

Supplementary information for:

## Design of orthogonal genetic switches based on a crosstalk map of $\sigma$ s, anti- $\sigma$ s, and promoters

Virgil A. Rhodius, Thomas H. Segall-Shapiro, Brian D. Sharon, Amar Ghodasara, Ekaterina Orlova, Hannah Tabakh, David H. Burkhardt, Kevin Clancy, Todd C. Peterson, Carol A. Gross, and Christopher A. Voigt

### I. Promoter Identification and Design

- I.A. *Identification of ECF promoters using genomic information*
- I.B. *ECF  $\sigma$  promoter modeling*
- I.C. *Predicted orthogonality of the complete promoter, as well as the -10 and -35 regions*
- I.D.  *$\sigma$ 70 promoter modeling and prediction*
- I.F. *Improving promoter response with synthetic UP elements*

### II. $\sigma$ and Anti- $\sigma$ Library Characterization

- II.A. *Selecting ECF  $\sigma$ s, anti- $\sigma$ s and promoters*
- II.B. *Complete  $\sigma$  screening data, including multiple  $\sigma$ s from each subgroup and non-orthogonal data*
- II.C. *Full transfer functions and cytometry data for promoter induction.*
- II.D. *Complete anti- $\sigma$  screening data*
- II.E. *Growth assays of  $\sigma$  and anti- $\sigma$  expression*
- II.F. *Simultaneous expression of multiple  $\sigma$ s*
- II.G. *Expression of a  $\sigma$  in *Klebsiella oxytoca**
- II.H. *Characterization of induction systems*

### III. Quantification of Anti- $\sigma$ Threshold Control

### IV. Creating Chimeric $\sigma$ s

- IV.A. *Design of chimeric  $\sigma$ s and promoters*
- IV. B. *Chimeric  $\sigma$  characterization*

### V. Plasmids

### VI. Supplementary Tables

### VII. References

## I. Promoter Identification and Design

### I.A. Identification of ECF promoters using genomic information

A three-step search strategy was used to maximize the identification of promoter sequences for each ECF  $\sigma$  subgroup and increase the number of promoters identified compared to the original work by Staron and co-workers. First, based on the observations that many ECF  $\sigma$  groups autoregulate their own gene expression<sup>1</sup> and that some  $\sigma$ s may regulate their own anti- $\sigma$ <sup>2</sup>, promoter motifs were searched for in sequences directly upstream of the  $\sigma$  gene,  $\sigma$  operon, and cognate anti- $\sigma$  gene. Second, upstream regulatory regions were extracted for all  $\sigma$ s within each subgroup to maximize the ability to find over-represented motifs. Third, BioProspector<sup>3</sup> was used to identify over-represented motifs in these upstream regulatory regions. BioProspector is a 2-block motif search algorithm that is ideally suited for bacterial promoters with variable length spacers between the -10 and -35 motifs.

All ECF  $\sigma$ s in subgroups 01-43 and their cognate anti- $\sigma$ s were identified from Staron and co-workers (their Table S5)<sup>1</sup>. To enable efficient retrieval of their upstream regulatory sequences, all 1232 complete bacterial genome sequences and annotations were downloaded from the NCBI FTP site (11/1/2010). Both  $\sigma$ s and anti- $\sigma$ s were identified from these genomes based on the annotation supplied by Staron *et al.*: source genome, gene ID (GI) and their listed amino acid sequence ( $\sigma$ s only). From the 1736 listed ECF  $\sigma$ s and cognate 1203 anti- $\sigma$ s listed by Staron *et al.*, 1329  $\sigma$ s and 880 anti- $\sigma$ s were successfully identified from the NCBI annotated genomes. The remaining  $\sigma$ s and anti- $\sigma$ s were from genomes not listed in the NCBI database and therefore were not used in this analysis.

For each ECF  $\sigma$  subgroup, three libraries of upstream regulatory sequences were extracted from: 1) directly upstream of the  $\sigma$  gene; 2) directly upstream of the  $\sigma$  gene operon ( $\sigma$  operons were defined as all consecutive genes adjacent to the  $\sigma$  gene, in the same orientation and separated by less than 50 nt from each other); 3) directly upstream of the cognate anti- $\sigma$  gene (if known). Most promoters occur near the start of genes but can be difficult to detect when searching long upstream regulatory sequences for over-represented motifs. To facilitate identification, different length upstream regulatory sequences were extracted for each library, from the start codon to 100, 150, 200 and 300 nt upstream. For each library, searches for over-represented motifs were performed using BioProspector with the short 100 nt upstream sequences first and then repeated with the successively longer sequences. Motif searches with BioProspector were performed only on the forward strand and the highest scoring motifs selected from 100 reinitializations. The search for 2-block motifs was typically of the form,  $W7 w5 G18 g15$ : where  $W$  and  $w$  denotes the length (nt) of the upstream and downstream blocks, respectively; and  $G$  and  $g$  denotes the maximum and minimum distances (nt) separating the two blocks, respectively. These parameters were varied iteratively to optimize the searches for different promoter motifs. From all the library, sequence and motif search combinations, the highest scoring 2-block motif was selected as the representative promoter motif for each ECF  $\sigma$ . These were typically from the 100 or 200 nt sequences upstream of the ECF  $\sigma$  gene or operon.

Promoters for ECF subgroups 05-10, 19, 27 and 32 listed in Staron *et al.* were not identified in our search. Subgroups 05-10 are not autoregulated<sup>1</sup> and the remaining subgroups only had a few  $\sigma$ s with highly related upstream sequences, making it difficult to search for over-represented motifs. For all of these cases, the promoter sequences were obtained from Staron *et al.* and BioProspector was used to redefine the -35 and -10 motifs. Promoter sequences and their -10/-35 motifs are listed in Supplementary Table S1.

### I.B. ECF $\sigma$ promoter modeling and prediction

For each ECF  $\sigma$  subgroup, the highest scoring 2-block motif identified by BioProspector was used to construct promoter models following the method described by Rhodius and Mutalik<sup>4</sup>. The upstream and downstream motif sequences were used to compile Position Weight Matrices (PWMs)<sup>5</sup> for the -35 and -10 motifs, respectively. Specifically, for the regions identified by BioProspector, the weights ( $W_{b,i}$ ) for each position ( $i$ ) and base ( $b$ ) were computed as

$$W_{b,i}^{motif} = \ln \left[ \frac{(n_{b,i} + 0.5)/(N + 2.0)}{P_b} \right] , \quad (S1)$$

where  $n_{b,i}$  is the number of times that the base  $b$  is found at position  $i$  in the promoter set,  $N$  is the number of promoters in the promoter set, and  $P_b$  is the probability of finding a specific base at any given position (assumed to be 0.25). Bayesian pseudocounts of 0.5 were added to each base to represent the relative uncertainty in the promoter sequences. To evaluate a motif in a promoter, the appropriate weights can be summed for a given sequence of bases  $b$  at positions  $i$  to obtain a complete -35 or -10 score. Additionally, the variable distances between the -35 and -10 motifs were used to construct spacer length histograms and to calculate a penalty score for suboptimal spacer lengths,

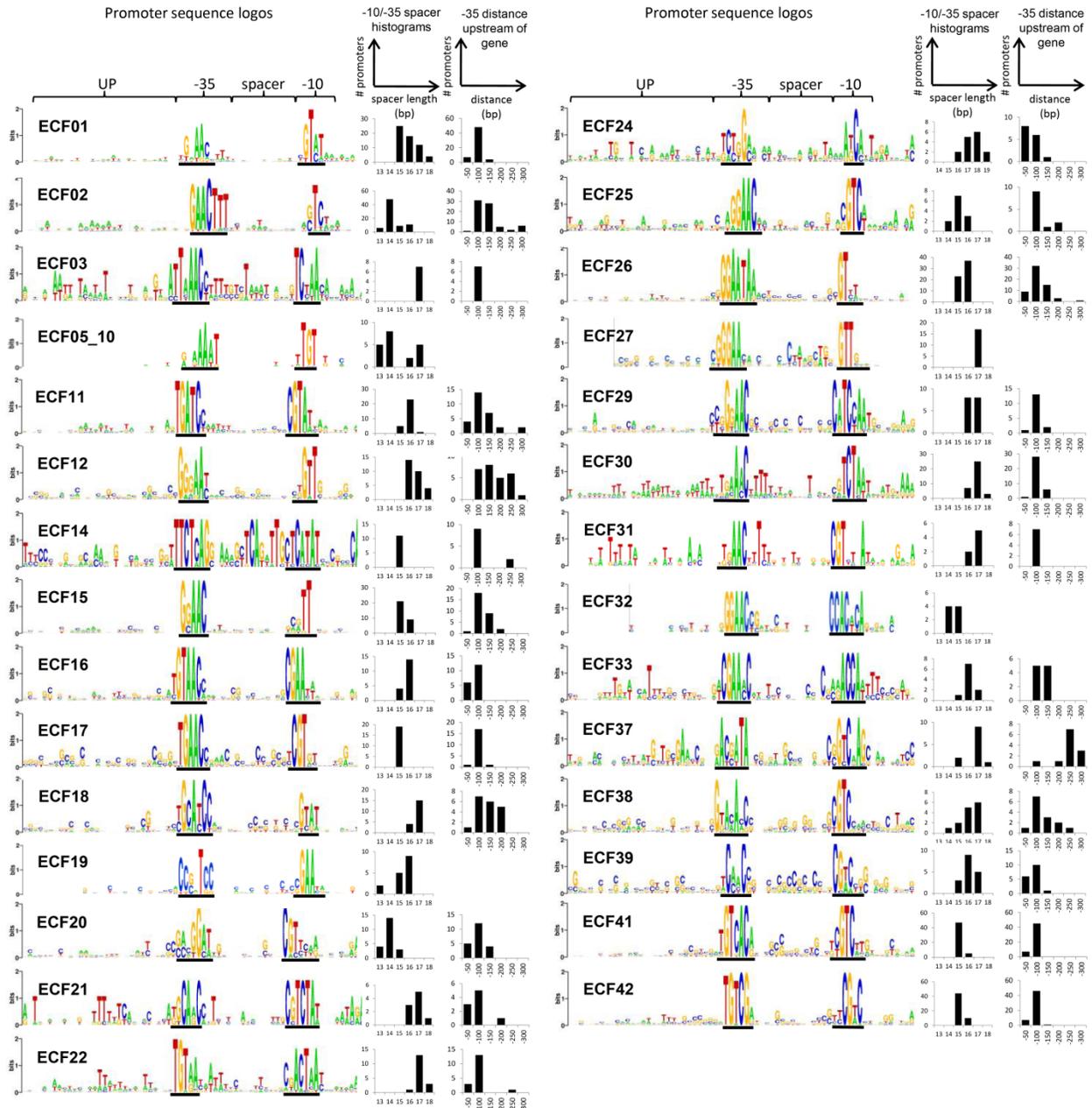
$$S = \ln \left[ \frac{f + 0.005F}{F + 0.005F} \right] , \quad (S2)$$

where  $F$  is the frequency of the most commonly observed (assumed to be optimal) spacer length in the promoter set, and  $f$  is the frequency of the spacer length in the promoter being evaluated. Bayesian pseudocounts of 0.5% of the frequency of the optimal spacer length were added to account for uncertainty.

The total promoter score was calculated as a sum of the -35 and -10 motifs evaluated with PWMs and the spacer length penalty

$$Score = \sum_{i=1}^{L_{-35}} W_{b,i}^{-35} + \ln \left[ \frac{f + 0.005F}{F + 0.005F} \right] + \sum_{i=1}^{L_{-10}} W_{b,i}^{-10} . \quad (S3)$$

When visualizing motifs, the sequence logos of aligned promoter sequences were generated using WebLogo 3 (<http://weblogo.threeplusone.com>; composition set to 50% GC<sup>6</sup>) for Figure 2, and WebLogo 2.8.2 (<http://weblogo.berkeley.edu/>; no small sample correction) for Supplementary Figure S1. For the Weblogos to compensate for the variable spacing between the -35 and -10 motifs for each promoter model, the distances between them was fixed to the most commonly observed spacer length. Figure 2 focuses on the -35 and -10 regions. Figure S1 contains the complete information for the promoter models, including more of the sequence flanking the -35 and -10 motifs, as well showing the distance from the promoter to the downstream target gene.



**Figure S1:** Complete promoter models are shown for each ECF subgroup. The models contain a sequence logo illustrating the upstream (UP) sequence, -35 sequence, spacer sequence, -10 sequence, and 10 bases following the -10. The histograms show, from all the analyzed promoters, the distance between the -35 and -10 motifs, and the distance between the -35 motif and the nearest downstream gene. The exact -35 and -10 sequences identified by the 2-block search algorithm, BioProspector, are underlined underneath each sequence logo, and were used to calculate the distances for the -10/-35 spacer histograms.

*I.C. Predicted orthogonality of the promoters in the library, as well as of their individual -35 and -10 regions.*

The 29 generated promoter models were used to analyze all 706 promoters in the promoter library. This analysis revealed a high level of predicted orthogonality between the ECF subgroups. A similar analysis was performed on just the -10 or -35 subsites, revealing far less predicted orthogonality. Equation S3 was used to evaluate the full promoters, and the first or third term of that equation was used for the -35 or -10 subsite analysis as appropriate.

