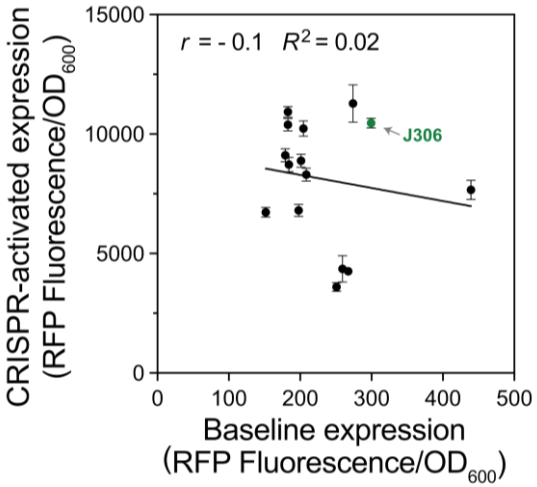
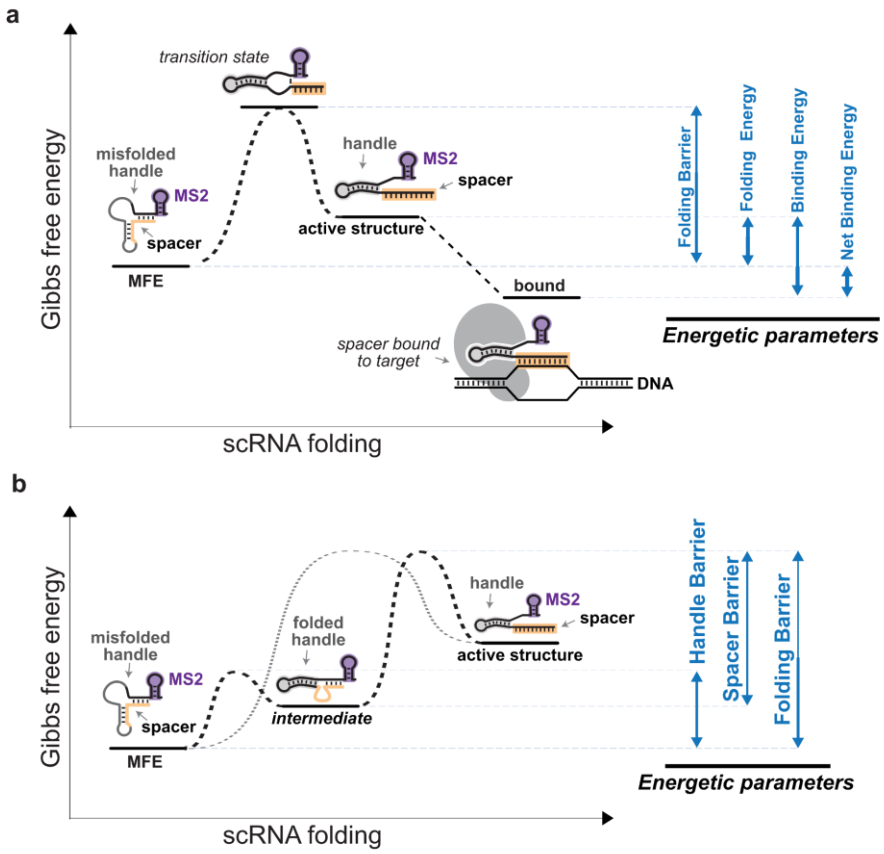
Guide RNA structure design enables combinatorial CRISPRa programs for biosynthetic profiling

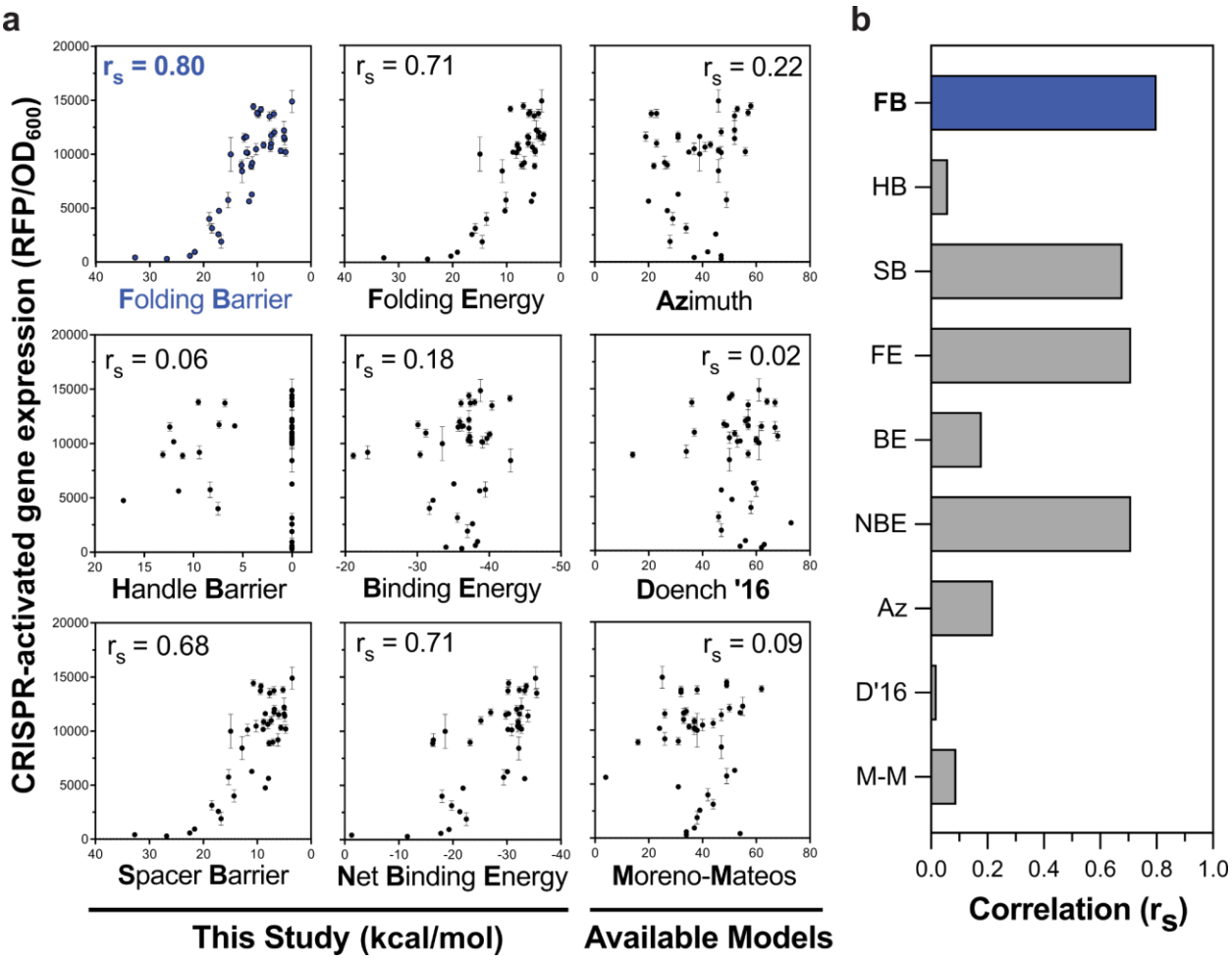
Fontana *et al.*



Supplementary Figure 1. Small differences in baseline expression in synthetic promoters with different scRNA target sites do not explain CRISPR-activated expression level. The CRISPR-activated gene expression of the synthetic promoters from Figure 2B is plotted against their baseline, unactivated expression. The J3 promoter was previously described[1]; each of the 14 other promoters contains a randomly selected scRNA target site and is targeted by a scRNA with complementary target sequence. Values on the *x*-axis represent the baseline expression of each promoter, obtained by measuring the Fluorescence/OD_{600} of strains harboring each synthetic promoter and expressing an off-target scRNA. Values on the *y*-axis depict the Fluorescence/OD_{600} of strains harboring each synthetic promoter and expressing the matching scRNAs, as in Figure 2B. The green dot indicated by the arrow represents a strain harboring the previously described J3 promoter and expressing its cognate J306 scRNA. r and R^2 represent, respectively, the Pearson correlation coefficient and the coefficient of determination for the linear fit between baseline expression and CRISPR-activated expression. Values represent the average ± standard deviation calculated from n = 3 biologically independent samples. Source data are provided as a Source Data file.

