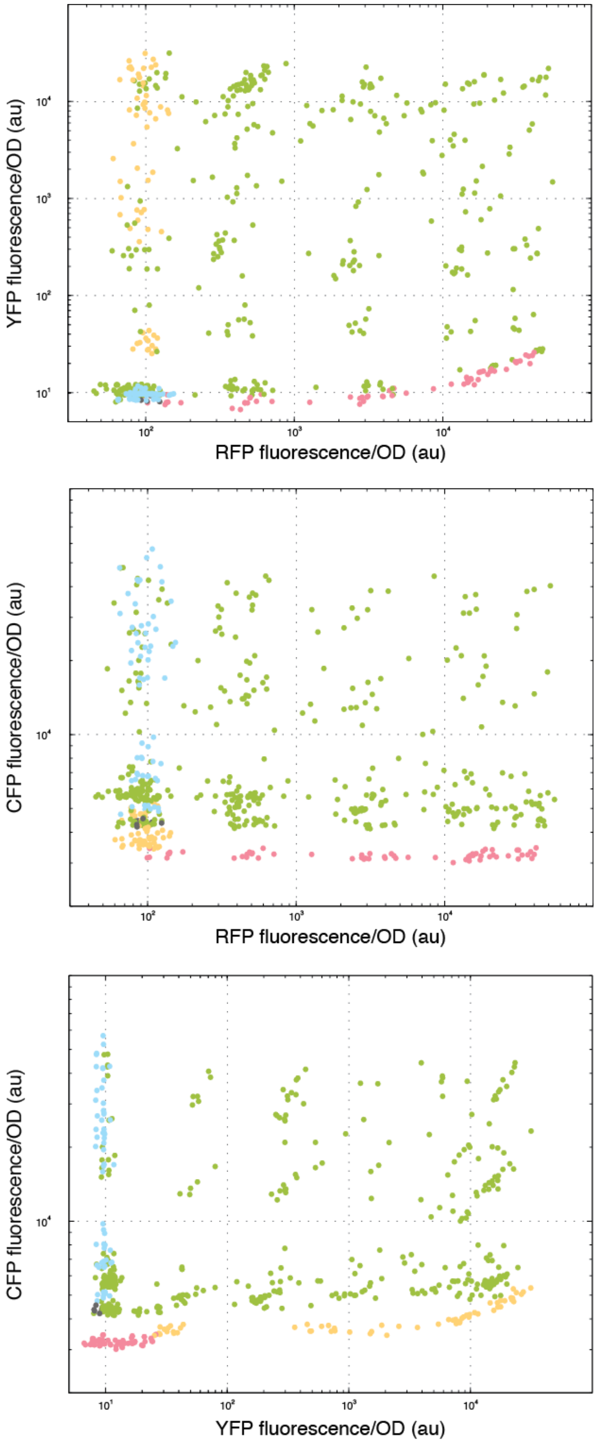
Supplementary Figure S2. Raw fluorescence data from a typical TRAC reaction plate. Each promoter is associated with a unique fluorescent dye, which is released and fluoresces after PCR amplification. A set of ninety-six library clones screened is shown, with each of the five wavelengths. Correctly assembled clonal isolates should fluoresce at precisely a single wavelength.



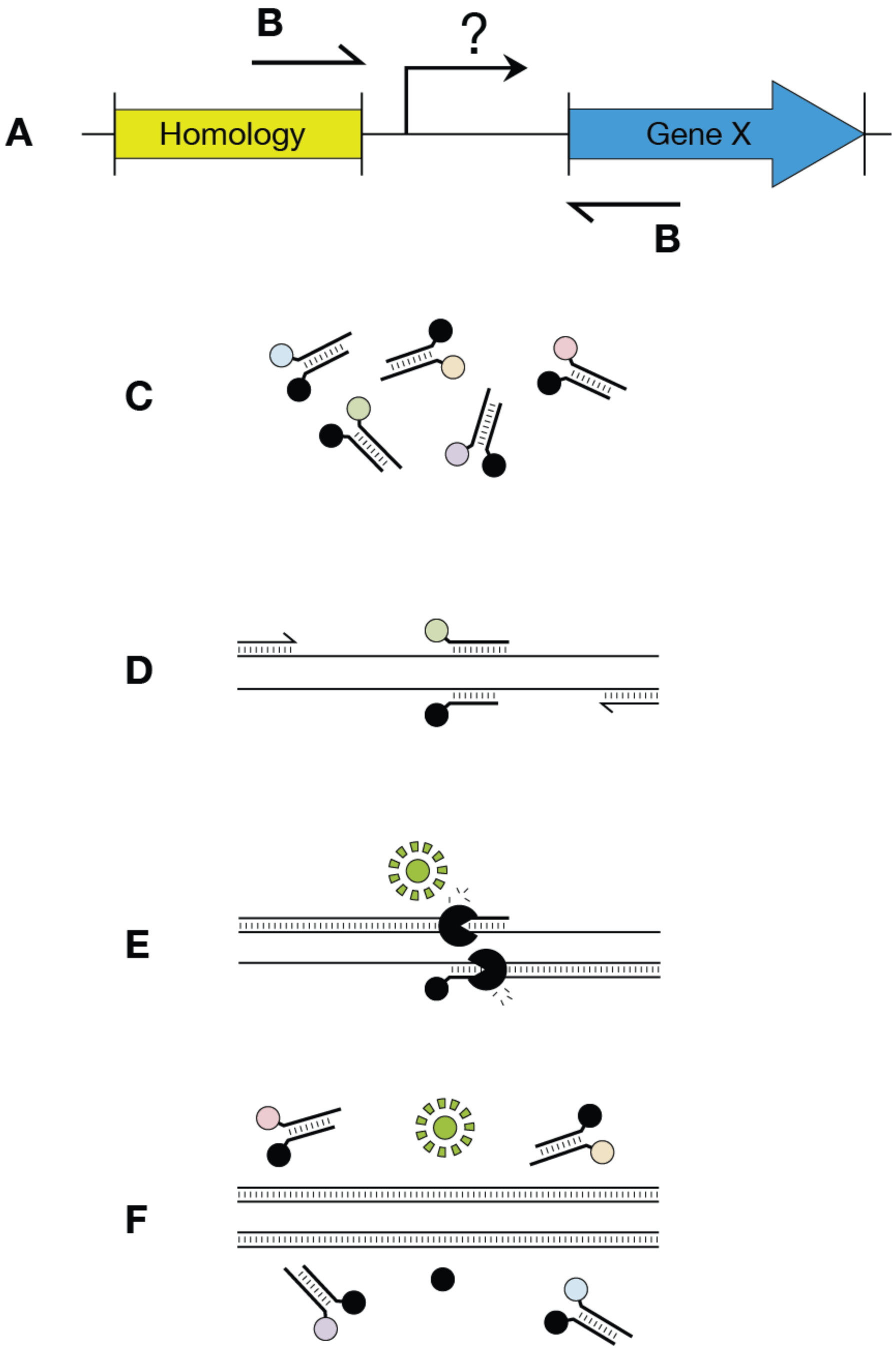
Supplementary Figure S3. Characterization of yeast constitutive promoters. Thirteen promoters cloned from the yeast genome driving