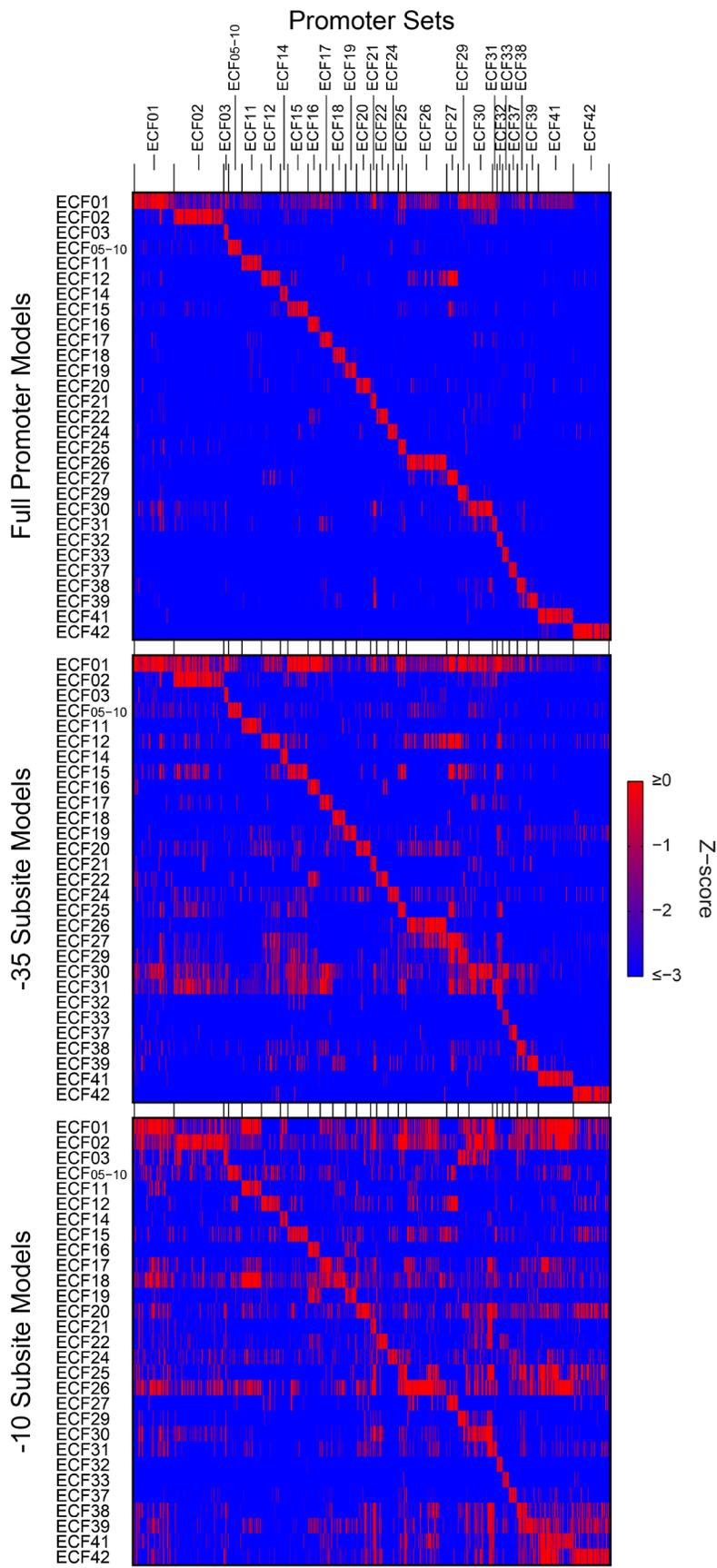
Z-scores were calculated from the full, -35, and -10 promoter scores (Figure S2) by normalizing to the predicted on-target promoter scores. The means and standard deviations of the scores for on-target promoters for each model were calculated, and the z-score calculated with the following equation

$$'Z - score' = \frac{\langle \text{On-target scores} \rangle - \text{Score}}{\sigma_{\text{On-target scores}}} . \quad (\text{S4})$$

Hence, the z-score for a given promoter represents the number standard deviations a promoter is from the mean score of the on-target promoter set.

The orthogonality of the full promoter model was quantitatively compared to the -10 and -35 models using Z-scores. First, the Z-scores were capped at a minimum of -3, as a promoter that is 3 or more standard deviations away from the on-target mean value has little chance to be active. Therefore, a score of less than -3 should generally be equivalent to -3 when considering crosstalk. Next, for each promoter model, the mean of the Z-scores of all the off-target promoters in the promoter library was calculated. This was repeated using just the -35 and -10 models to get three sets of 29 mean off-target Z-scores.

While the mean values of the mean off-target scores are similar (full model= -2.91, -10 model= -2.48, -35 model= -2.72), the full model is shown to be significantly better at generating orthogonality. Two-tailed, paired t-tests between the set of mean off-target scores from the full model and the set of scores from the -10 or -35 models yield p values  $\ll 0.05$  in both cases. This difference remains significant ( $p < 0.05$ ) for both the -10 and -35 model when not comparing pairwise by promoter model (ie, using an unpaired t-test), but it is not as strong for the -35 model. In all analyses, the -10 model is much worse at producing orthogonally than either the -35 model or the full model.



**Figure S2:** The predicted orthogonality of ECF  $\sigma$  promoter models and individual subsite models is shown. Heatmap of Z-score analysis using the 29 promoter models to score all 706 identified ECF  $\sigma$  promoters. The blue to red color range covers Z-scores from -3 to 0, representing scores ranging from the mean on-target score to three standard deviations below. Z-score analysis demonstrates how the -35 and -10 promoter models are both needed for specificity.

#### I.D. $\sigma$ 70 promoter modeling and prediction

We built a  $\sigma$  70 promoter model to screen promoter constructs for potential overlapping  $\sigma$  70 promoter sequences. This model was used to determine whether the identified ECF  $\sigma$  promoters had a possible overlapping  $\sigma$  70 binding site and therefore should not be included in our library (See II.A.). The  $\sigma$  70 promoter model was constructed from 674 known  $\sigma$  70 promoter sequences with experimentally determined transcription starts obtained from RegulonDB<sup>7</sup> (<http://regulondb.ccg.unam.mx/>). Since the -10 and -35 motifs of  $\sigma$  70 promoters are poorly conserved, work by Shultzaberger *et al.*<sup>8</sup> was used as a guide for identifying the motifs. A 2-step search using the 1 block function of BioProspector was used. First, the -10 motif was identified as a 6 mer between positions -16 to -5 (a large window was used to allow for inaccuracies mapping the start site). Next, the -35 motif was identified as a 6 mer 15-20 nt upstream of the identified -10 motif. Four PWMs were constructed using the method of Rhodius and Mutalik<sup>4</sup>. As in Section I.C., a PWM<sub>-35</sub> is built for the -35 motif (aTTGaca) and a PWM<sub>-10</sub> for the -10 motif (TATAaT). In addition, a PWM<sub>spacer</sub> was built for a 10-mer block aligned from -21 to -13 aligned with the -10. This incorporates the putative Zn finger contact from the  $\beta'$  subunit of RNA polymerase (-21 to -18)<sup>9</sup>, -17/-16 dyad and -15/-14 TG motif<sup>10,11</sup>. Finally, PWM<sub>start</sub> is included to capture the transcription start site (-1/+1). All of these PWMs were built using Equation S1. Two spacer penalties were constructed<sup>4</sup> using Equation S2 based on distance histograms between the -35, -10 and start motifs: a spacer penalty (-35 to -10) and a discriminator penalty (-10 to +1). Upstream sequences were scored using counts of overlapping A- and T-tracts between positions -57 to -37, assuming the 5' end of the -35 motif is at position -36<sup>12</sup>. From these terms, the total  $\sigma$  70 promoter score was calculated as:

$$\begin{aligned} \text{Score} = & (\text{UP model}) + \text{PWM}_{-35} + \text{PWM}_{\text{spacer}} + \text{PWM}_{-10} + \text{PWM}_{\text{start}} + (\text{Spacer penalty}) \\ & + (\text{Discriminator penalty}) \end{aligned} \quad (\text{S5})$$

Inserting the terms described in Section I.B yields,

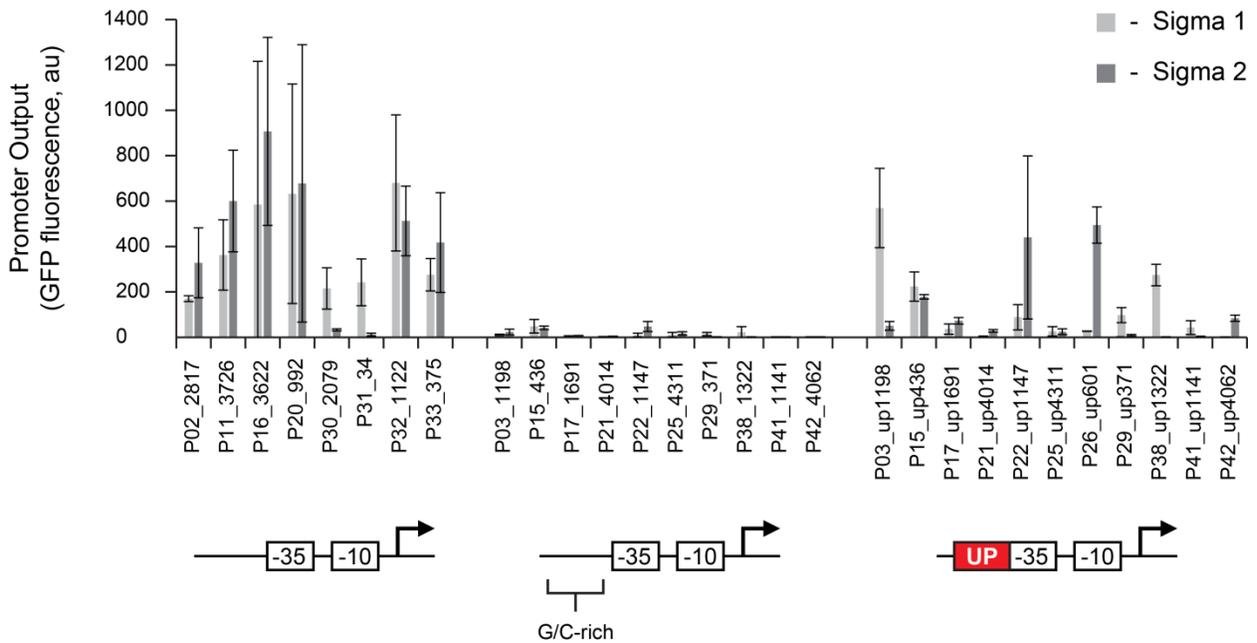
$$\begin{aligned} \text{Score} = & (N_{AAA} + N_{TTT}) + \sum_{i=1}^7 W_{b,i}^{-35} + \sum_{i=1}^{10} W_{b,i}^{\text{spacer}} + \sum_{i=1}^6 W_{b,i}^{-10} + \sum_{i=1}^2 W_{b,i}^{\text{start}} + \ln \left[ \frac{(f_{\text{spacer}} + 0.005F_{\text{spacer}})}{(F_{\text{spacer}} + 0.005F_{\text{spacer}})} \right] + \\ & + \ln \left[ \frac{(f_{\text{discrim}} + 0.005F_{\text{discrim}})}{(F_{\text{discrim}} + 0.005F_{\text{discrim}})} \right] \quad , \quad (\text{S6}) \end{aligned}$$

where  $N_{AAA}$  is the number of AAAs, and  $N_{TTT}$  is the number of Ts proceeding the -35 site.

Note that this promoter model is more complex than that used for the ECF  $\sigma$ s for several reasons. First, the additional PWM<sub>spacer</sub> term is based on several contacts between  $\sigma$ <sup>70</sup>-RNA polymerase and the promoter region that are not known to occur with ECF  $\sigma$ s<sup>9-11</sup>. Second, the discriminator penalty and PWM<sub>start</sub> scores rely on the correct identification of the transcriptional start site for each promoter. This was experimentally established for the  $\sigma$ 70 promoters, but is unknown for the ECF  $\sigma$  promoters. Third, the UP model was not applied to the ECF  $\sigma$  promoter models for determining promoter orthogonality. This is because the upstream sequence does not distinguish promoter specificity between different ECF  $\sigma$ s. However, the UP model was used to optimize ECF  $\sigma$  promoters for function in *E. coli* (see next section).

### I.E. Improving promoters with synthetic UP elements

Promoter sequences were initially tested for activity against both cognate  $\sigma$ s from their own ECF  $\sigma$  group. (See section II.A. for information on how these specific promoters and  $\sigma$ s were selected.) Many non-functional promoter constructs were from GC-rich organisms and consequently had poor upstream sequences with little or no AAA- and TTT-tracts. These were scored by counting the number of overlapping AAA- and TTT-tracts within the sequence window -35 to -57 (assuming that the 5' end of the -10 motif is at position -10). For these promoters, the sequence between -60 to -35 was replaced with a synthetic UP-element derived from the upstream region of the Pefc02\_2817 promoter; CATGACAAAATTTTTAGATGCGTT, which generates a score of 6. The A- and T-tracts were designed predominantly in the proximal  $\alpha$  binding site (-47 to -57) to mimic the location of the observed A- and T-rich sequences of the active  $\sigma$  promoters (data not shown). Adding the UP-element greatly increased the function of a number of the nonfunctional promoters (Figure S3), and the UP-element was added to all promoters except for those that proved functional without it in this test (Pefc02\_2817, Pefc11\_3726, Pefc16\_3622, Pefc20\_992, Pefc30\_2079, Pefc31\_34, Pefc32\_1122, Pefc33\_375). UP-element modified promoters were used in all following experiments.



**Figure S3: Improvement of promoter activity by adding UP-elements.** Promoter sequences were tested for activity against both cognate  $\sigma$ s from their ECF  $\sigma$  group. ( $\sigma$  1 denotes the  $\sigma$  with a lower number in the library, while  $\sigma$  2 denotes the higher one. For example, with P02\_2817,  $\sigma$  1 is ECF02\_915 and  $\sigma$  2 is ECF02\_2817. Inactive promoters tended to contain G/C-rich upstream sequences. These sequences were replaced with synthetic UP-element (CATGACAAAATTTTTAGATGCGTT; -60 to -35), improving promoter activity. In vivo assays were performed by inducing  $\sigma$  expression with 100  $\mu$ M IPTG for 6 hr and measuring promoter activity from GFP fluorescence using flow cytometry. Each bar represents the average promoter output from two independent assays, except for: P15\_436  $\sigma$  1 and 2, P21\_4014  $\sigma$  2, P25\_4311  $\sigma$  1 and 2, P41\_1141  $\sigma$  1 and 2, and P41\_4062  $\sigma$  1 and 2, which are the average promoter output from three independent assays, and P21\_4014  $\sigma$  1, P29\_371  $\sigma$  1 and 2, and P38\_1322  $\sigma$  1 and 2, which are the average promoter output from four independent assays. Error bars represent standard deviations.

## II. $\sigma$ and Anti- $\sigma$ Library Characterization

### II.A. *Selecting $\sigma$ s, anti- $\sigma$ s and promoters*

Libraries of  $\sigma$ s and anti- $\sigma$ s were constructed using several criteria. To maximize phylogenetic diversity, 2  $\sigma$ s were selected from each of the 43 ECF subgroups defined by Staron *et al*<sup>1</sup>. to create a library of 86  $\sigma$ s. Within each subgroup,  $\sigma$ s were preferentially selected from genomes closely related to *E. coli* to maximize the likelihood of binding to *E. coli* RNAP. Since some ECF subgroups only contain  $\sigma$ s from genomes phylogenetically distant to *E. coli* this still resulted in a  $\sigma$  library spanning 6 bacterial classes.  $\sigma$ s were also selected if they had a known cognate anti- $\sigma$ <sup>1</sup>. Not all  $\sigma$ s have known anti- $\sigma$ s; however, this enabled the creation of a library of 62 anti- $\sigma$ s cognate to 62  $\sigma$ s from the main  $\sigma$  library.