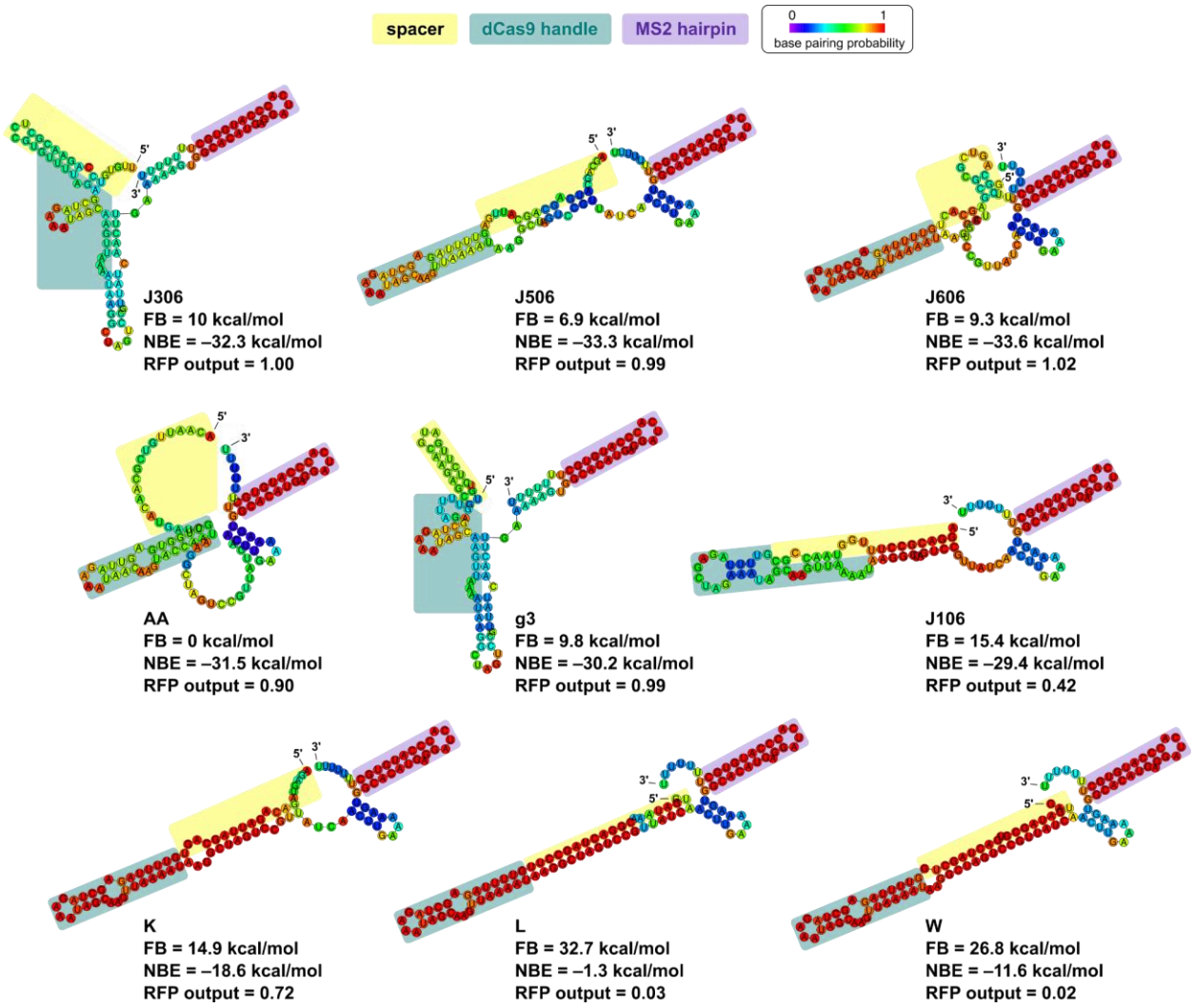


Supplementary Figure 2. Kinetic and thermodynamic parameters provide a quantitative description of scRNA folding. a Kinetic and thermodynamic parameters computationally analyze scRNA folding. The *y*-axis represents the Gibbs free energy of representatives states in scRNA folding. The *x*-axis represents the scRNA folding coordinate. The MFE state indicates the Minimum Free Energy structure adopted by an scRNA, which in this example contains a misfolded dCas9-binding handle and a sequestered spacer. The active structure indicates the correctly folded scRNA in which the spacer is unstructured and the dCas9-binding handle and MS2 hairpin are correctly folded. Folding Barrier represents the free energy barrier that must be overcome to fold the scRNA from the MFE state to the active structure, and is calculated as the free energy difference between the MFE state and the transition state to the active structure. The Folding Energy represents the free energy difference between the MFE state and the active structure. The bound state illustrates the binding of the scRNA to the target DNA and is approximated as the free energy of the scRNA in the active structure bound to a RNA 20-mer complementary to the spacer sequence. The Binding Energy represents the free energy difference between the active state and the bound state. The Net Binding Energy represents the free energy difference between the MFE state and the bound state. **b** Additional metrics were investigated to understand the role of kinetics in scRNA folding. An intermediate state between the MFE state and the active structure was defined in which the dCas9-binding handle is correctly folded. Handle Barrier represents the free energy barrier between the MFE state and the intermediate state with the dCas9-binding handle folded. Spacer Barrier represents the free energy barrier between the intermediate state and the active structure.