expression of RFP, YFP, and CFP. The five promoters used in all subsequent experiments are colored. Error bars in all plots indicate s.e.m. *n*=8.



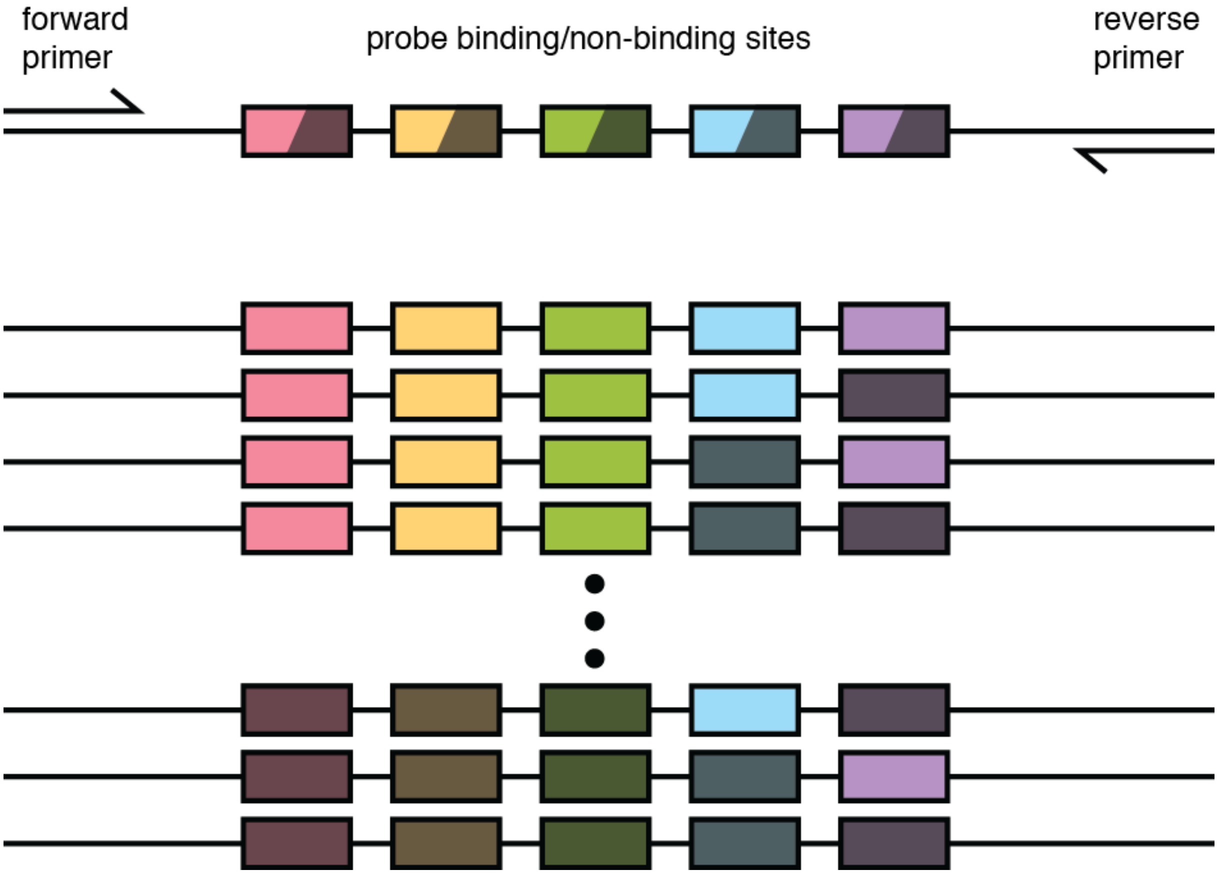
Supplementary Figure S4. Gibson assembly of multi-gene constructs. Expression cassettes comprising of a promoter (library), gene, and transcriptional terminator are flanked by unique DNA homology sequences. Homology allows for specific assembly of multiple (shown here, three) cassettes into a recipient vector backbone.



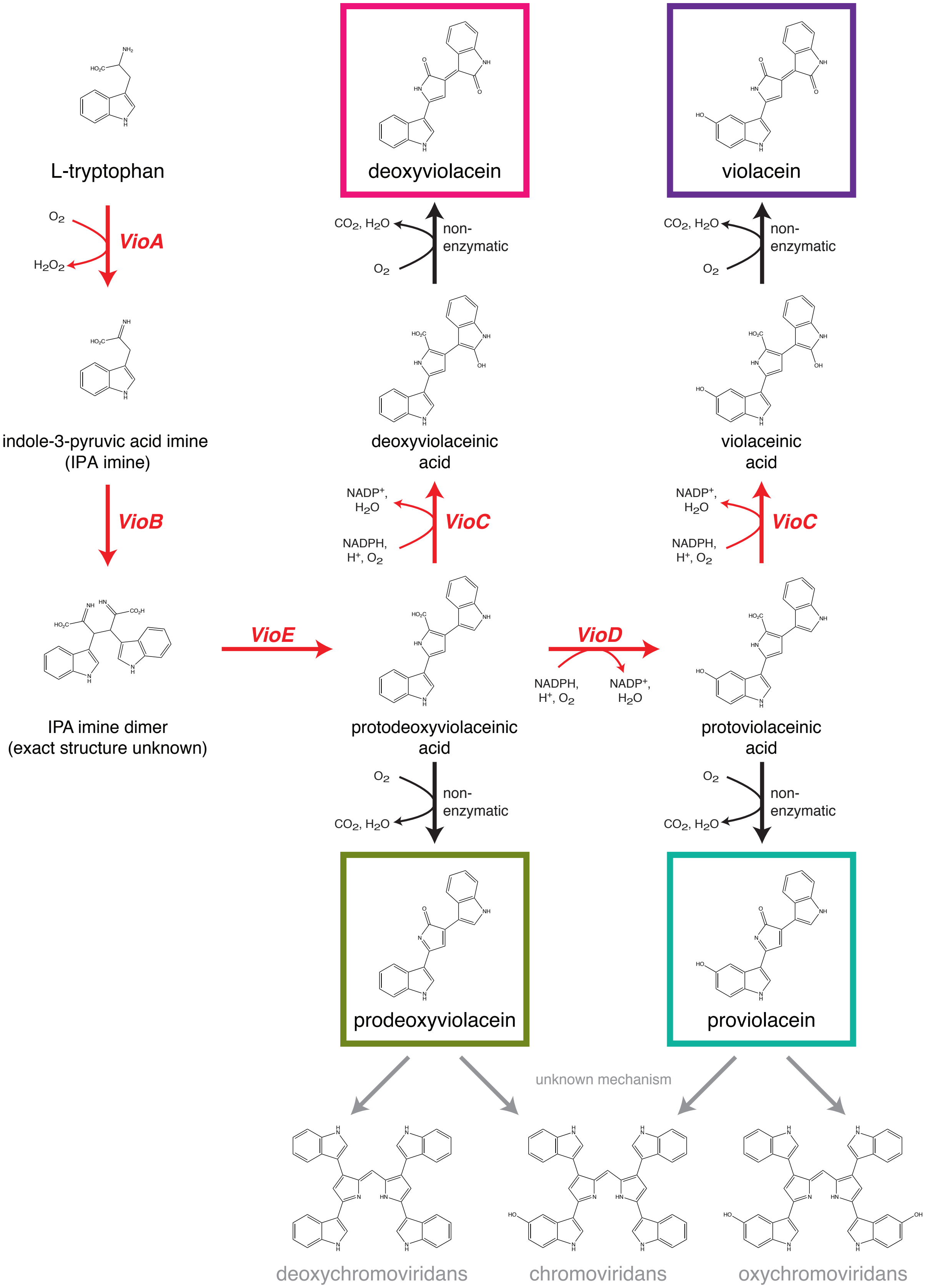