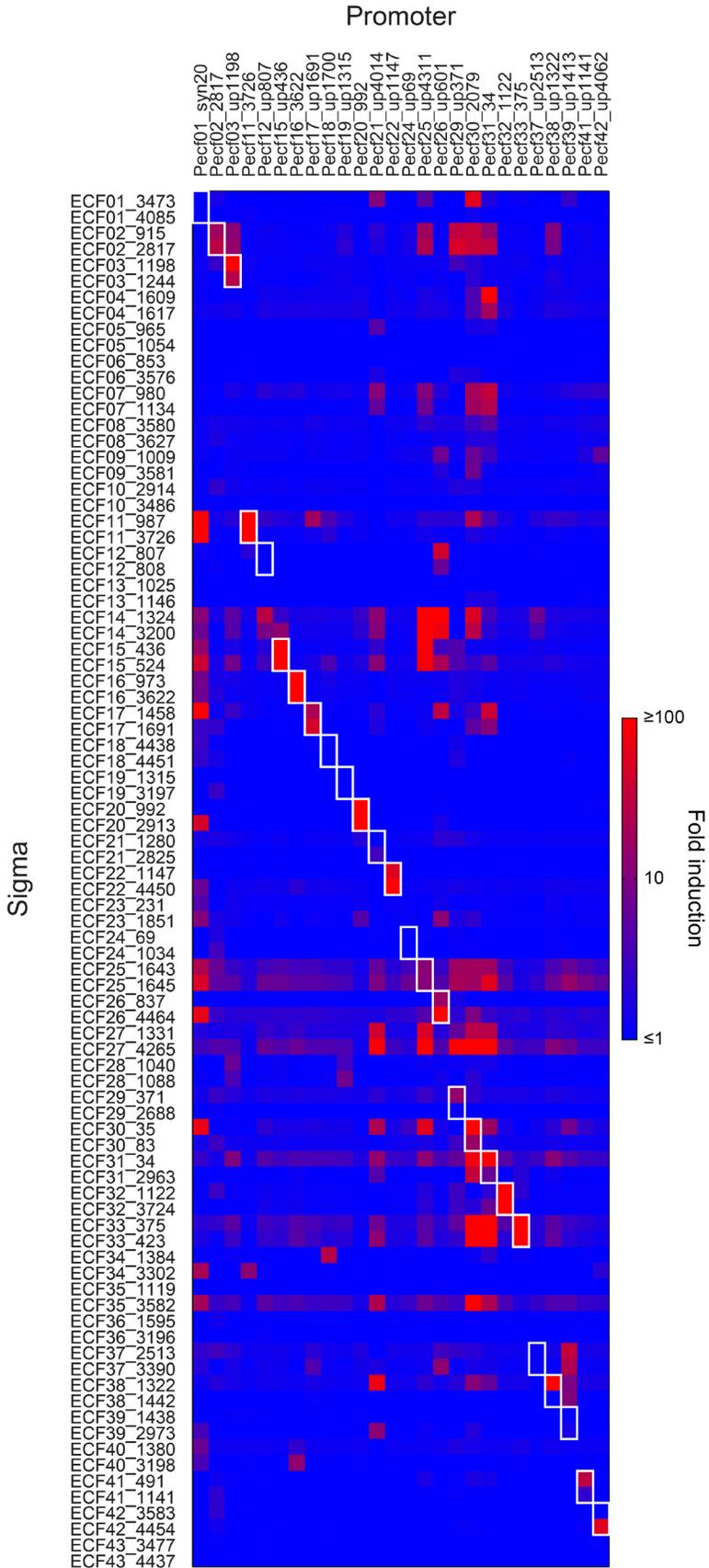
Where possible, a given  $\sigma$  was paired with the putative promoter for that  $\sigma$  from the same genome. In these cases the promoter and  $\sigma$  have the same unique ID (*e.g.*, ECF02\_2817 and P<sub>02-2817</sub>). However, this pairing was not always possible and several criteria were used in selecting the final promoters for each  $\sigma$  group:

- 1) Promoters were preferentially selected that were discovered upstream of  $\sigma$  genes also present in our  $\sigma$  library.
- 2) Preference was given to promoters that were predicted to be orthogonal against the other ECF  $\sigma$ s: *i.e.* scored highly in their own promoter model and scored poorly against the other promoter models.
- 3) Promoters were also screened against any overlapping host promoter sequences using an *E. coli*-specific  $\sigma^{70}$  promoter model for the housekeeping  $\sigma$  and the ECF05-10 promoter model for Fecl. This was especially important for promoters selected from A/T-rich genomes, since they often contained weak overlapping  $\sigma^{70}$  promoter signals that are also A/T-rich.

All promoters were named using the convention P<sub>xx\_yyyy</sub>, where “XX” and “YYYY” denotes the subgroup and unique ID of the downstream parent  $\sigma$  gene (*e.g.*, P<sub>02\_2817</sub> is the promoter upstream of  $\sigma$  ECF02\_2817).

### II.B. *Complete $\sigma$ screening data, including multiple $\sigma$ s from each subgroup and non-orthogonal data*

After promoter optimization, activity assays were performed combinatorially between all optimized promoters and all members of the  $\sigma$  library (Figure S4). For each promoter, cells containing the promoter-gfp construct were transformed with the entire  $\sigma$  library in 96 well format, recovered, induced, and fluorescence measured with flow cytometry. Fluorescence measurements were compared to controls lacking  $\sigma$ s (but including the promoter-gfp construct) to calculate fold induction (Methods). This testing was used to identify the active  $\sigma$ s and promoters in the library, even in cases where the promoter models did not match their intended subgroup. Additionally, these results allowed the selection of a subset of orthogonal  $\sigma$ :promoter pairs, which could be used in the same engineered system without crosstalk. The orthogonal subset of this data is shown in Figure 3e.

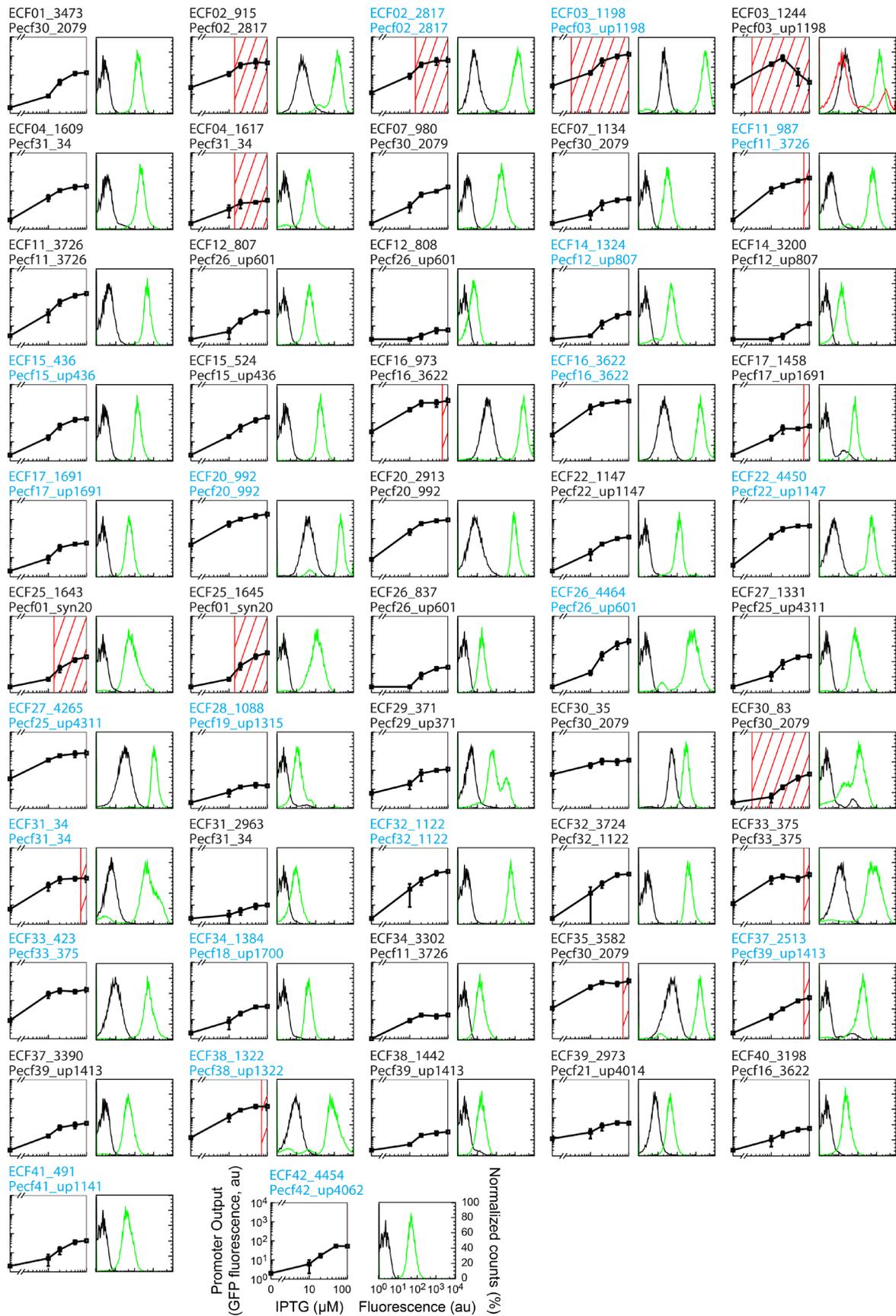


**Figure S4:** *Activity of all  $\sigma$ s with all active promoters.* A heatmap of *in vivo* activities for the complete library of 86  $\sigma$ s tested against 26 active promoters is shown. Assays were performed by inducing  $\sigma$  expression with 100  $\mu$ M IPTG for 6 hr and measuring promoter activity from GFP fluorescence using flow cytometry. White boxes indicate predicted cognate  $\sigma$ :promoter pairs. Each square represents the average fold induction from at least two independent assays.

### II.C. Full transfer functions and cytometry data for promoter induction.

Based on the combinatorial  $\sigma$ :promoter matrix (Figure S4), 58 members of the  $\sigma$  library were found to activate a promoter by at least 5-fold. Of these, 52 were chosen for further testing. Each of these 52  $\sigma$ s was paired with its most active promoter from the combinatorial assay, and induced at multiple levels of IPTG to determine promoter activity at multiple levels of  $\sigma$  (*i.e.*, the induction curve) (Figure S5). Measurements were performed in a similar manner to the combinatorial assay, at 0, 10, 20, and 50  $\mu$ M IPTG in addition to 100  $\mu$ M. These induction curves show a wide range of activities and are generally as expected. A subset of this induction curve data consisting of one member from each active ECF subgroup is shown in Figure 3b.

**Figure S5 (next page):  $\sigma$  titration assays.** Each of the chosen 52 active  $\sigma$ s was induced at 0, 10, 20, 50 and 100  $\mu$ M IPTG with their most active cognate promoter *in vivo*. Assays were performed with 6 hr induction and promoter activity measured from GFP fluorescence using flow cytometry. Plots represent the average promoter activity from three independent assays and error bars represent one standard deviation. Red hash marks indicate toxicity (as judged by a fall in the 8hr OD600 to 75% or lower of wild type; see Section II.D. and Fig. S8; Note that growth measurements were only taken at 0, 10, 20, and 100  $\mu$ M IPTG) occurring at different levels of IPTG induction. Cytometry distributions show typical log-scale histograms at 0 (black) and 100  $\mu$ M IPTG (green). One  $\sigma$ :promoter pair, ECF03\_1244:Pecf03\_up1198, showed extreme toxicity at full induction, so three histograms are shown: black = 0  $\mu$ M, green = 20  $\mu$ M, red = 100  $\mu$ M. All cytometry histograms are shown with log scaling and normalized to the mode of the distribution. A portion of the cells falls on the Y-axis and is not visible when plotted in this way. Light blue labels indicate that that  $\sigma$ :promoter set was included in the orthogonal subset of the  $\sigma$  library (Figure 3e).

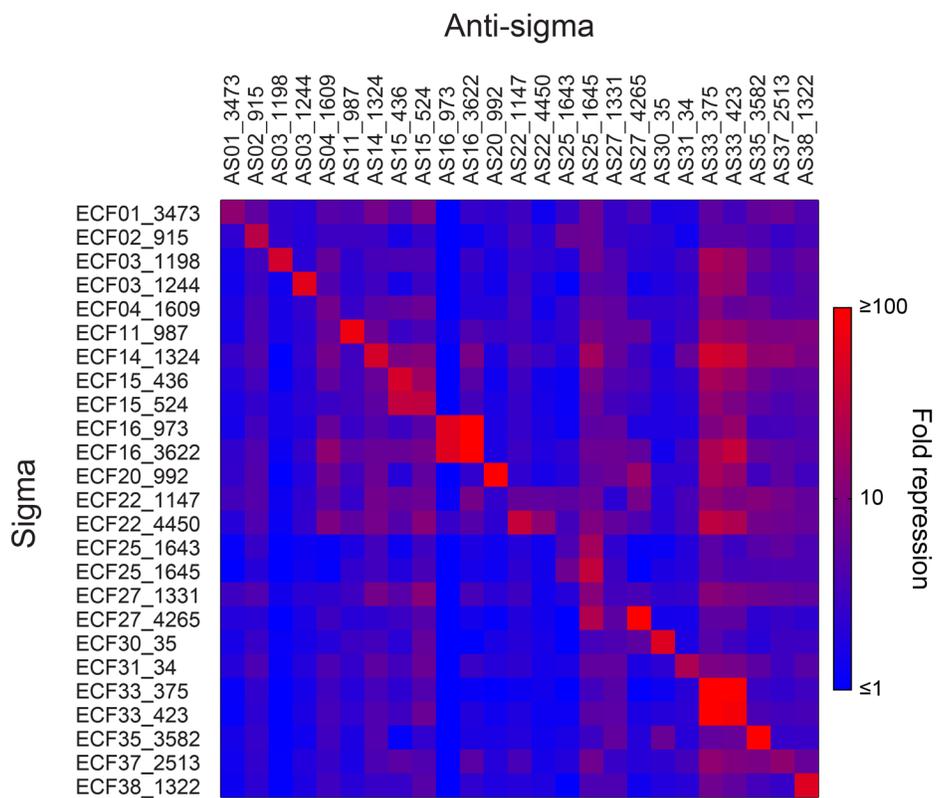


#### II.D. Complete anti- $\sigma$ screening data

This section outlines the initial screen for anti- $\sigma$  activity; more detailed titration curves for those deemed active are presented in Section III. Of the 58  $\sigma$ s shown to activate a promoter by more than 5-fold (Figure S5), 47 have cognate anti- $\sigma$ s in the synthesized library. Based on the strength and orthogonality of the  $\sigma$ :promoter interaction, the most promising 35 anti- $\sigma$ s were chosen for further testing. To check for anti- $\sigma$  activity, titrations of the  $\sigma$  and anti- $\sigma$  were performed with the promoter:reporter construct most activated by the  $\sigma$ . These assays were performed using four levels of induction for the  $\sigma$  (0, 5, 20, and 100  $\mu$ M IPTG), and three for the anti- $\sigma$  (0, 10, and 50 nM HSL) in addition to a control lacking the anti- $\sigma$  expression plasmid. This test showed that 32 of the 35 tested anti- $\sigma$  were able to repress their cognate  $\sigma$  by at least 2-fold (Figure S6).