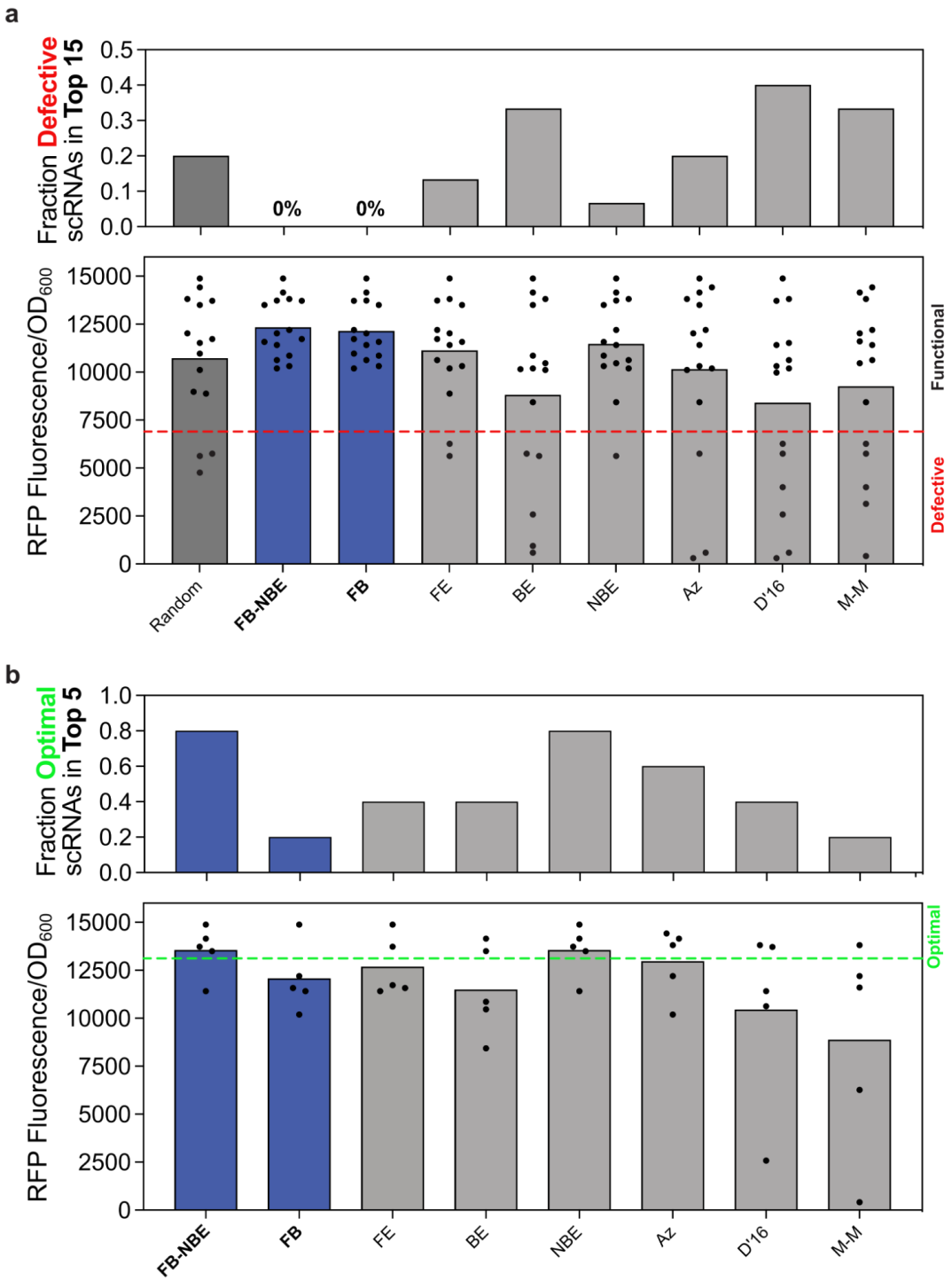


Supplementary Figure 3. Folding barrier outperforms popular gRNA selection models for predicting scRNA activity in CRISPRa experiments. **a** Comparison of kinetic and thermodynamic parameters from this study with three popular gRNA selection models for scRNA design. Individual plots describe the performance of the indicated parameter or model. A schematic of each parameter from this study is provided in Supplementary Figure 2. The gRNA selection models tested are Azimuth, Doench '16[2] and Moreno-Mateos[3], which were the three best performing models from the set of models found in the CRISPOR webserver[4]. Each plot contains 39 data points, corresponding to the scRNAs targeting each of the synthetic promoters described in Figure 2d. In each plot, the *y*-axis shows the CRISPR-activated RFP expression of strains harboring each synthetic promoter and expressing the matching scRNA. The *x*-axes represent different computed parameters for scRNA function. For the parameters described in this study, the *x*-axis is kcal/mol on a reverse scale (high values on the left, low values on the right). The reverse scale places scRNAs predicted to be more active to the right side of each plot. The depicted parameters from this study are Folding Barrier, Handle Barrier, Spacer Barrier, Folding Energy, Binding Energy, and Net Binding Energy. Among those, Folding Barrier, Folding Energy, and Net Binding Energy were most correlated with expression. For the Azimuth, Doench '16, and Moreno-Mateos models, values on the *x*-axis represent scores assigned by each model when the scRNA target site for each of the synthetic promoters described in Figure 2D is used as

an input. Again, more active predictions should be to the right side of the plot. r_s represents the Spearman's rank order correlation coefficient between the predicted scRNA activity based on the *x*-axis values and experimental CRISPR-activated expression. **b** Correlation coefficients between predicted scRNA activity and experimental results. Bars represent the Spearman's rank order correlation coefficient between predicted scRNA activity and the corresponding experimental CRISPR-activated expression for each parameter depicted in panel **a**. Labels are abbreviations of the metrics from this study and the gRNA selection models. FB: Folding Barrier, HB: Handle Barrier, SB: Spacer Barrier, FE: Folding Energy, BE: Binding Energy, NBE: Net Binding Energy, Az: Azimuth. D'16: Doench '16, M-M: Moreno-Mateos. Values in panel a represent the average ± standard deviation calculated from $n = 3$ biologically independent samples. Source data are provided as a Source Data file.

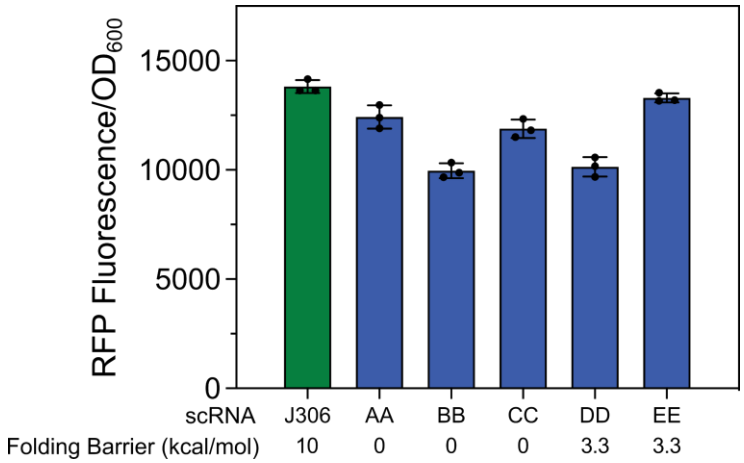


Supplementary Figure 4. Select scRNA MFE structures. Minimum Free Energy (MFE) structure predictions were generated using the RNAfold WebServer[5], and the resulting dot-bracket notations are also included in Supplementary Table 3. The scRNAs used in this study's combinatorial expression library—J306, J506, and J606—are shown in the top row, while the rest are roughly in ascending order of Folding Barrier and descending order of RFP output. Colors indicating base pairing probability near 0 indicate structures that probably require little energy to break and reform into an MFE-adjacent structure, as suggested by low Folding Barriers. High FBs, accompanied by low output, are often the result of strong hybridization of the spacer sequence to an internal sequence, exemplified by K, L, and W.

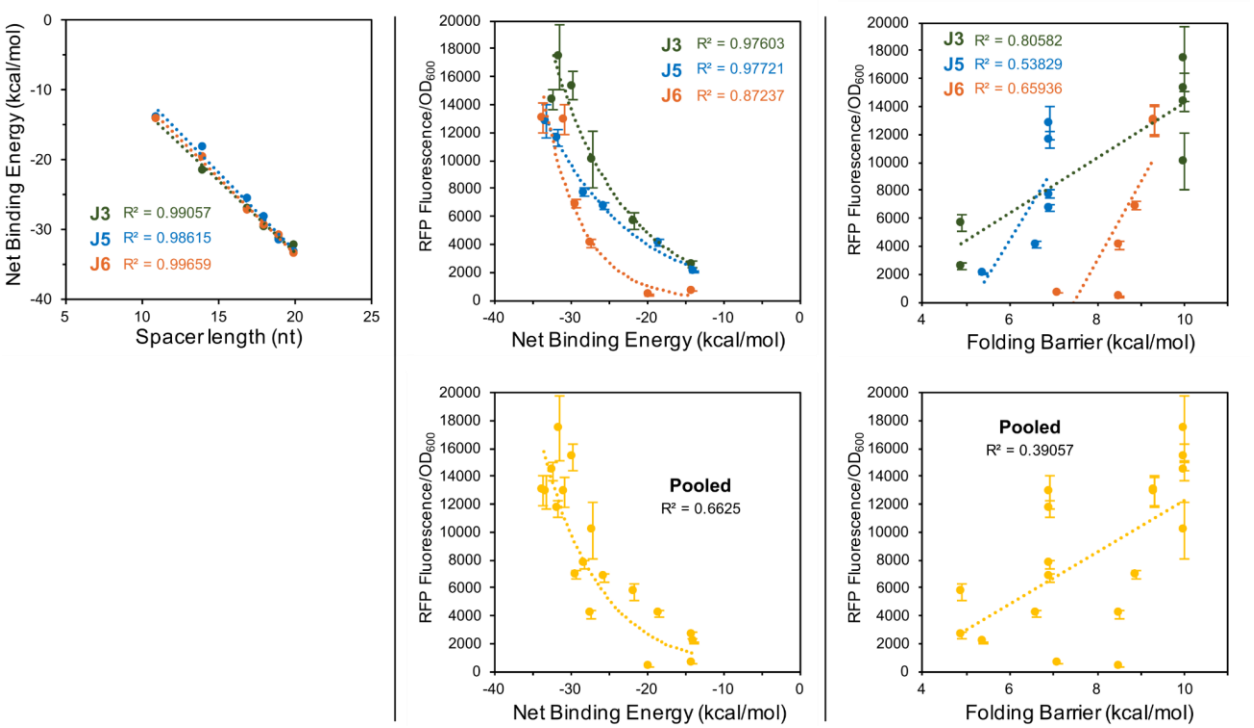


Supplementary Figure 5. Selecting scRNAs based on Folding Barrier and Net Binding Energy parameters avoids defective scRNAs and enriches for optimal scRNAs. a Selecting scRNAs with a low Folding Barrier avoids defective scRNAs. A schematic of each parameter from this study is provided in Supplementary Figure 2. Of the 39 scRNAs described in Figure 2D, the 15 scRNAs predicted by each parameter to perform the best are shown here. For the parameters described in this study, better performance is predicted by lower Gibbs free energy (Net Binding