Supplementary Figure S5. Combinatorial assembly of a fluorescent protein library. Two-dimensional projections of the data shown in **Figure 1D**. RFP-only library (red dots, LEU2), YFP-only library (yellow dots, HIS3), CFP-only library (blue dots, URA3), triple FP library (green dots, MET15), empty vector (black dots, MET15). *N.b.*, the individual libraries and triple library are expressed from plasmids carrying different auxotrophic markers, which may contribute to the lower baseline CFP fluorescence observed for the RFP and YFP libraries.



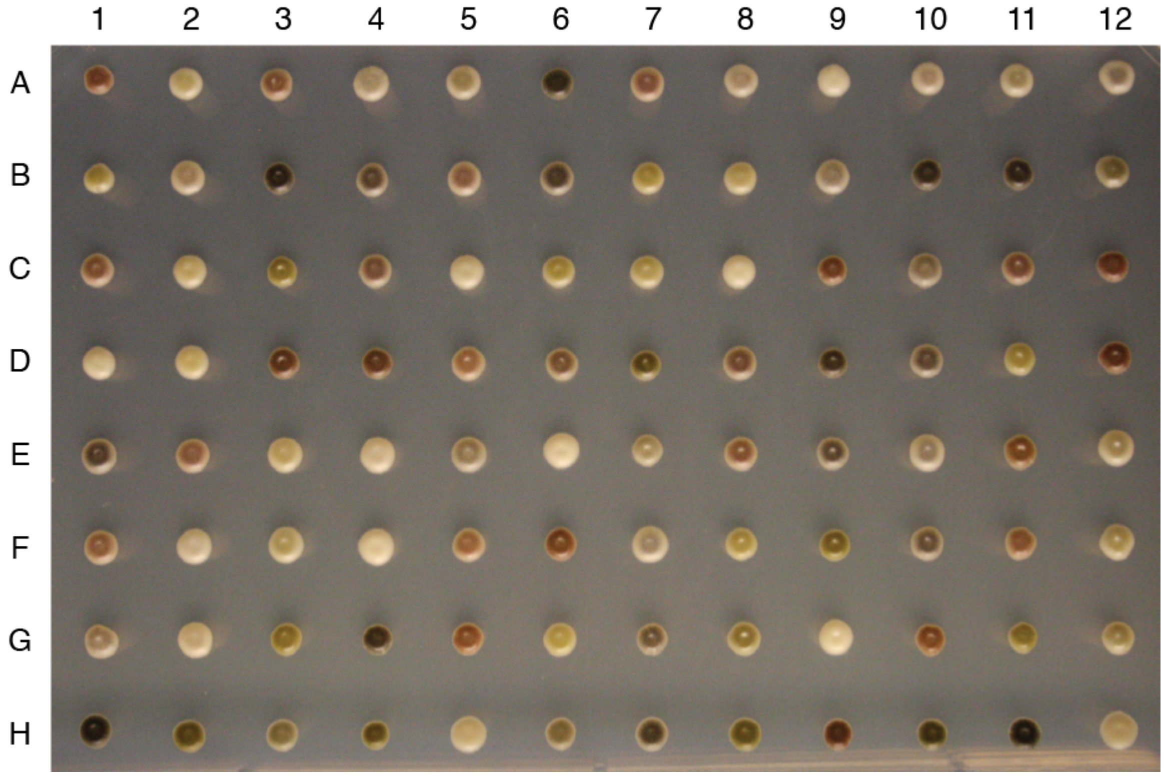
Supplementary Figure S6. TaqMan Rapid Analysis of Combinatorial assemblies. Schematic for how TRAC can detect the genotype of an unknown promoter in a single PCR reaction and fluorescent measurement. **A.** A template (colony) with an unknown promoter at position X. **B.** Gene-specific primers amplify the promoter region. **C.** Duplex probes are included in the amplification reaction. **D.** Specific probes anneal to the amplified DNA. **E.** 5′-3′ exonuclease activity of Taq DNA polymerase cleaves annealed probes. **F.** Only the single probe specific for the promoter is released, resulting in a single fluorescent signal.