The 25 anti- $\sigma$ s with the best repression of their cognate  $\sigma$  from the titration assay were chosen for combinatorial orthogonality testing (Figure S7). In this test, the set of 25 anti- $\sigma$ s plus a no anti- $\sigma$  control was tested against the 25  $\sigma$ :promoter pairs targeted by the anti- $\sigma$ s. In order to better observe any repression effects, the  $\sigma$ s were induced to an intermediate level (10  $\mu$ M IPTG), while the anti- $\sigma$ s were induced to a high level (50 nM HSL). This assay shows that a number of the anti- $\sigma$ : $\sigma$  interactions appear to be fairly orthogonal. However, there are also a number that were less specific and affected many  $\sigma$ s. These less specific anti- $\sigma$ s often greatly reduce growth (See Section II.D.).





**Figure S7:** *Repression of  $\sigma$  by all tested anti- $\sigma$ s.* A heatmap of fold repression of  $\sigma$  activity by different anti- $\sigma$ s is shown. Assays were performed by inducing  $\sigma$  expression with 10  $\mu$ M IPTG and anti- $\sigma$  expression with 50 nM HSL for 6 hr and fold repression calculated from fluorescence measurements. Each square represents the average fold repression from at least two independent assays.

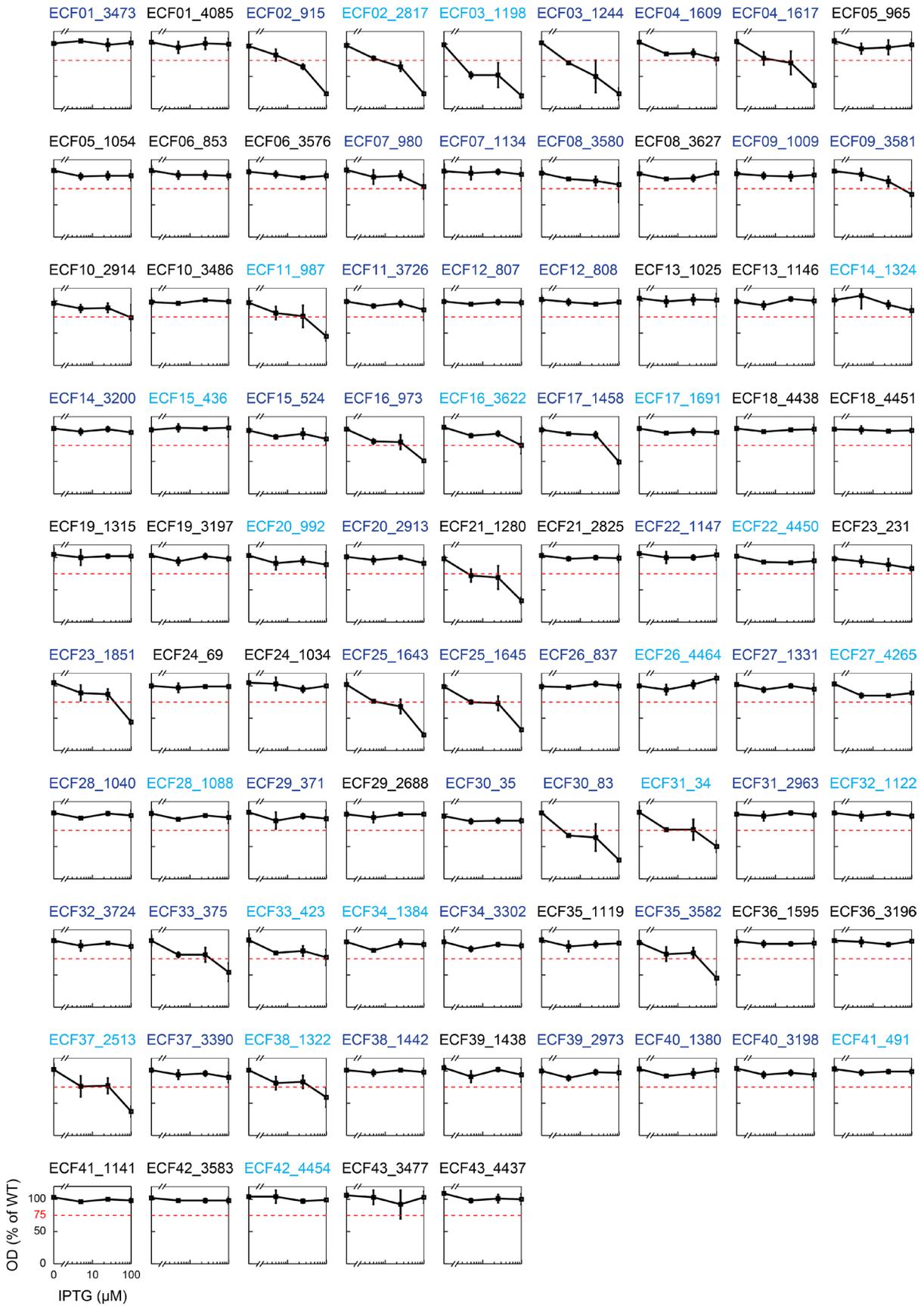
## II.E. $\sigma$ and anti- $\sigma$ library growth assays

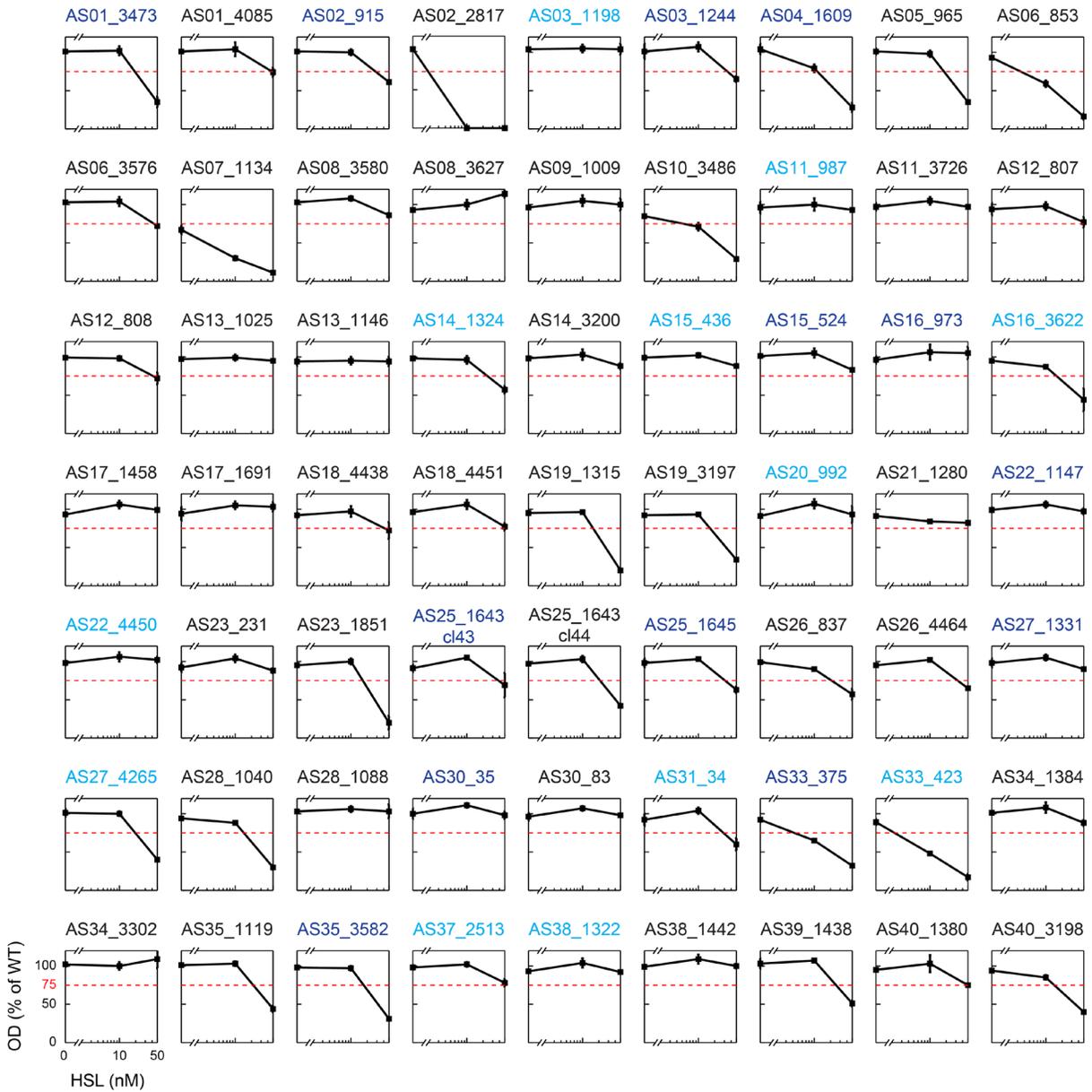
Both the  $\sigma$  and anti- $\sigma$  libraries were tested for toxic effects occurring with expression in *E. coli* DH10 $\beta$ . Toxicity can be due to aberrant gene expression or titration of host RNAP by the  $\sigma$ s, or by interaction of the anti- $\sigma$ s with essential host  $\sigma$ s such as  $\sigma^E$ . The effects of expressing the  $\sigma$ s and anti- $\sigma$ s were measured using 3 types of growth assays across a range of inductions: 1) transition phase culture density in liquid LB media; 2) exponential growth rates in liquid LB media; 3) colony size on LB agar plates (Figures 3c; S8; S9; Supplemental Tables S2.4, S2.5). For each condition, growth assays were performed from at least 3 separate transformations and across a range of inducer concentrations: 0, 10, 20 and 100  $\mu$ M IPTG for the  $\sigma$  library; 0, 10 and 50 nM HSL for the anti- $\sigma$  library. The  $\sigma$  library assay strains were freshly transformed *E. coli* DH10 $\beta$  cells carrying pN565 with the pVRa plasmid library and plasmid pET21a (Novagen) as a no  $\sigma$  control; the anti- $\sigma$  library assay strains were *E. coli* DH10 $\beta$  cells freshly transformed with the pVRc plasmid library and pACYC184<sup>13</sup> as a no anti- $\sigma$  control.

Under low levels of induction (10  $\mu$ M IPTG or 10 nM HSL for the  $\sigma$  and anti- $\sigma$  libraries, respectively) 95% of the  $\sigma$  library and 86% of the anti- $\sigma$  library exhibited near wild type growth levels by all metrics (>75% wild type growth). Under high induction levels (100  $\mu$ M IPTG or 50 nM HSL for the  $\sigma$  and anti- $\sigma$  libraries, respectively) most growth defects were observed during transition phase and by colony size. For the  $\sigma$  library, 99% exhibited near wild type growth levels (>75% wild type growth) during exponential growth, whilst 77% and 90% exhibited near wild type growth measured in transition phase or by colony size, respectively. A similar pattern was observed with the anti- $\sigma$  library but with slightly larger defects: 83%, 49% and 46% exhibited near wild type growth levels during exponential growth, in transition phase and by colony size, respectively. In general, transition phase and colony size yield a similar pattern of growth defects in both states across the  $\sigma$  and anti- $\sigma$  libraries, likely due to the transition/stationary phase growth properties of cells in the center of colonies.

Both  $\sigma$ s from subgroup 02 exhibited the highest toxicity. *E. coli*  $\sigma^E$  is also from subgroup 02 and is represented by the candidate ECF02\_2817 in the  $\sigma$  library. *E. coli*  $\sigma^E$  is toxic when highly expressed<sup>14</sup>; consequently, the toxic effects of high expression of both ECF02  $\sigma$  members in the library (ECF02\_2817 and ECF02\_915) suggest similar function. *E. coli*  $\sigma^E$  is also essential<sup>15,16</sup>; accordingly, high expression of its cognate anti- $\sigma$  AS02\_2817 is lethal due to repression of host  $\sigma^E$  activity. Interestingly, high expression of anti- $\sigma$  AS02\_915 from the same subgroup only gave reduced growth levels, suggesting that this anti- $\sigma$  has reduced specificity for host  $\sigma^E$ . Both  $\sigma$  pairs from subgroups 03 and 25, and anti- $\sigma$  pairs from subgroups 19, 33 and 35 were also highly toxic (<50% wild type growth), indicating similar activities of each member within the subgroup. There were also several instances of where just one subgroup member was toxic, indicating different functionality in an *E. coli* host (*e.g.* ability to bind *E. coli* RNAP). Importantly, the lack of toxicity of most library members suggests that they could have utility as orthogonal regulators in *E. coli*.

**Figure S8 (next page):  $\sigma$  transition phase growth assay.** *OD*<sub>600</sub> measurements were taken after 8 hours of growth at different levels of expression of each of the  $\sigma$ s in the library. Plots represent the average measurement from three independent assays normalized to cells not expressing a  $\sigma$ , and error bars represent one standard deviation. Dark blue titles indicate 'active'  $\sigma$ s with more than 5-fold induction of a promoter in the library, and light blue titles indicate a  $\sigma$  in the final orthogonal set (Figure 3e).