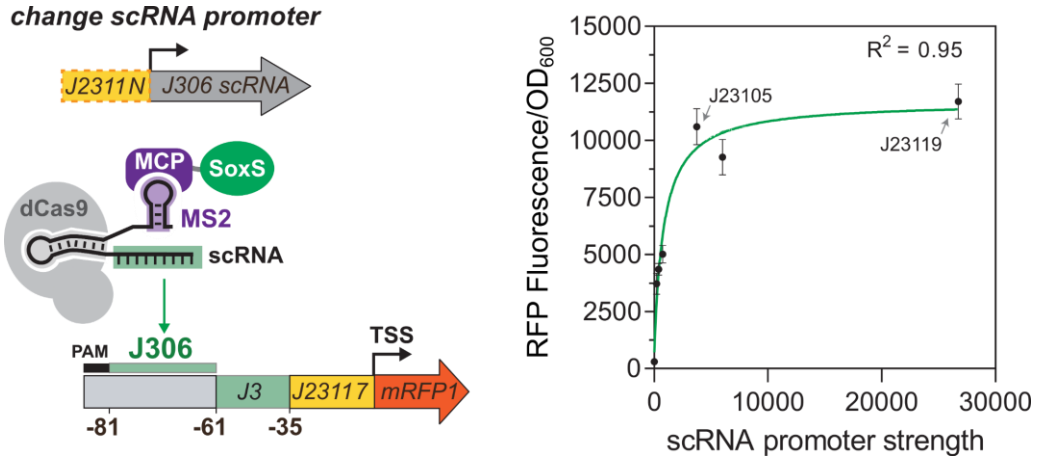
Energy, Folding Energy, Binding Energy) or smaller free energy barriers (Folding Barrier). FE-NBE was obtained by first screening the scRNAs for Folding Barriers ≤10 kcal/mol, and then ranking the remaining scRNAs based on their Net Binding Energy (Supplementary Method 1). For the three available gRNA selection models, the highest-scoring 15 scRNAs are shown here[2,3] (using the scRNA target site as an input). The top graph shows the fraction of defective scRNAs found in the top 15 scRNAs selected by each parameter. Defective scRNAs are defined as those used in strains that yielded ≤50% of J306-level CRISPR-activated expression (Fluorescence/OD_{600}). The bottom graph shows the CRISPR-activated expression of strains expressing the top 15 scRNAs selected by each parameter. Dots represent individual strains, while bars represent the average expression from each set. The dashed red line shows the threshold for defective scRNAs. **b** Screening scRNAs with low Folding Barrier and minimizing Net Binding Energy yields optimal scRNAs. The top graph shows the fraction of optimal scRNAs found in the top 5 scRNAs selected by each parameter. scRNAs were selected in the same way described in panel **a**. Optimal scRNAs are defined as those used in strains that yielded ≥95% of J306-level CRISPR-activated expression (Fluorescence/OD_{600}). The bottom graph shows the CRISPR-activated expression of strains expressing the top 5 scRNAs selected by each parameter. FB-NBE and NBE happened to select the same scRNAs in this case, because they were all under the FB threshold. FB still has value for avoiding defective scRNAs in other cases, though, as seen in panel **a**. Inclusion of NBE ranking in FB-screened subsets does yield slightly more active scRNAs in this case (see Supplementary Method 1 for additional discussion). Dots represent individual strains, while bars represent the average expression from each set. The dashed green line shows the threshold for optimal scRNAs. FB: scRNA Folding Barrier, NBE: Net Binding Energy, HB: Handle Barrier, SB: Spacer Barrier, FE: Folding Energy, BE: Binding Energy, Az: Azimuth. D'16: Doench '16, M-M: Moreno-Mateos. Source data are provided as a Source Data file.



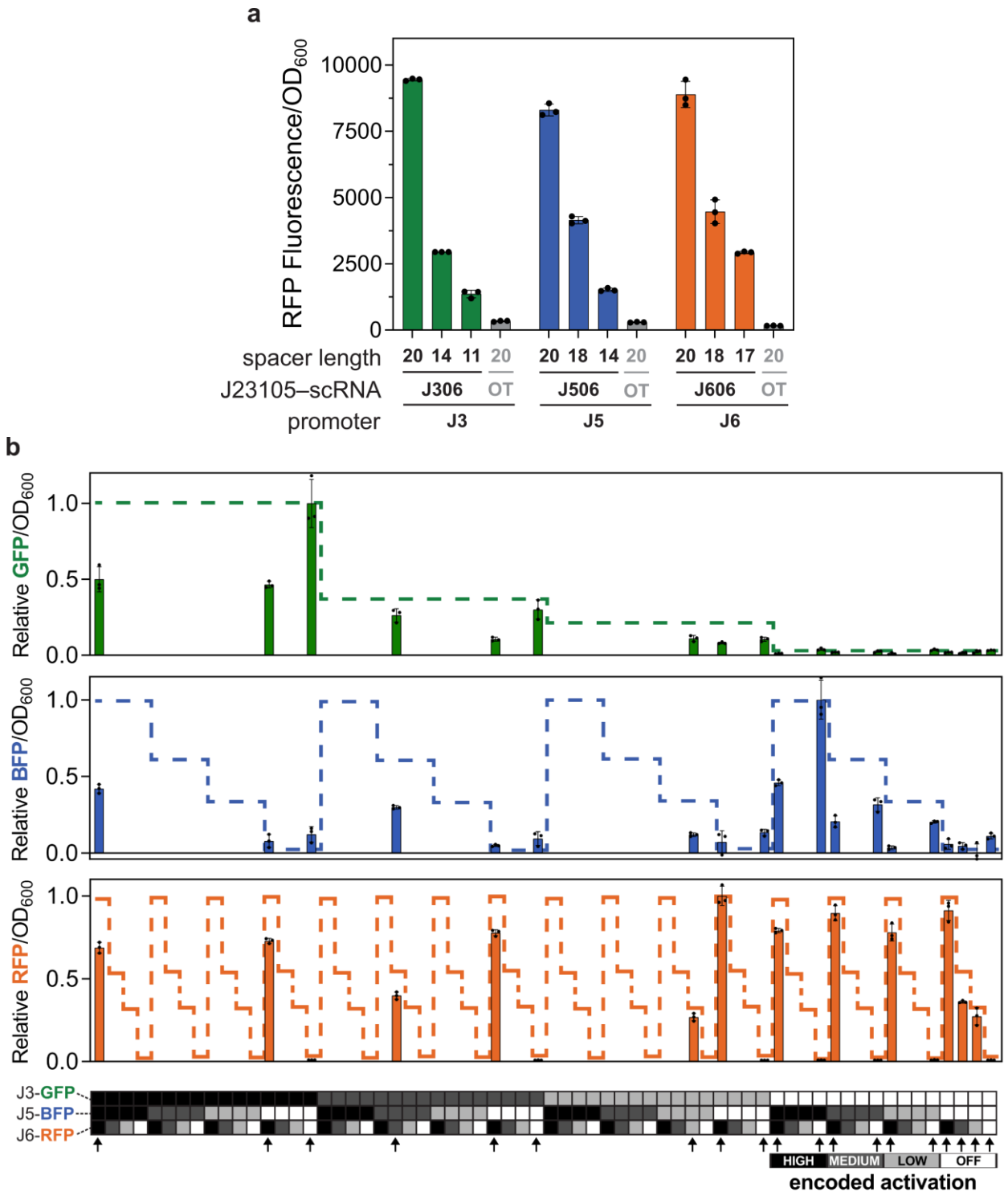
Supplementary Figure 6. Forward engineering of scRNAs with minimal Folding Barriers. As a test of Folding Barrier's predictive power for forward design of highly functional scRNAs, five additional scRNAs were specifically designed to have Folding Barriers below 5 kcal/mol, and as close to 0 kcal/mol as possible. To preserve the structure of the dCas9-binding handle, we had to provide additional stability to the handle by altering the sequence relative to wild-type (Supplementary Table 3). For AA, BB, and CC, the MFE was predicted to be the active structure, hence the zero barrier. CRISPRa using these scRNAs was very active for all five, as expected from Folding Barriers lower than the 10 kcal/mol threshold, but interestingly they were not the highest-performing scRNAs in the overall set included in this work. This relates to the sigmoidal fit proposed in Figure 2d, where Folding Barrier loses some of its predictive power below the threshold of FB ≤ 10 kcal/mol. Values represent the average calculated from n = 3 biologically independent samples. Source data are provided as a Source Data file.



Supplementary Figure 7. Parameter-output correlations in J3/J5/J6 truncation response. Despite strong correlation between spacer length after truncation and Net Binding Energy (NBE) (left), correlation between NBE and RFP output from the J3, J5, and J6 promoters (center) is not strong enough for precise prediction of truncation effects on output. When individually fit to exponential functions (center top), the relatively sensitive truncation response of J6 results in a different function and a poorer correlation than that of J3 or J5. This means that when all the points are pooled into one correlation (center bottom), the overall correlation is poor. For the purpose of forward scRNA design, strong predictive power of a computational parameter would have come from a good correlation between that parameter and output, across all tested promoters. Therefore, our ability to predict spacer truncation effects on output remains poor, even though NBE correlates well with spacer length. Folding Barrier (FB) response to truncation (right) is much more stepwise than the NBE response, and does not correlate to output in a useful way. In the absence of computational prediction of truncation response, experimental mapping of that response (as in Figure 3c) remains necessary. *y*-axis values represent the average ± standard deviation calculated from $n = 3$ biologically independent samples. Source data are provided as a Source Data file.