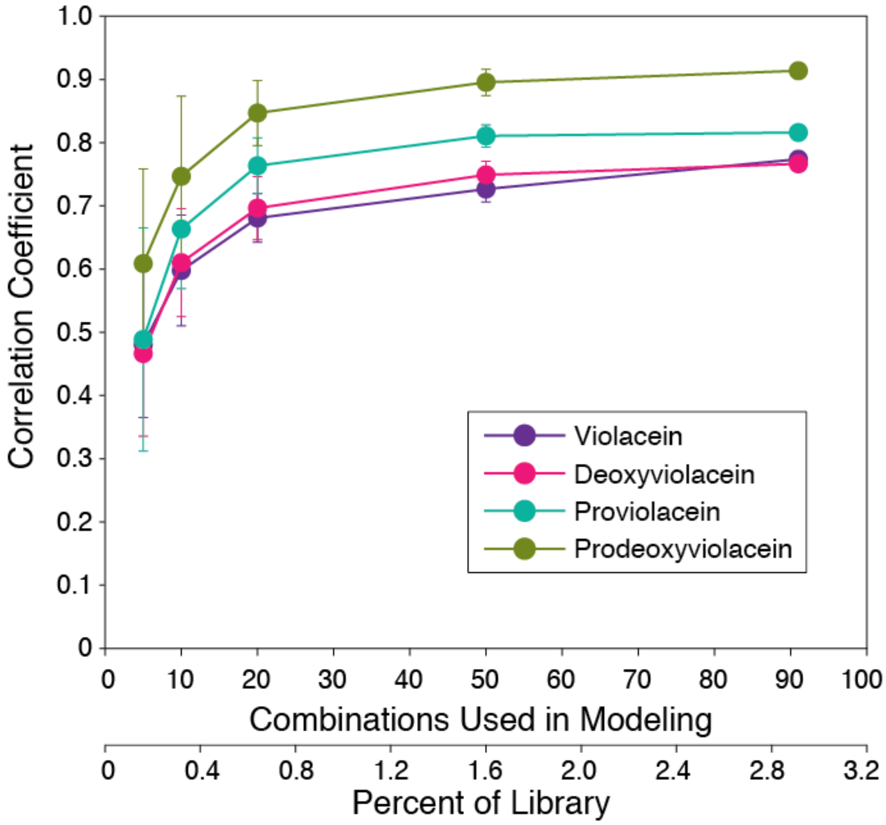
Supplementary Figure S7. "TRAC barcode" design. Barcodes were cloned to include either a complementary or non-complementary sequence for all five TRAC probes ($2^5 =$ 32 possible sequences) and flanking PCR primer binding sites. When a TRAC reaction was performed, combinations of zero to five fluorescent dyes were cleaved depending on whether the complementary sequence for a particular probe was present in the template. Fluorescence was measured on a plate reader as per a typical TRAC reaction, and all thirty-two unique barcodes were successfully identified.



Supplementary Figure S8. Chemical structures of the violacein biosynthetic pathway.



Supplementary Figure S9. Violacein biosynthesis expression library. Ninety-six unique clones from a combinatorial expression library of the violacein biosynthetic pathway. The first ninety-one (A01-H07) were used as training data for the regression model. The last five are controls containing *pTDH3* driving: *vioABE* (H08), *vioABEC* (H09), *vioABED* (H10), *vioABEDC* (H11), empty vectors (H12).



Supplementary Figure S10. Effect of training set size on model accuracy. The original training set included ninety-one combinations; random subsets (of size five, ten, twenty, fifty, and the full ninety-one) were used to retrain the model, and the mean correlation coefficient of one hundred trials is shown. Larger training sets improve correlation, although the marginal benefit decreases as the number of samples increases. Error bars indicate s.d., *n*=100.