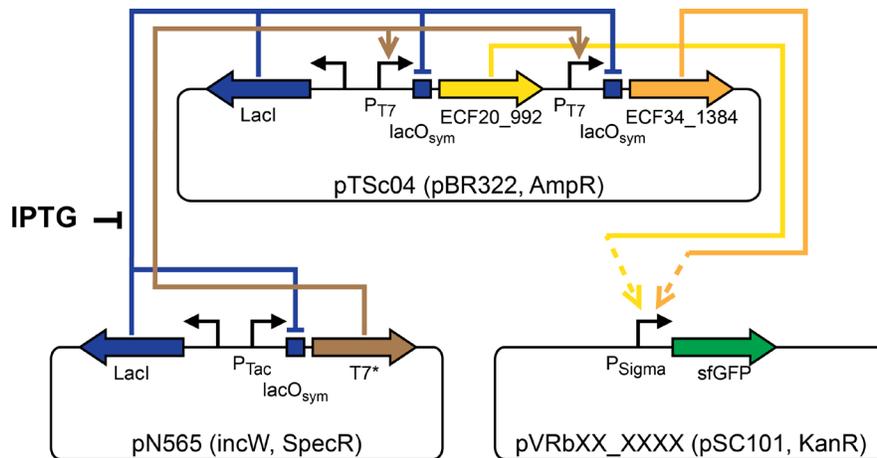




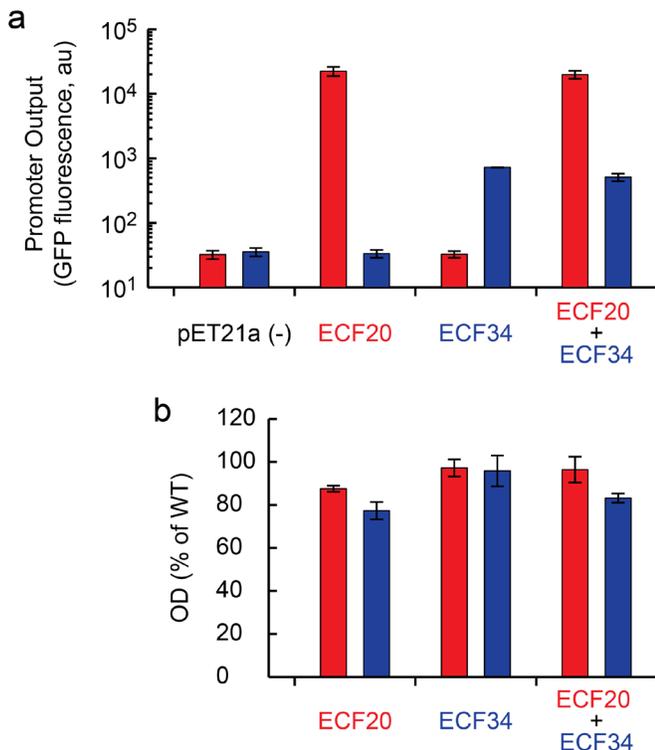
**Figure S9:** *Anti- $\sigma$  transition phase growth assay.* OD<sub>600</sub> measurements were taken after 8 hours of growth at different levels of expression of each of the anti- $\sigma$  in the library. Plots represents the average measurement from three independent assays normalized to cells not expressing an anti- $\sigma$ , and error bars represent one standard deviation. Dark blue titles indicate the 25 anti- $\sigma$ s tested for orthogonality (Fig. S7), and light blue titles indicate anti- $\sigma$ s that repress  $\sigma$  in the final orthogonal set (Figure 5c).

## II.F. Simultaneous expression of multiple $\sigma$

Two  $\sigma$ s from the library were simultaneously expressed to test whether they would interfere with each other's function. One relatively high activity  $\sigma$  (ECF20\_992) and one relatively low activity  $\sigma$  (ECF34\_1384) were selected, and a plasmid constructed that expresses each from a pT7 promoter (pTSc04). This plasmid was transformed into Z-competent *E. coli* DH10 $\beta$  cells (see chimeric  $\sigma$  assay for method) containing plasmid pN565 and either the pECF20\_992 or pECF18\_up1700 (activated by ECF34\_1384) reporters (pVRb plasmids) and induced fully (diluted 1:200 into LB + Antibiotics + 100  $\mu$ M IPTG, grown at 1000 rpm, 37  $^{\circ}$ C). The activity and toxicity of the two  $\sigma$ s was compared to cells with a negative control plasmid (pET21a) and plasmids expressing only one  $\sigma$  (pVRa plasmids).



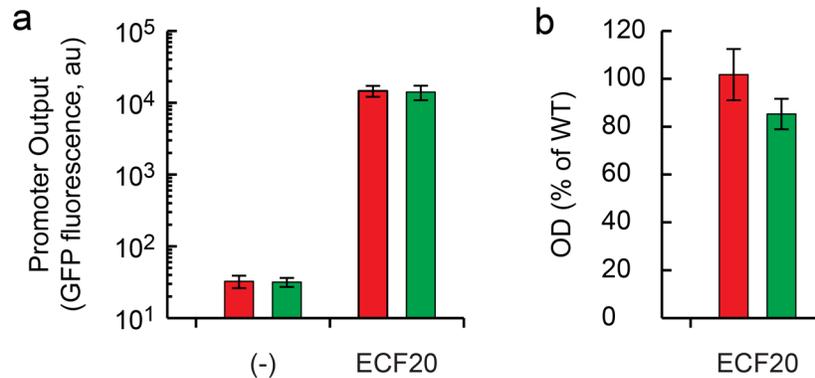
**Figure S10:** Plasmids for simultaneous expression of multiple  $\sigma$ s. pTSc04 is based on pVRa, with both ECF20\_992 and ECF34\_1384 expressed from pT7 promoters. It is driven by T7 RNA polymerase from pN565, and the activity of the  $\sigma$ s measured using plasmid series pVRb.



**Figure S11:** Simultaneous expression of multiple  $\sigma$ s. Cells carrying either the reporter for ECF20\_992 (p20\_992, red bars) or for ECF34\_1384 (p18\_up1700, blue bars) were transformed and assayed with plasmids expressing either ECF20\_992, ECF34\_138, both of them simultaneously, or a negative control. (a) After induction with 100  $\mu$ M IPTG for 6 hours, fluorescence was measured with flow cytometry to gauge  $\sigma$  activity. (b) After a total of 8 hours of induction, OD<sub>600</sub> measurements were taken with a Synergy H1 plate reader and normalized to cells not expressing a  $\sigma$ . For both measurements, each bar represents the mean from three independent experiments, and error bars show one standard deviation.

## II.G. Expression of a $\sigma$ in *Klebsiella oxytoca*

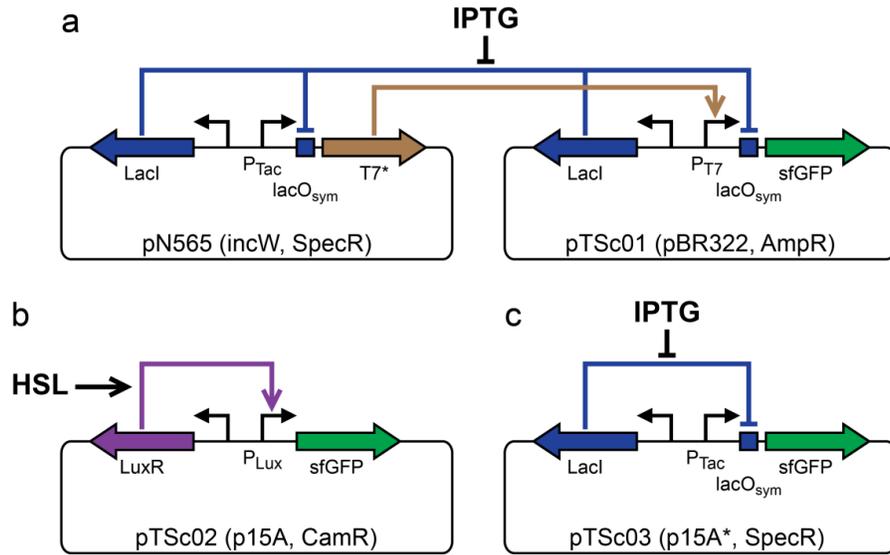
A strongly active  $\sigma$ , ECF20\_992, was expressed in *Klebsiella oxytoca* M5a1 to assess how the  $\sigma$ s may perform in a different species. The  $\sigma$  library plasmid expressing ECF20\_992 (pVRa20\_992) was transformed into Z-competent *Klebsiella oxytoca* M5a1 cells (see chimeric  $\sigma$  assay for method) containing pN565 and the ECF20\_992 reporter (pVRb20\_992), induced to a moderate amount (diluted 1:200 from overnights into LB + Antibiotics + 10  $\mu$ M IPTG, grown at 1000 rpm, 30  $^{\circ}$ C), and the fluorescence and OD600 measured and compared to cells transformed with a negative control plasmid (pET21a). Simultaneous measurements were done with *E.coli* DH10 $\beta$  in an identical fashion (except grown at 37  $^{\circ}$ C) for comparison.



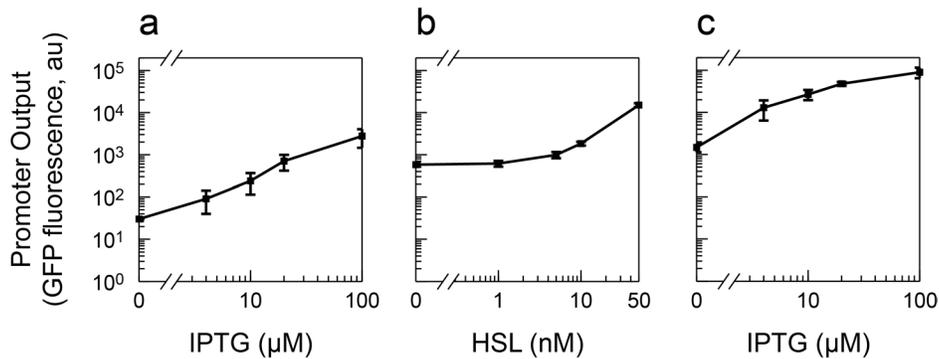
**Figure S12:** Expression of an  $\sigma$  in *Klebsiella oxytoca* M5a1. The activity and toxicity of ECF20\_992 was assayed in both *E.coli* DH10 $\beta$  (red bars) and *K.oxytoca* M5a1 (green bars) carrying a p20\_992 reporter plasmid. (a) After 6 hours of induction, fluorescence was measured with flow cytometry. (b) OD<sub>600</sub> was measured after 8 hours of induction with a Synergy H1 plate reader and normalized to cells carrying a negative control plasmid. All bars represent the mean from three independent experiments and error bars show one standard deviation.

## II.H. Characterization of induction plasmids

In order to characterize the relative strengths of the expression systems in our assays, plasmids were constructed that express sfGFP in place of a  $\sigma$  or anti- $\sigma$ . (Figure S13). These plasmids were induced using IPTG (0, 4, 10, 20, and 100  $\mu$ M) or HSL (0, 1, 5, 10, and 50 nM) as appropriate and the fluorescence resulting from GFP expression was measured.



**Figure S13:** Plasmids used for expression system characterization. (a) pTSc01 is based on the pVRa plasmid series with sfGFP expressed in place of a  $\sigma$ . (b) pTSc02 is based on the pVRc plasmid series with sfGFP expressed in place of an anti- $\sigma$ . (c) Plasmid pTSc03 is based on the pTSa plasmid series with sfGFP expressed in place of a parental or chimeric  $\sigma$ .

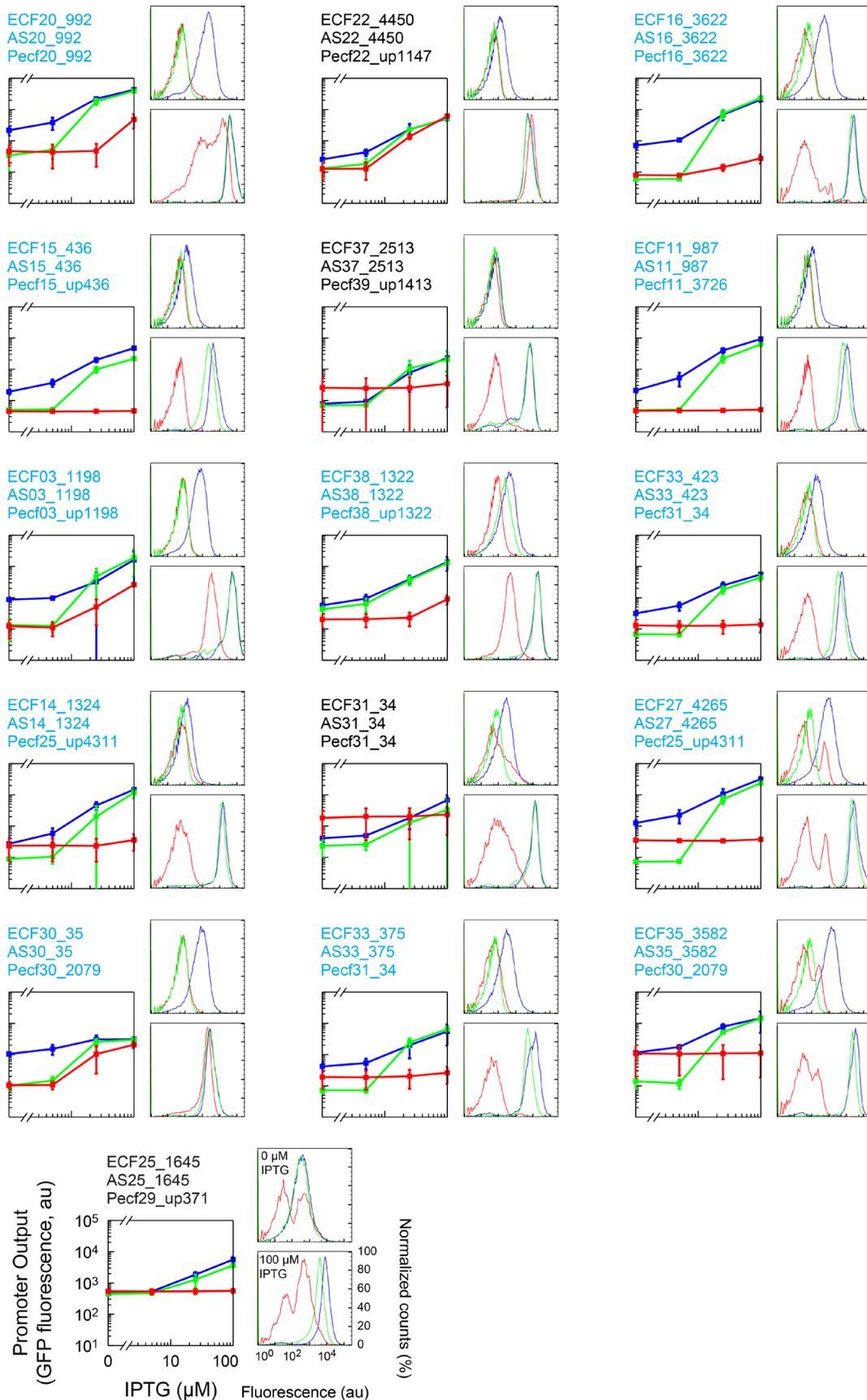


**Figure S14:** Induction plasmid transfer functions. (a) *E.coli* DH10 $\beta$  cells carrying pTSc01 and Nif565 were induced with IPTG (0, 4, 10, 20, and 100  $\mu$ M) for 6 hours. (b) *E.coli* DH10 $\beta$  cells carrying pTSc02 were induced with HSL (0, 1, 5, 10, and 50 nM) for 6 hours. (c) *E.coli* DH10 $\beta$  cells carrying pTSc03 were induced with IPTG (0, 4, 10, 20, and 100  $\mu$ M) for 6 hours. In all cases, fluorescence was measured with flow cytometry, and data points represent the average of three measurements taken on separate days, with error bars showing one standard deviation.