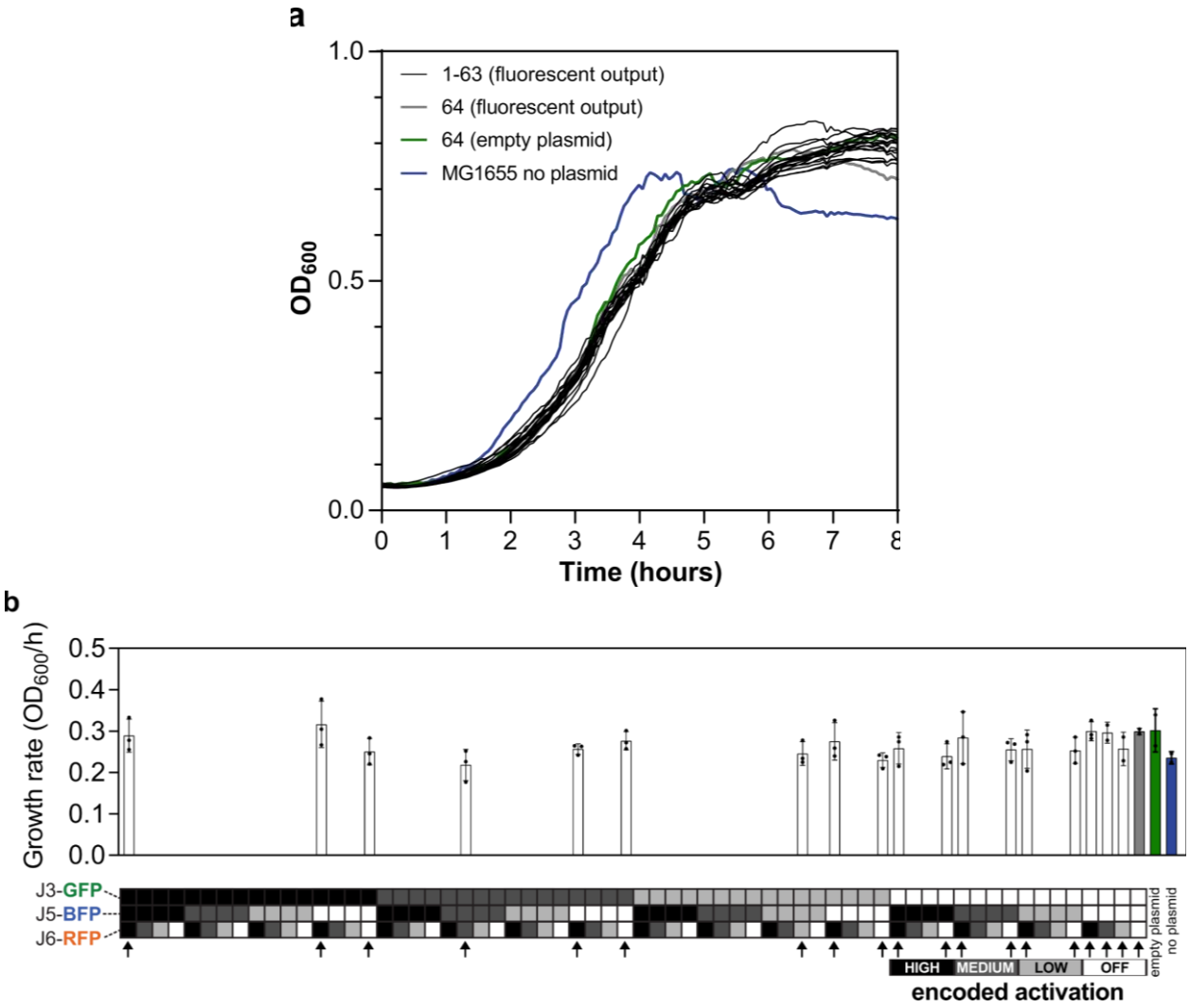


Supplementary Figure 8. Activation dependence on scRNA expression level. CRISPR-activated fluorescence resulting from a panel of six promoter strengths directing scRNA expression is compared to expression level of that scRNA. All RFP expression was directed by the J3 synthetic promoter; fluorescence changes result from various amounts of J306 scRNA. Reducing scRNA expression by ~7-fold results in almost no defect in CRISPRa activity compared to the stronger, commonly-used J23119 promoter. Especially in multiple-scRNA systems, use of J23119 could therefore represent wasted expression capacity. Below the amount of scRNA directed by the J23105 promoter, output fluorescence from J3 drops dramatically, so J23105 was chosen as the minimum promoter strength directing expression of the triple-scRNA library. Promoter strength on the *x*-axis is quantified by the RFP fluorescence observed if that promoter was directly expressing mRFP1. These values are derived from Figure 3a in a previous report[1]. Green line represents a dose-response function fit to the data. Values on the *y*-axis represent the average ± standard deviation calculated from n = 3 biologically independent samples. Source data are provided as a Source Data file.

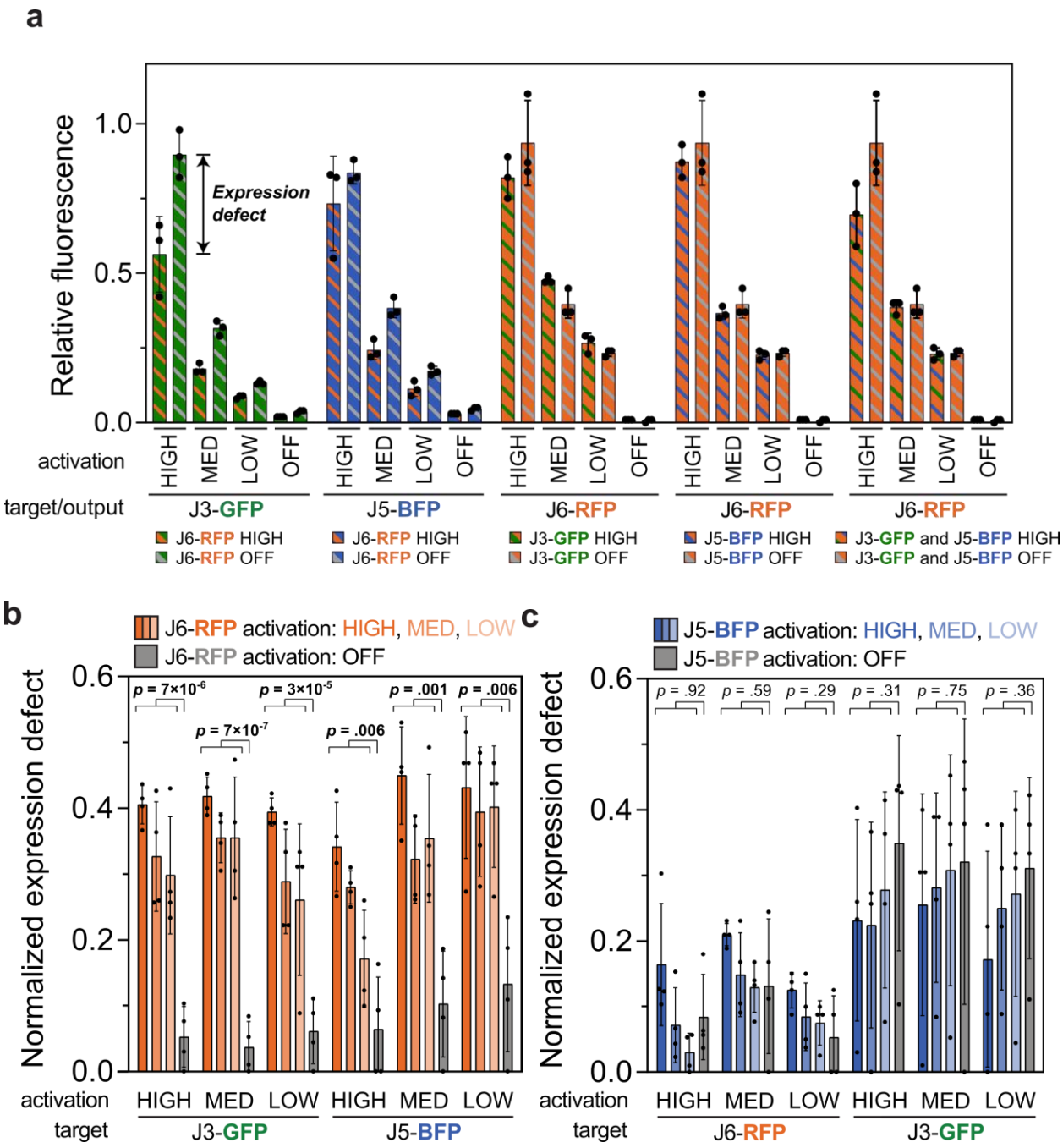


Supplementary Figure 9. Truncation response to lower scRNA expression level and plate reader output from triple-fluorescent reporter. **a** Effects of scRNA expression level on truncation-based tuning. Adopting the J23105 promoter for expressing scRNAs in the triple-scRNA library format, here we verify that tuning CRISPR activation by truncating spacer lengths still results in the previously calibrated RFP output response (Figure 3c). Only truncation levels defined as high, medium, and low for the combinatorial scRNA library are included here. Bars

represent RFP Fluorescence/OD_{600} of strains harboring the J3, J5, or J6 synthetic promoters controlling RFP and also expressing the J306, J506, and J606 scRNAs truncated to different lengths. Grey bars represent the baseline expression of the promoters and were obtained by measuring the Fluorescence/OD_{600} of strains harboring each promoter and expressing an off-target scRNA. Values represent the average ± standard deviation calculated from n = 3 biologically independent samples. The full sequence of the J3, J5, and J6 promoters is described in Supplementary Table 5 and Supplementary Software 1. **b** Bulk measurement of scRNA truncation effects on triple-fluorescent output in a subset of the combinatorial scRNA library. For the strains indicated by arrows on the *x*-axis heatmap, output of each fluorescent protein was measured in bulk on a plate reader. This measurement is in contrast to the per-cell flow cytometry measurement in Figure 4b, and in this case is normalized by OD_{600}. BFP signal appears more difficult to accurately measure on the plate reader compared to flow cytometry, but even so, correlations are high between the normalized flow cytometry and normalized plate reader values. Pearson correlation coefficients are r = 0.9822 for GFP, r = 0.913 for BFP, r = 0.9884 for RFP, and r = 0.973 overall. Values represent the average ± standard deviation calculated from n = 3 biologically independent samples. Dashed lines are the same as in Figure 4b and represent the Relative Fluorescence/OD_{600} of strains harboring only one of the three fluorescent reporters and only the cognate scRNA. Source data are provided as a Source Data file.

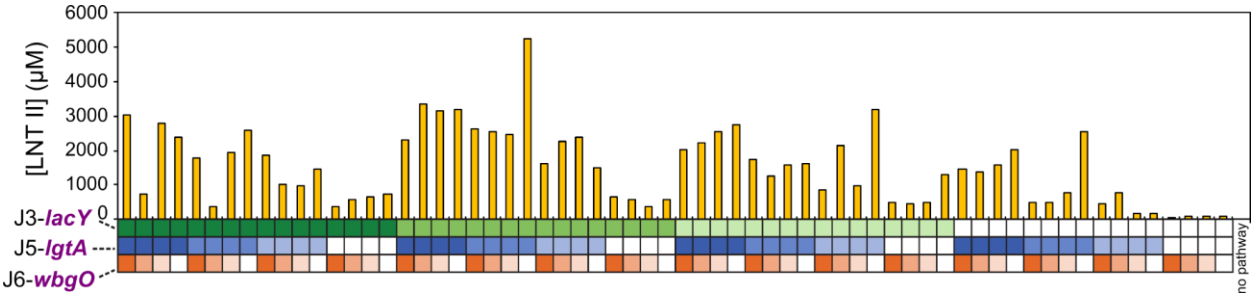


Supplementary Figure 10. Growth effects of triple-fluorescent output expression. **a** Growth rates appear similar across the subset of strains within the combinatorial scRNA library, despite ten strains having high expression of at least one output. The subset shown here is the same subset as in Supplementary Figure 9b. In this growth assay, the effect of plasmid burden (green vs. blue) is apparent, suggesting that output expression burden has minimal growth effect relative to plasmid maintenance burden. Each line is an average of $n = 3$ biologically independent cultures, except strains 62, 63, 64, empty plasmid, and no plasmid ($n = 2$). Error shading is omitted for clarity, but standard deviations are included in the Source Data file. **b** Quantitative comparison of growth rates derived from the curves in panel **a**. The library subset tested here is the same as in Supplementary Figure 9b, again indicated by the arrows on the *x*-axis heatmap. Values represent the average ± standard deviation of growth rate between 3–3.5 hours calculated from $n = 3$ biologically independent cultures. Source data are provided as a Source Data file.

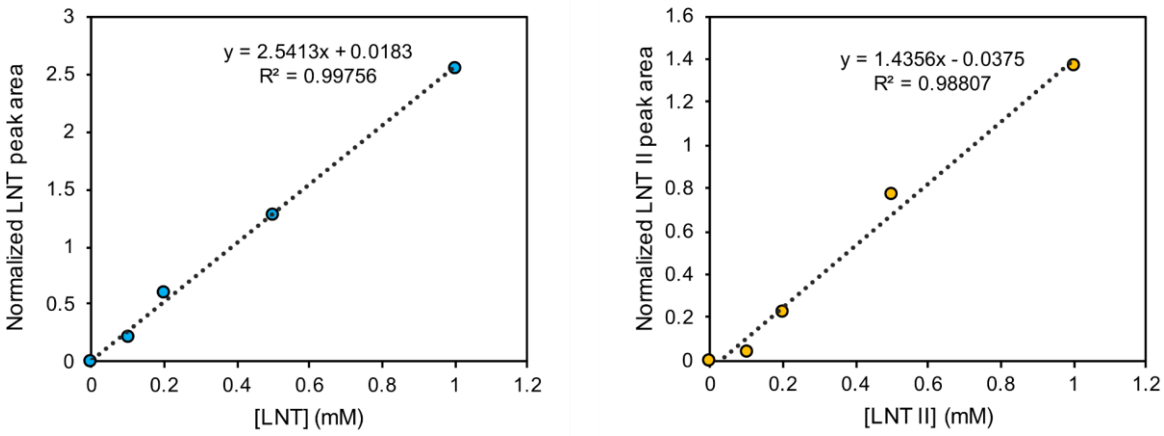


Supplementary Figure 11. RFP expression burden affects simultaneous expression of GFP and BFP. **a** Select strains from the triple-fluorescent combinatorial library often indicate lower non-RFP outputs when RFP is highly expressed. The full scRNA truncation series for each output is shown, with other outputs either expressed highly (20 bp spacer) or minimally (off-target spacer). The expression of high RFP results in the indicated expression defect in GFP (left group) and BFP (second group from left) outputs, but the defect is much smaller or zero when observing effects of high GFP or BFP expression on RFP output (middle and second-from-left groupings, respectively). RFP's relative resilience to expression defect suggests that it has dominant expression burden relative to the other outputs in this set. Only when all three outputs are highly