### III. Quantification of Anti- $\sigma$ Threshold Control

A subset of the anti- $\sigma$ : $\sigma$  pairs were assayed in more detail to determine their capability to implement ultrasensitivity through sequestration<sup>17</sup> (Figure S15). Sixteen of the  $\sigma$ :anti- $\sigma$ :promoter sets previously tested were selected based on either: 1. targeting one of the  $\sigma$ s in the orthogonal subset, or 2. having an induction curve that suggests switch-like behavior in Figure S6. These sets were induced at four levels (0, 5, 25, and 100  $\mu$ M IPTG) of  $\sigma$  and three levels (no anti- $\sigma$  plasmid, 0 nM HSL induction, 50 nM HSL induction) of anti- $\sigma$  and the promoter activities were measured via fluorescence. High expression of the anti- $\sigma$  often significantly reduced the promoter output at all levels of  $\sigma$  induction, in many cases also causing highly toxic effects. In contrast, the lower anti- $\sigma$  induction showed the desired threshold effect in many cases. At this level of anti- $\sigma$ , the higher levels of  $\sigma$  induction showed promoter activity close to the no anti- $\sigma$  control, while the lower levels of  $\sigma$  had much lower activity than the equivalent induction points with no anti- $\sigma$ . This differential repression is characteristic of a threshold system, and increases the utility of these proteins in applications where a more digital-like signal response is desired. The 9 anti- $\sigma$ : $\sigma$ :promoter sets that have the best induction curves and correspond to an orthogonal  $\sigma$  are shown in Figure 5d.

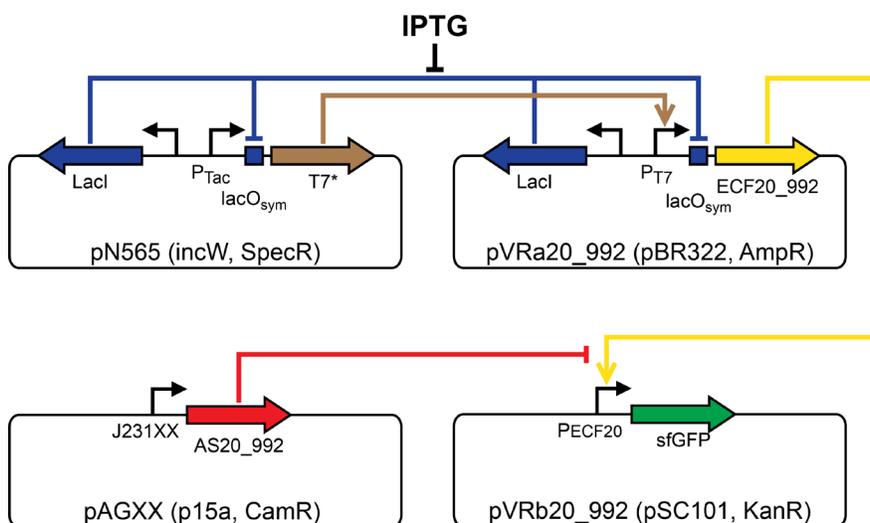
**Figure S15 (next page): Anti- $\sigma$  thresholding.** Sixteen sets of anti- $\sigma$ ,  $\sigma$ , and promoter were assayed at varying inductions of both anti- $\sigma$  and  $\sigma$  to characterize their ability to implement thresholding. The anti- $\sigma$ s were either not present (blue), or induced at 0nM HSL (green) or 50nM HSL (red), while the  $\sigma$ s were induced at 0, 5, 25, and 100  $\mu$ M IPTG. Plots represent the average promoter activity from three independent assays and error bars represent one standard deviation. Histograms show representative distributions from a 0 and 100  $\mu$ M induction for each set. Light blue titles indicate data shown in Figure 5d.



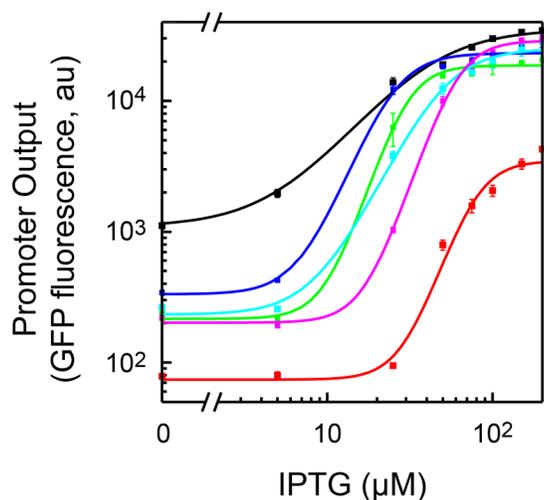
To precisely model the impact of the anti- $\sigma$  sequestration on switch cooperativity, a threshold-gated switch was constructed using ECF20\_992 and AS20\_992 and characterized more thoroughly (Figure S17). The inducible anti- $\sigma$  system was supplemented by a set of plasmids constitutively expressing the anti- $\sigma$  AS20\_992 at a number of levels (Figure S16). Changing the strength of a constitutive promoter allowed for finer control over the expression level of anti- $\sigma$ . This system was tested at 8 induction levels of the  $\sigma$  (0, 5, 25, 50, 75, 100, 150, 200  $\mu$ M IPTG) to characterize the transfer function. Finally, the Hill equation was used to fit the data,

$$\frac{y-y_{min}}{y_{max}-y_{min}} = \frac{x^n}{x^n+K^n} \quad , \quad (S5)$$

where  $x$  is the IPTG induction concentration,  $y$  is the output (promoter activity),  $y_{max}$  is the maximum output,  $y_{min}$  is the minimum output,  $K$  is the half-maximum, and  $n$  is the Hill coefficient. The optimization was weighted to minimize the relative least-squares error so that the model fit both the low and high ends of the data.



**Figure S16:** *Alternate plasmid set used in anti- $\sigma$  threshold experiments.* An alternate series of plasmids, pAGXX, replaces the pVRcXX\_XXXX plasmid series used the anti- $\sigma$  library testing experiments (Fig. S23). This plasmid series expresses the anti- $\sigma$  AS20\_992 from a range of constitutive promoters instead of from the  $P_{lux}$  promoter.



**Figure S17:** *Hill fits to AS/ECF20\_992 anti- $\sigma$ : $\sigma$  threshold switch.* Threshold-gated switches were built using  $\sigma$  ECF20\_992 and its corresponding anti- $\sigma$  and promoter. ECF20\_992 was induced with varying levels of IPTG (0, 5, 25, 50, 75, 100, 150, and 200  $\mu$ M) in the presence of plasmids expressing AS20\_992 from varying promoters: No anti- $\sigma$  (black), BBa\_J23117 (blue), BBa\_J23105 (green), BBa\_J23101 (cyan), BBa\_J23100 (pink), pLux induced with 50 nM HSL (red). Equation S5 was fit to the data using a relative least-squares algorithm and the fit is shown. The Hill coefficient  $n$  is higher when the anti- $\sigma$  is present:  $n = 1.7$  (No anti- $\sigma$ ),  $n = 3.3$  (BBa\_J23117),  $n = 4.1$  (BBa\_J23105),  $n = 2.6$  (BBa\_J23101),  $n = 3.9$  (BBa\_J23100 and pLux). Plots represent the average promoter activity from three technical replicates (No anti- $\sigma$ ) or three independent assays (all others) and error bars represent one standard deviation.

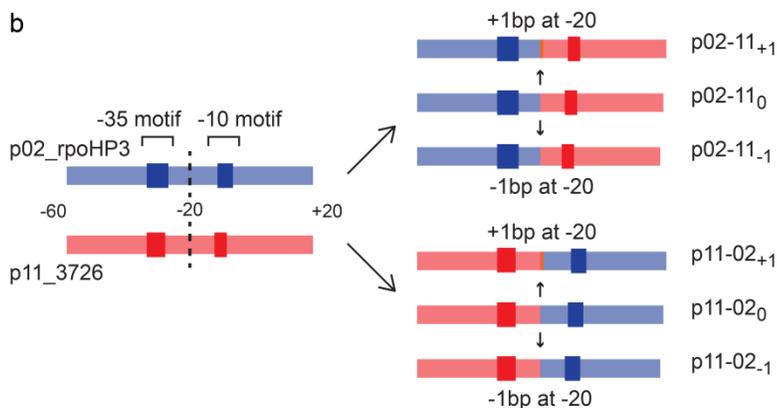
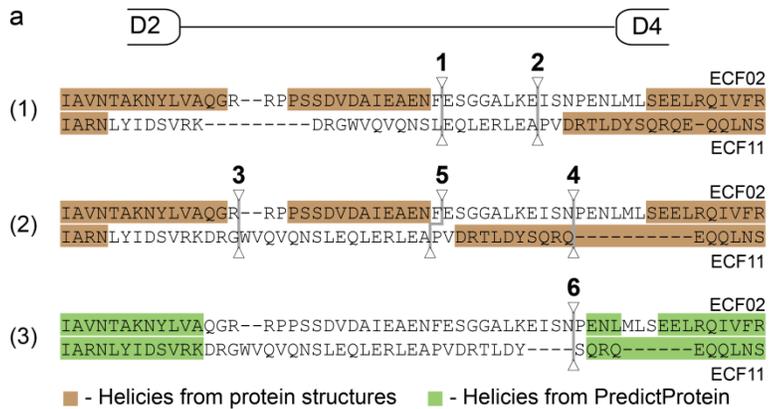
## IV. Creating Chimeric $\sigma$

### IV.A. Design of chimeric $\sigma$ s and promoters

A combination of protein alignment, structural information, and secondary structure prediction algorithms were used to generate chimeric  $\sigma$ s from ECF02\_2817 and ECF11\_3726 (Figures 3g and S18a). These parental  $\sigma$ s were chosen since they have high activity in *E. coli* and there is protein structural information available that could be used to guide the construction of the chimeras (ECF02\_2817 (*E. coli*  $\sigma^E$ )<sup>18</sup> and *R. sphaeroides*  $\sigma^E$ , which belongs to the same subgroup as ECF11\_3726<sup>19</sup>). Chimeras of both combinations (N-terminal ECF02\_2817 / C-terminal ECF11\_3726 and N-terminal ECF11\_3726 / C-terminal ECF02\_2817) were created by recombining the parental proteins at six 'crossover seams' located in the flexible linker region between the conserved domains 2 and 4, which recognize the -10 and -35 promoter subsites, respectively. While domains 2 and 4 play the most important roles in promoter recognition, the linker region between these regions in Group I  $\sigma$ s plays an important role in abortive initiation and promoter escape<sup>20</sup>, and likely plays a similar role in the ECF  $\sigma$ s. Consequently, the choice of crossover seams within the linkers of the ECF02\_2817 and ECF11\_3726  $\sigma$ s may affect the functionality of the resultant chimeras. The structure and precise boundary of the linker region in the ECF $\sigma$ s is ambiguous for two reasons: 1. in both structures the  $\sigma$ s are bound to their cognate anti- $\sigma$ , distorting the structure of the linker and 2. the amino acid sequence of the linker region is poorly conserved, making accurate alignments challenging.

In order to select a range of potentially functional crossover seams, the full library of 86  $\sigma$ s was initially aligned using clustalW (<http://www.ebi.ac.uk/Tools/msa/clustalw2/>)<sup>21</sup>. The alignment of ECF02\_2817 and ECF11\_3726 was then tweaked by hand based on the protein structures mentioned previously. Crossover seams 1 and 2 were located at either end of the flexible linker in this alignment. Due to some uncertainties in the structural analysis (specifically, that the linkers were too distorted by binding anti- $\sigma$ s for proper structural analysis) crossover seams 4, 5, and 6 were generated from the unaltered clustalW alignment near the beginning, middle, and end of the linker. Finally, a secondary structure prediction algorithm, PredictProtein<sup>22</sup>, was used to analyze ECF02\_2817 and ECF11\_3726 for  $\alpha$ -helices. Crossover seam 6 was placed one residue before the beginning of the first  $\alpha$ -helix after the linker region in both proteins.

Chimeric promoters were similarly created by crossing over cognate promoters for ECF02\_2817 and ECF11\_3726 between the -10 and -35 boxes (Figure S18b). The promoter rpoHP3<sup>23</sup> from *E. coli* was used as the parental pECF02 promoter, with a 1 bp mutation (T-34G) made from the WT sequence to differentiate it more from ECF11 promoters. (Note that this promoter contains an overlapping  $\sigma$ 70 promoter<sup>23</sup>, which likely partly accounts for the high background induction level and low dynamic range of activation by ECF02.) The pECF11\_3726 promoter from the  $\sigma$  library was chosen as the parental pECF11 promoter. In each case, to -60 to +20 region of the promoter was used, and these parental promoters were crossed over between -20 and -21 to make chimeric promoters. While the initially engineered chimeric promoters were functional, they were relatively weak when compared to the parental promoters. One explanation for reduced activity is that while the -10 and -35 recognition sites are identical to the parental plasmids, the spacing between them may not be optimal for the chimeric proteins. This is made even more likely because of uncertainties in identifying the -10 and -35 sites in the promoter, and because the ECF02 and 11 promoter models have different optimal spacings (Figures 2 and S1: ECF02 has optimal spacing 14, and ECF11 has optimal spacing 16). For these reasons, additional chimeric promoters were engineered with the -10 and -35 sites moved either 1bp closer or 1bp farther apart.



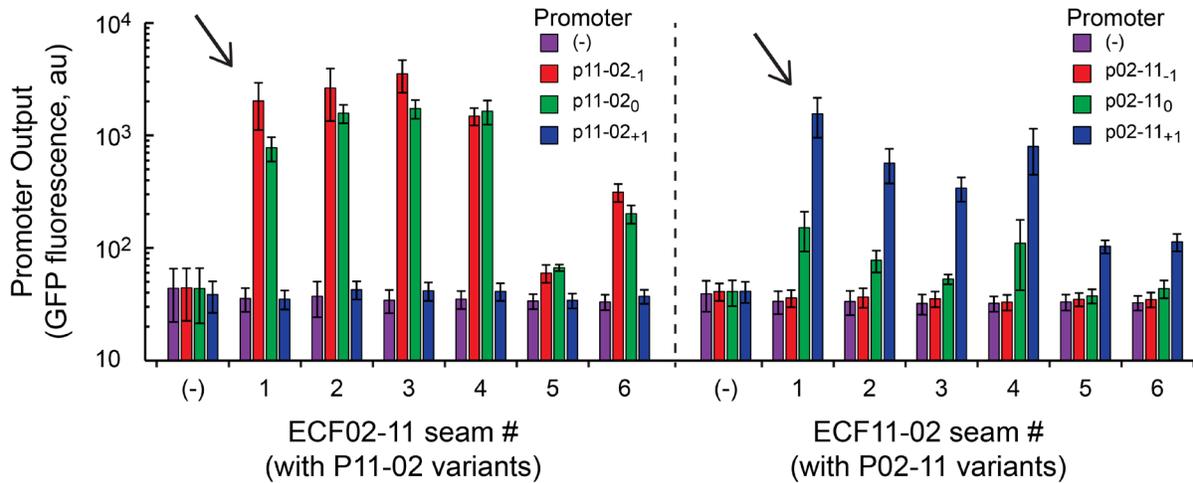
**Figure S18: Chimeric  $\sigma$  and promoter engineering.** (a) Three alignments of the flexible linker region connecting domains 2 and 4 of the parental  $\sigma$ s are shown with the crossover seams derived from each alignment marked. Alignment 1 was generated through protein sequence alignment, hand-adjusted based on crystal structures. Alignment 2 was generated through protein sequence alignment alone. Alignment 3 was created with secondary structure prediction by PredictProtein. Each of the six marked seams was used to engineer one chimeric  $\sigma$  of either orientation, for a total of 12 chimeric  $\sigma$ s. (b) Three chimeric promoters of each orientation were similarly engineered. pECF02\_rpoHP3 and pECF11\_3726 were recombined between -20 and -21 to make p02-11<sub>0</sub> and p11-02<sub>0</sub>. To correct for any differences in optimal spacing between the chimeras and parental  $\sigma$ s, 1 bp was added or removed at the -20/-21 seam to make additional promoters with longer or shorter spacers.