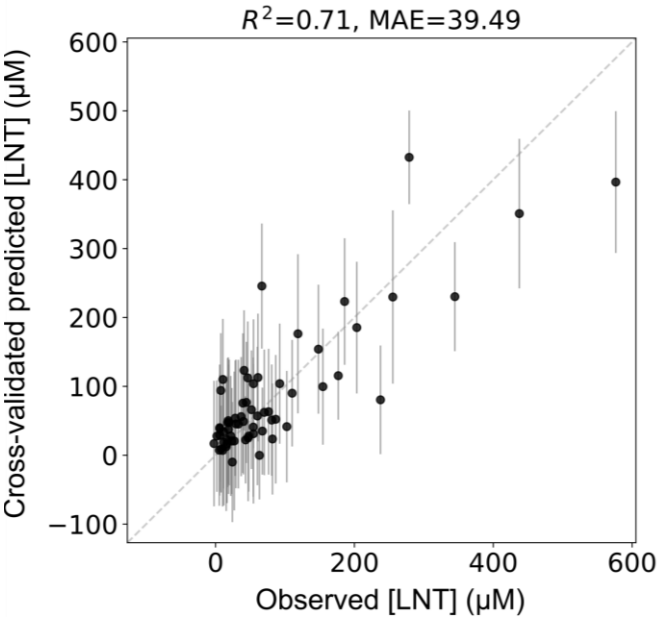
expressed (left bar of rightmost grouping) is a substantial defect in RFP expression observed, suggesting that total combined expression could also play a role at high levels of these outputs. Points represent median flow cytometry values from Figure 4b, and bars represent the average ± standard deviation calculated from $n = 3$ biologically independent samples. **b** Defect in GFP and BFP expression at noted RFP expression levels in the full set of strains expressing the combinatorial multi-scRNA library. Defect is calculated as the reduction in fluorescence within the activation state noted on the *x*-axis (e.g. high, medium, low expression), relative to the strain with the maximum fluorescence for that state. Defects in GFP and BFP expression tend to be significantly higher when RFP is activated (red) than when RFP is not activated (grey), suggesting that RFP has more of an effect on overall expression burden than the other outputs. **c** Defect in RFP and GFP expression at noted BFP expression levels in strains expressing the combinatorial multi-scRNA library. Compared to the RFP effect in panel **b**, BFP activation has little effect on defect in RFP or GFP expression: J5-OFF strains (grey) have as much defect as J5-activated strains (blue). Each point in **b** and **c** is a ratio of two averages of flow cytometry data (Figure 4b), each calculated from $n = 3$ biologically independent samples. Bar values in **b** and **c** represent the average ± standard deviation calculated from 4 such ratios, because there are four activation states for each bar that are not specified by the *x*-axis and legend. Each *p*-value is derived from comparison between the 12 activated strains and the 4 unactivated strains contained in each *x*-axis label. Source data and test statistics are provided as a Source Data file.



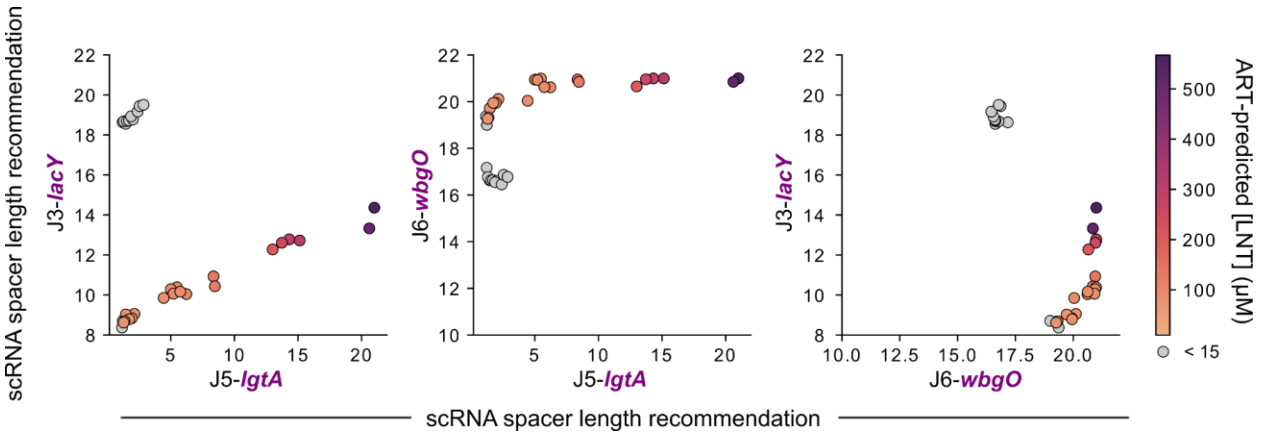
Supplementary Figure 12. LNT II titers across the library. HPLC analysis of supernatant from cultured library members indicates LNT II production levels from the scRNA combinations. The high LNT II titers relative to LNT titers, with relatively little dependence on *wbgO* expression, suggest galactosyltransferase activity as a limiting factor in the pathway. This bottleneck results in accumulation of LNT II and export to the supernatant, where it is inaccessible to WbgO. The *x*-axis heatmap is color coded to indicate the encoded promoter expression for each strain, as in Figure 6B. The 65th strain at right is a no-pathway control culture carrying an empty vector. Refer to Figure 6a for the pathway overview. Source data are provided as a Source Data file.



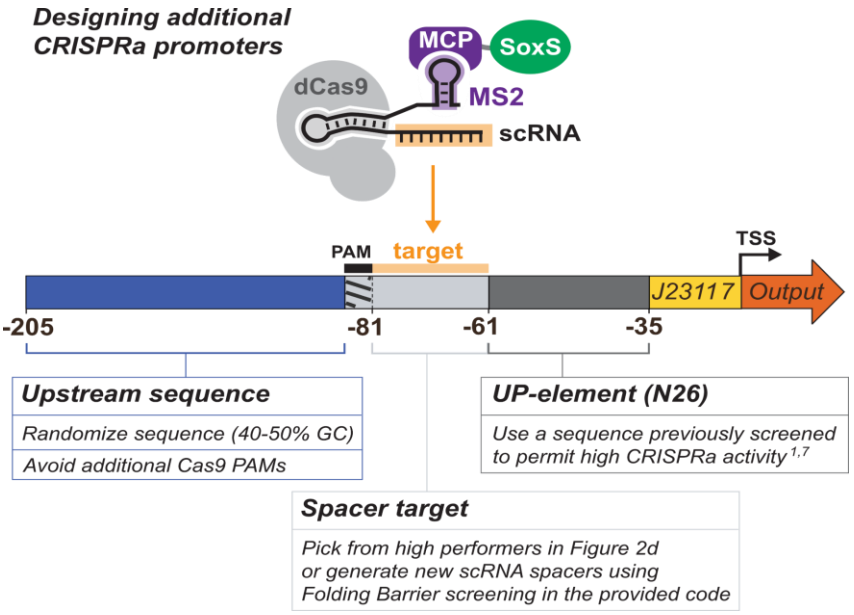
Supplementary Figure 13. Standard curves for HPLC quantification of LNT and LNT II. Peak areas resulting from LNT (10.6 min) or LNT II (11.4 min) were normalized by an endogenous peak (9.1 min) before conversion to molarity.



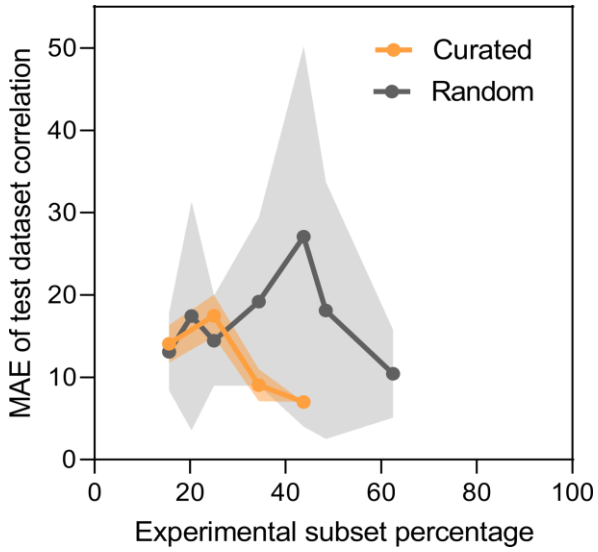
Supplementary Figure 14. ART training and cross-validation for predicting LNT production. Experimental observations of LNT production from the combinatorial scRNA library (*x*-axis) are used to train an Automated Recommendation Tool (ART) model. In the representative training results shown here, we performed 10-fold cross-validation and plotted the predicted LNT production of each data point when it was in the test set (*y*-axis). When trained in this way, ART predictions correlate well with experimental values (R^2 = 0.71) for unseen training data. ART tended to underpredict the titer of the highest performing strains. Generally, retraining on the full dataset is then performed before ART returns strain recommendations in the form of a set of input scRNA spacer lengths and a predicted LNT titer. Error bars indicate the 95% credible interval of the predictive posterior distribution. Source data are provided as a Source Data file.