#### IV.B. Chimeric $\sigma$ characterization

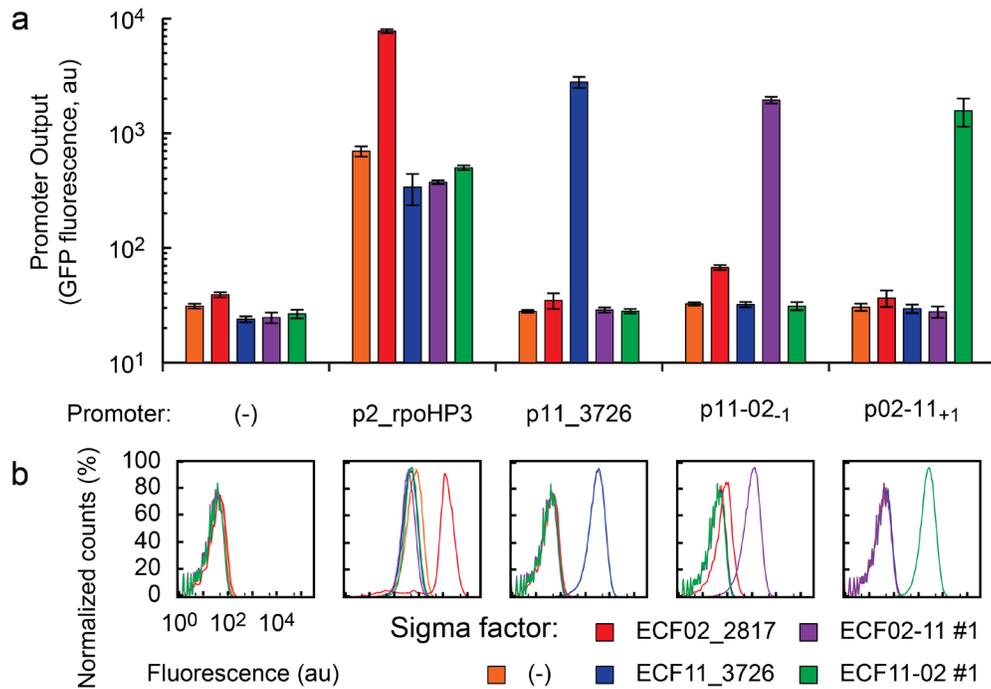
The chimeric  $\sigma$ s and promoters were first assayed to determine which crossover seams and promoter variants were most successful (Figure S19). Each of the six versions of each chimera was paired with each of the three versions of its cognate promoter, and the promoter activity determined *in vivo*. From this assay, it seems as though  $\sigma$ s can tolerate chimeragenesis in many different areas and alignments within the linker. Despite the differing alignments used to design the chimeras, seams 1-4 produced very active chimeras for both ECF02-11 and ECF11-02, seams 5 and 6 were slightly active in ECF11-02, and seam 6 was slightly active in ECF02-11. Of these seams, seam 1 was chosen for further experimentation as it was the most active variant of ECF11-02, and one of the more active variants of ECF02-11. In contrast to the flexibility on protein crossover location, the chimeric promoter spacing had an extreme effect on  $\sigma$  chimera activity. The initially built chimeric promoters had an intermediate level of activity, while pECF02-11<sub>-1</sub> and pECF11-02<sub>+1</sub> were greatly improved. In contrast, pECF11-02<sub>-1</sub> and pECF02-11<sub>+1</sub> were inactive, indicating that 1-2 bps of change in the distance between the -10 and -35 sites is enough to abrogate promoter activity. This result also demonstrates that optimal promoter spacing is determined by the source of the parent -10/Domain 2 of the chimeric promoter and  $\sigma$ . Based on these results, ECF11-02 #1, ECF02-11 #1, pECF02-11<sub>-1</sub> and pECF11-02<sub>+1</sub> were chosen as the chimeric  $\sigma$ s and promoters to be used for further chimera testing.

Next, using the optimized chimera constructs, the parental and chimeric  $\sigma$ s and promoters were tested with each other to check their orthogonality (Figures S20, S21, and 3e). Each of the two parental  $\sigma$ s and chimeras from the most active seam was tested with each of the two parental promoters and best chimeric promoters. Promoter activity was measured *in vivo*, and fold induction of each promoter was calculated using a negative control plasmid that does not express a  $\sigma$ . This assay demonstrates that both the -10 and -35 sites must be recognized for promoter activation. The chimeric  $\sigma$ s activated their promoters by more than 50-fold more than the parental  $\sigma$ s, and the parental  $\sigma$  promoters likewise only recognized the parental  $\sigma$ s. ECF02\_2817:pECF02\_rpoHP3 displayed the weakest fold induction at ~4-10-fold, however, this is due to extremely high background activation and toxicity from overexpression. This assay was run in two different strains, *E. coli* CAG22216, which is deficient in ECF02\_2817 (a native *E. coli*  $\sigma$ ), and *E. coli* DH10 $\beta$ . The results were consistent across both strains and assay conditions, with the chimeric  $\sigma$ s remaining orthogonal to each other and their parents.

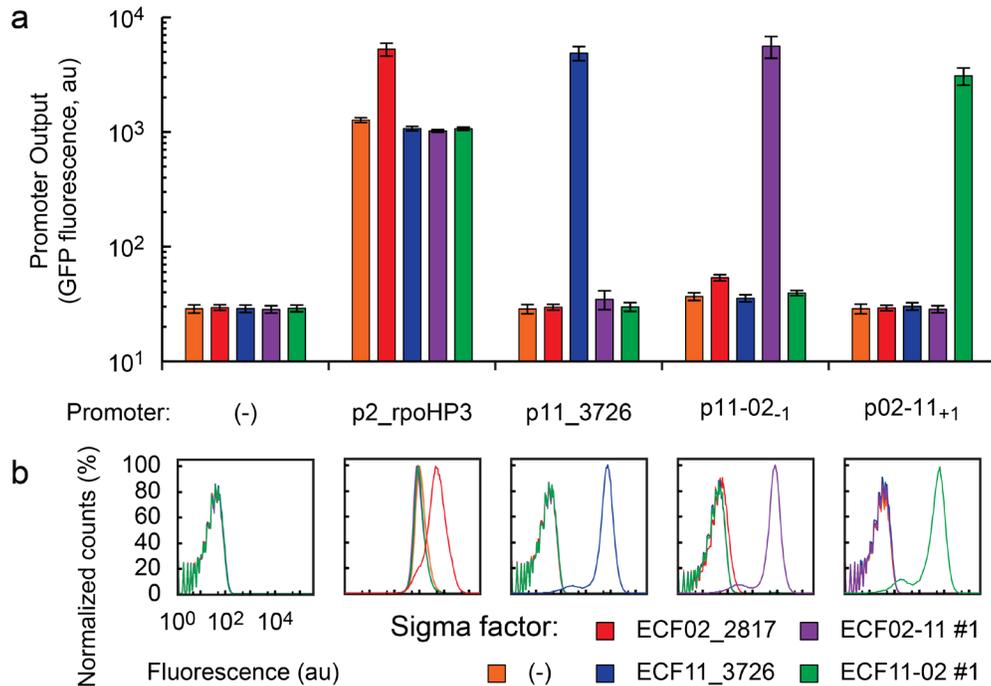
Finally, the toxicity of the chimeric  $\sigma$ s was assayed and compared to the parental  $\sigma$ s. The two parental  $\sigma$ s and the two chimeras with seam #1 were transformed into *E. coli* DH10 $\beta$  cells carrying a negative control reporter plasmid and the transition phase (8hr) OD measured after induction. The assay shows that neither of the chimeric  $\sigma$ s exhibits the extreme toxicity of ECF02\_2817, one of their parents.



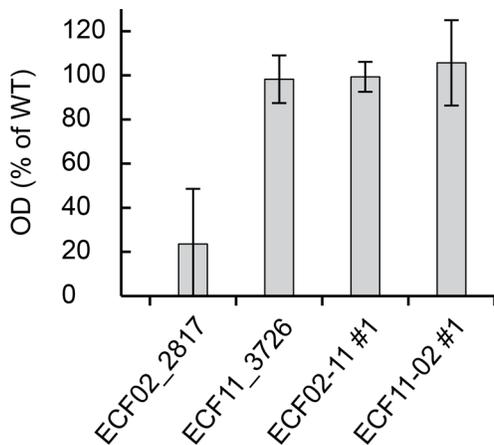
**Figure S19: Testing of chimeric  $\sigma$  and chimeric promoter variants.** Each of the engineered chimeric  $\sigma$ s was tested with each of the variants of its cognate chimeric promoter. Negative controls represent an empty plasmid lacking either a  $\sigma$  coding region, or a  $\sigma$  promoter:reporter cassette. Arrows point to the  $\sigma$  chimera: chimera promoter pairs chosen for orthogonality testing. Assays were performed *in vivo* in *E. coli* CAG22216 cells at 30°C, induced for 6 hours with 10  $\mu$ M IPTG, and promoter activity was measured by flow cytometry. Each bar represents the average promoter activity from three independent assays, and error bars represent one standard deviation.



**Figure S20: Orthogonality testing of chimeric and parental  $\sigma$ s in *E. coli* CAG22216 cells.** (a) Initial orthogonality testing of the chimeric  $\sigma$ s in *E. coli* CAG22216 cells and (b) representative cytometry data. Assays were performed *in vivo* in *E. coli* CAG22216 cells at 30°C, induced for 6 hours with 10  $\mu$ M IPTG, and promoter activity was measured by flow cytometry. (-) controls represent cells carrying a plasmid identical to the chimeric  $\sigma$  expression or reporter plasmids, but with the chimeric  $\sigma$  gene, or promoter/reporter removed. Each bar represents the average promoter activity from three independent assays, and error bars represent one standard deviation. Histograms show representative flow cytometry distributions for each promoter in the presence of all of the  $\sigma$ s.



**Figure S21:** **Orthogonality testing of chimeric and parental  $\sigma$ s in *E. coli* DH10 $\beta$  cells.** (a) Promoter output data corresponding to Figure 3G. Each bar represents the mean fluorescence from three independent assays measured after 6 hours of induction with 10  $\mu$ M IPTG. Error bars show the standard deviation. (b) Representative cytometry distributions from the assay showing each promoter in the presence of all of the  $\sigma$ s.



**Figure S22:** **Toxicity of chimeric  $\sigma$ s.** Plasmids expressing ECF02\_2817, ECF11\_3726, the two most active chimeras, and a negative control with no  $\sigma$  were transformed into DH10 $\beta$  cells, induced with 10  $\mu$ M IPTG, and their OD600 measured after 8 hours. Each bar shows the mean OD from four independent assays normalized to the OD of the cells containing the negative control. Error bars show one standard deviation.

## V. Plasmids

A 4-plasmid system was used for expressing the  $\sigma$ , promoter and anti- $\sigma$  libraries (Figure S23). Plasmid pN565 encodes an IPTG-inducible low processivity T7 RNA polymerase enzyme. This was used to weakly express the  $\sigma$  library under control of a T7-regulated promoter encoded on the pVRa plasmid series. The pVRb plasmid series carries the  $\sigma$ -dependent promoters fused to the fluorescent reporter, superfolder GFP<sup>24</sup>. The pVRc plasmid series encodes the anti- $\sigma$  library under control of HSL. Plasmid modifications were performed using Type II restriction enzyme cloning, PCR and one-step isothermal DNA assembly<sup>25</sup>. The  $\sigma$  and anti- $\sigma$  gene libraries were codon optimized for *E. coli* K12 MG1655, constructed by gene synthesis and assembled into their parent vectors by GeneArt, Life Technologies.

Plasmid pN565 (incW (2-3 copies)<sup>25</sup>, SpecR) is a variant of the low processivity T7 RNA polymerase expression vector, pN249<sup>26</sup> and is tightly regulated by IPTG. The plasmid encodes T7 RNAP with a GTG initiation codon for low translation, an N-terminal degradation tag and the active site mutation R632S. T7 RNAP is expressed from a weak RBS sequence tuned to 50 units using the RBS calculator<sup>27</sup> and a modified P<sub>tac</sub> promoter with a symmetrical LacO operator sequence (aattgtgagcgctcacaatt), enabling near complete promoter repression in the absence of IPTG. The plasmid also encodes LacI.

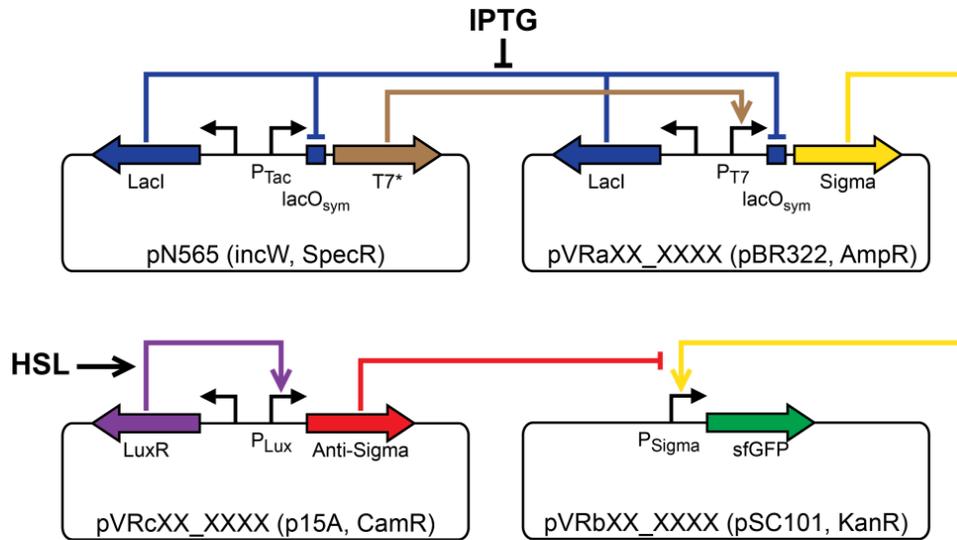
Plasmid series pVRa (pBR322 (15-20 copies/cell)<sup>28</sup>, AmpR) expresses the  $\sigma$  library from a T7-lacO promoter. The plasmids are derived from pET15b (Novagen) in which the thrombin cleavage site was replaced with a PreScission protease cleavage site. The series encodes codon optimized  $\sigma$  genes on *NdeI-HindIII* fragments in frame with an N-terminal His<sub>6</sub> tag and intervening PreScission site. The plasmids and amino acid sequences of the  $\sigma$ s are listed in Supplementary Table S1.1.

Plasmid series pVRb (SC101 (~5 copies/cell)<sup>28</sup>, KanR) carries the  $\sigma$ -dependent promoter library fused to superfolder GFP (*sfgfp*)<sup>24</sup>. The plasmids are derived from the GFP expression vector, pUA66<sup>29</sup>, in which the reporter gene *gfpmut2* was replaced with *sfgfp* on a *BamHI-PstI* fragment. Promoter sequences from -60 to +20 with respect to the transcription start site were inserted upstream of *sfgfp* into the *BbsI-BamHI* sites of pVRb (the 5' end of the -10 motif was assumed to be at position -10). For each promoter, DNA fragments were assembled from 4 overlapping 45-mer DNA oligos that corresponded to native promoter sequence, and 2 flanking vector specific oligos. The oligos were assembled by PCR to generate 120 bp fragments in which the 80 nt promoter sequence is flanked by 20 nt of vector sequence. The fragments were gel purified and assembled into purified pVRb *BbsI-BamHI* vector using one-step isothermal DNA assembly. The plasmids and promoter sequences are listed in Supplementary Table S1.2.

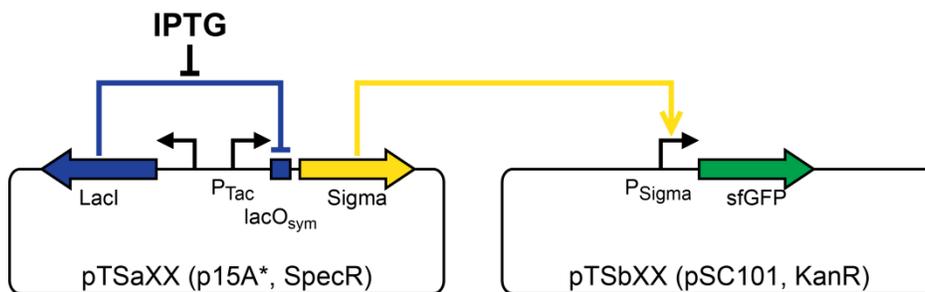
Plasmid series pVRc (p15a (10-12 copies/cell)<sup>28</sup>, CmR) expresses the anti- $\sigma$  library from a HSL-regulated P<sub>lux</sub> promoter. The plasmids contain *cat* and *LuxR* under constitutive control, and replicate via a p15a origin. The plasmids and amino acid sequences of the anti- $\sigma$ s are listed in Supplementary Table S1.3.

A 2-plasmid system was used to test the chimeric  $\sigma$ s and their cognate promoters (Figure S24). Plasmid series pTSaXX (p15a\*, SpecR) expresses parental (ECF02\_2817, ECF11\_3726) and chimeric  $\sigma$ s under the control of a modified P<sub>tac</sub> promoter with a symmetrical LacO operator sequence. These plasmids were derived from pSB3C5<sup>30</sup>, and contain a mutation in the origin that appears to cause them to be maintained at a higher copy number than wild-type p15a. Plasmid series pTSbXX (pSC101, KanR) contains parental and chimeric  $\sigma$ -dependent promoters driving expression of *sfgfp*. These plasmids are very similar to plasmid series pVRb, with only the promoter region varying. All construction of these plasmid series was done with one-step isothermal DNA assembly or PCRs and blunt ligations.

pVRa, pVRb, and pVRc plasmids containing active  $\sigma$ s, anti- $\sigma$ s, and promoters will be made available upon request at Addgene ([http://www.addgene.org/Christopher\\_Voigt/](http://www.addgene.org/Christopher_Voigt/)).



**Figure S23:** *Plasmids used for  $\sigma$  and anti- $\sigma$  characterization.* Low processivity T7 RNA polymerase (T7\*) is expressed from pN565 using an IPTG-inducible P<sub>Tac</sub> promoter with a symmetric lac operator (lacO<sub>sym</sub>). T7\* is used to express the  $\sigma$  library via a T7 and IPTG induced promoter (consisting of the P<sub>T7</sub> promoter sequence followed by a symmetric lac operator) from the pVRaXX\_XXX plasmid series.  $\sigma$ -dependent promoters (P <sub>$\sigma$</sub> ) are carried on the pVRbXX\_XXX plasmid series fused to superfolder gfp. The anti- $\sigma$  library carried on the pVRcXX\_XXX plasmid series is under HSL control via the P<sub>Lux</sub> promoter. XX\_XXX in each of the library names represents which anti- $\sigma$  /  $\sigma$  / or  $\sigma$ -dependent promoter the plasmid carries. For example, the set pVRa20\_992, pVRb20\_992, pVRc20\_992 carries ECF20\_992, AS20\_992, and pECF20\_992, respectively.



**Figure S24:** *Plasmids used for chimeric  $\sigma$  characterization.* A two plasmid system was used to characterize the chimeric  $\sigma$ s. Plasmid library pTSaXX expresses the parental or chimeric  $\sigma$ s under the control of an IPTG inducible P<sub>Tac</sub> promoter with a symmetric lac operator (lacO<sub>sym</sub>). Plasmid library pTSbXX contains the parental and chimeric  $\sigma$ -dependent promoters driving sfGFP.

## VI. Supplementary Tables

The following tables are included as separate files:

### **Supplementary Table 1 – Libraries and plasmids.**

1.1 –  $\sigma$  library. A summary of the full  $\sigma$  library, including source organism, sequence, and plasmid name.

1.2 – Promoter library. Details of the promoter library, including sequences and plasmid names.

1.3 – Anti- $\sigma$  library. A summary of the anti- $\sigma$  library, including source organism, sequence, and plasmid name.

1.4 – Plasmids. A list of all plasmids used.

### **Supplementary Table 2 – $\sigma$ and anti- $\sigma$ library data.**

2.1 –  $\sigma$  orthogonality. The full map of  $\sigma$  : promoter interactions that is shown in Figure 3e and Supplementary Figure S4. Includes average fold induction values, standard deviations, and the number of replicates for each point.

2.2 –  $\sigma$  titrations. Induction curves of 52 active  $\sigma$ s from the library against their most active target promoter. This data is shown in Supplementary Figure S5.

2.3 – Anti- $\sigma$  orthogonality. The full map of anti- $\sigma$  :  $\sigma$  interactions that is shown in Figure 5c and Supplementary Figure S7. Includes average fold induction values, standard deviations, and the number of replicates for each point.

2.4 –  $\sigma$  growth assays. Three measures of the toxicity of the  $\sigma$  library: exponential growth (Supplemental Figure S8), transition phase OD, and colony size on solid media.

2.5 – Anti- $\sigma$  growth assays. Three measures of the toxicity of the anti- $\sigma$  library: exponential growth (Supplemental Figure S9), transition phase OD, and colony size on solid media.

### **Supplementary Table 3 – Analysis of natural $\sigma$ occurrences.**

3.1 – Analysis of co-occurrence. The data from Tables 3.2 and 3.3 is used to statistically determine whether orthogonal  $\sigma$  subgroups are more likely to naturally be found with other  $\sigma$  subgroups.

3.2 – ECF  $\sigma$  co-occurrence. Data from Staron, et. al.<sup>1</sup> was used to generate a table of how frequently the different subgroups of ECF  $\sigma$ s occur in the same genome.

3.3 – Natural sigma orthogonality. The data from Table 3.2 and our  $\sigma$  orthogonality map was used to determine which subgroups of  $\sigma$ s do not crosstalk with the other subgroups they are found with.

3.4 – Examples of natural  $\sigma$  cascades. Our crosstalk analysis and 3.2 is used to predict a few examples of subgroups of  $\sigma$ s that may form cascades.

## VII. References

1. Staroń, A. *et al.* The third pillar of bacterial signal transduction: classification of the extracytoplasmic function (ECF)  $\sigma$  factor protein family. *Mol. Microbiol.* **74**, 557–581 (2009).
2. Rhodius, V. A., Suh, W. C., Nonaka, G., West, J. & Gross, C. A. Conserved and Variable Functions of the  $\sigma$ E Stress Response in Related Genomes. *PLoS Biol* **4**, e2 (2005).
3. Liu, X., Brutlag, D. L. & Liu, J. S. BioProspector: discovering conserved DNA motifs in upstream regulatory regions of co-expressed genes. *Pac. Symp. Biocomput. Pac. Symp. Biocomput.* 127–138 (2001).
4. Rhodius, V. A. & Mutalik, V. K. Predicting strength and function for promoters of the Escherichia coli alternative sigma factor,  $\sigma$ E. *Proc. Natl. Acad. Sci.* **107**, 2854–2859 (2010).
5. Staden, R. Computer methods to locate signals in nucleic acid sequences. *Nucleic Acids Res.* **12**, 505–519 (1984).
6. Crooks, G. E., Hon, G., Chandonia, J.-M. & Brenner, S. E. WebLogo: A Sequence Logo Generator. *Genome Res.* **14**, 1188–1190 (2004).
7. Salgado, H. *et al.* RegulonDB v8.0: omics data sets, evolutionary conservation, regulatory phrases, cross-validated gold standards and more. *Nucleic Acids Res.* **41**, D203–D213 (2013).
8. Shultzaberger, R. K., Chen, Z., Lewis, K. A. & Schneider, T. D. Anatomy of Escherichia coli  $\sigma$ 70 promoters. *Nucleic Acids Res.* **35**, 771–788 (2007).
9. Yuzenkova, Y., Tadigotla, V. R., Severinov, K. & Zenkin, N. A new basal promoter element recognized by RNA polymerase core enzyme. *EMBO J.* **30**, 3766–3775 (2011).
10. Burr, T., Mitchell, J., Kolb, A., Minchin, S. & Busby, S. DNA sequence elements located immediately upstream of the –10 hexamer in Escherichia coli promoters: a systematic study. *Nucleic Acids Res.* **28**, 1864–1870 (2000).
11. Mitchell, J. E., Zheng, D., Busby, S. J. W. & Minchin, S. D. Identification and analysis of ‘extended –10’ promoters in Escherichia coli. *Nucleic Acids Res.* **31**, 4689–4695 (2003).
12. Rhodius, V. A., Mutalik, V. K. & Gross, C. A. Predicting the strength of UP-elements and full-length E. coli  $\sigma$ E promoters. *Nucleic Acids Res.* **40**, 2907–2924 (2012).
13. Chang, A. C. & Cohen, S. N. Construction and characterization of amplifiable multicopy DNA cloning vehicles derived from the P15A cryptic miniplasmid. *J. Bacteriol.* **134**, 1141–1156 (1978).
14. Nitta, T., Nagamitsu, H., Murata, M., Izu, H. & Yamada, M. Function of the  $\zeta$ E Regulon in Dead-Cell Lysis in Stationary-Phase Escherichia coli. *J. Bacteriol.* **182**, 5231–5237 (2000).
15. Hiratsu, K., Amemura, M., Nashimoto, H., Shinagawa, H. & Makino, K. The rpoE gene of Escherichia coli, which encodes sigma E, is essential for bacterial growth at high temperature. *J. Bacteriol.* **177**, 2918–2922 (1995).
16. Peñas, A. D. L., Connolly, L. & Gross, C. A. SigmaE is an essential sigma factor in Escherichia coli. *J. Bacteriol.* **179**, 6862–6864 (1997).
17. Buchler, N. E. & Cross, F. R. Protein sequestration generates a flexible ultrasensitive response in a genetic network. *Mol. Syst. Biol.* **5**, (2009).
18. Campbell, E. A. *et al.* Crystal Structure of Escherichia coli  $\sigma$ E with the Cytoplasmic Domain of Its Anti- $\sigma$  RseA. *Mol. Cell* **11**, 1067–1078 (2003).
19. Campbell, E. A. *et al.* A Conserved Structural Module Regulates Transcriptional Responses to Diverse Stress Signals in Bacteria. *Mol. Cell* **27**, 793–805 (2007).
20. Murakami, K. S., Masuda, S. & Darst, S. A. Structural Basis of Transcription Initiation: RNA Polymerase Holoenzyme at 4 Å Resolution. *Science* **296**, 1280–1284 (2002).
21. Larkin, M. A. *et al.* Clustal W and Clustal X version 2.0. *Bioinformatics* **23**, 2947–2948 (2007).
22. Rost, B., Yachdav, G. & Liu, J. The PredictProtein server. *Nucleic Acids Res.* **32**, W321–W326 (2004).
23. Erickson, J. W., Vaughn, V., Walter, W. A., Neidhardt, F. C. & Gross, C. A. Regulation of the promoters and transcripts of rpoH, the Escherichia coli heat shock regulatory gene. *Genes Dev.* **1**, 419–432 (1987).

24. Pédelacq, J.-D., Cabantous, S., Tran, T., Terwilliger, T. C. & Waldo, G. S. Engineering and characterization of a superfolder green fluorescent protein. *Nat. Biotechnol.* **24**, 79–88 (2005).
25. Gibson, D. G. *et al.* Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nat. Methods* **6**, 343–345 (2009).
26. Temme, K., Hill, R., Segall-Shapiro, T. H., Moser, F. & Voigt, C. A. Modular control of multiple pathways using engineered orthogonal T7 polymerases. *Nucleic Acids Res.* (2012). doi:10.1093/nar/gks597
27. Salis, H. M., Mirsky, E. A. & Voigt, C. A. Automated design of synthetic ribosome binding sites to control protein expression. *Nat. Biotechnol.* **27**, 946–950 (2009).
28. Sambrook, J. J. & Russell, D. D. W. *Molecular cloning: a laboratory manual. Vol. 2.* (CSHL Press, 2001).
29. Zaslaver, A. *et al.* Just-in-time transcription program in metabolic pathways. *Nat. Genet.* **36**, 486–491 (2004).
30. Shetty, R., Endy, D. & Knight, T. Engineering BioBrick vectors from BioBrick parts. *J. Biol. Eng.* **2**, 5 (2008).