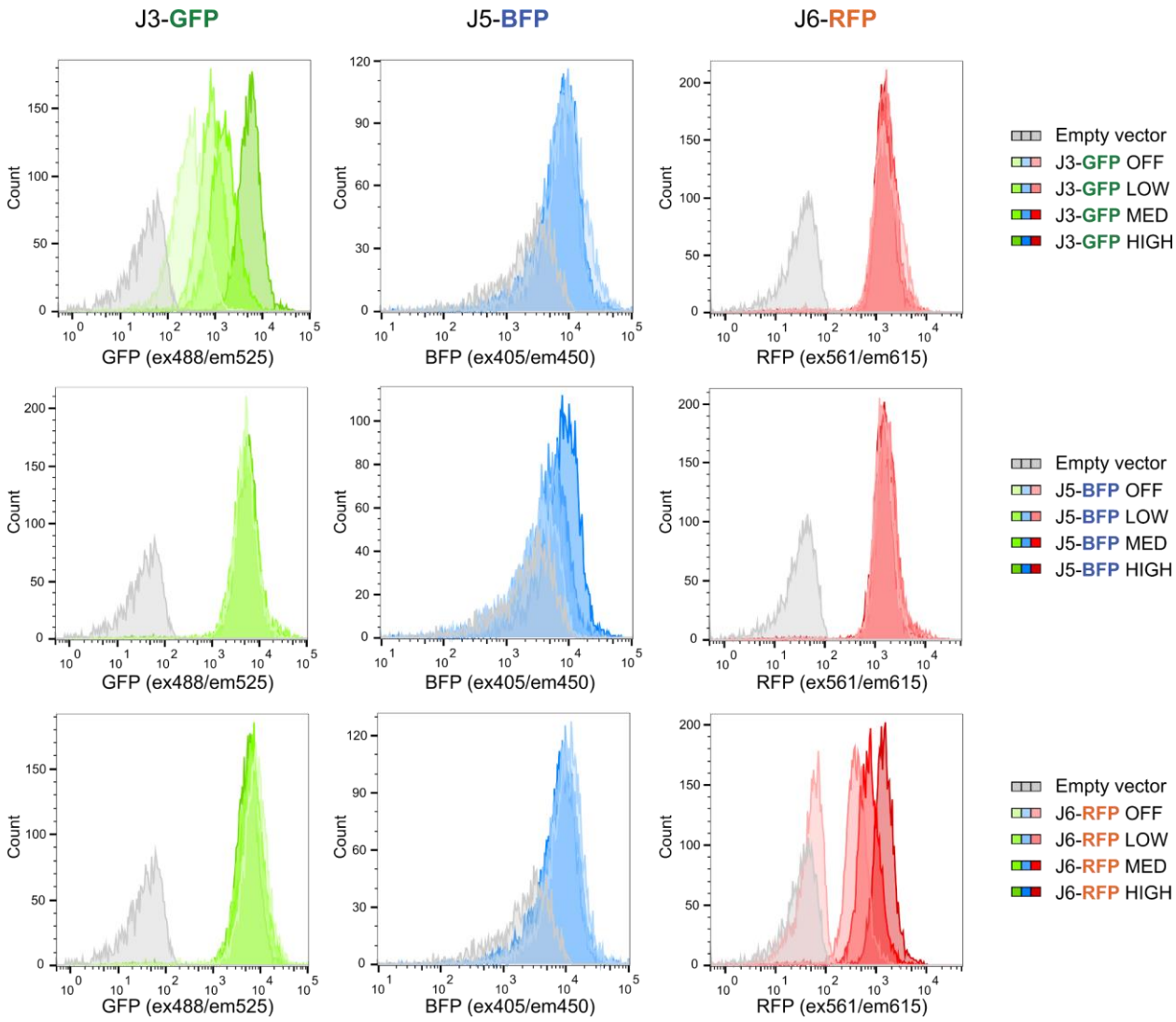
Supplementary Figure 15. Scatter plots of ART recommendation combinations. Each strain recommendation from the computational ART analysis in Figure 6d is plotted to demonstrate the relationship between each channel's spacer length recommendation. As in Figure 6d, the 20 strains with highest predicted LNT titer are highlighted in color on each sub-graph, while the 12 with lowest predicted LNT titer are shown in grey. Points are colored by their predicted LNT concentration: the same predictions as on the *y*-axis in Figure 6d. Source data are provided as a Source Data file.



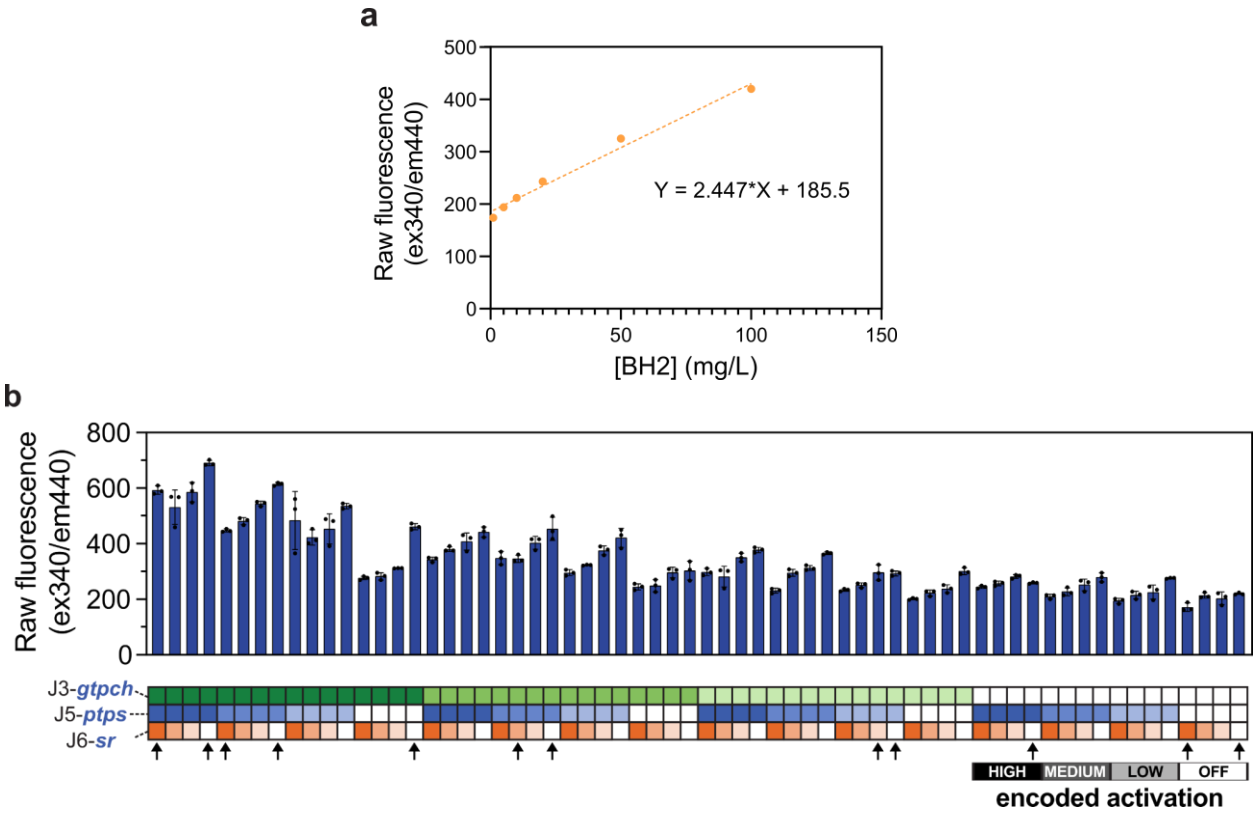
Supplementary Figure 16. Design of additional CRISPRa promoters. This work provides three promoters (J3, J5, and J6) that have been systematically characterized. To construct programs with additional orthogonal promoters, new scRNA-promoter pairs can be designed using modular sequence elements that either are pre-screened (Figure 2d) or can easily be screened from random sequences. CRISPRa efficacy is largely influenced by spacer target, UP-element, and minimal promoter sequence, while orthogonality is supplied by spacer sequence orthogonality. New promoters will require a new spacer target sequence and a new UP-element, but can reuse BBa_J23117 as a minimal promoter. Previously-unused spacer sequences—and therefore their cognate target sequences—can be picked from this work's existing set of spacers (Figure 2d and Supplementary Data 1), 16 of which showed >75% of J306 activity and four of which (g6, g10, g3, and g2) showed J306-equivalent (>98%) activity. Entirely new spacers can be screened from random sequences using the Folding Barrier metric (see accompanying code). An optimal minimal promoter for plasmid-based systems (BBa_J23117) has been described previously.[1,7] Note that single-copy genome integrated constructs may require a stronger minimal promoter. The 26 bp UP-element can be chosen from an existing set of sequences pre-screened for high CRISPRa dynamic range, as described in previous work[1,7] (see Figure 3c and Supplementary Figure 3 in Fontana *et al.*[1], in which this sequence element is called N26; and Figure 2, Figure S5, and Table S1 in Alba Burbano *et al.*[7], in which this sequence element is called UP-element). Choosing a new UP-element for a new CRISPRa promoter is not critical for orthogonality to other promoter-scRNA pairs, but can be useful for minimizing repeated sequences in the system. Finally, we note that including roughly 120 bp of randomized sequence upstream of the spacer target helps reduce repeated sequences in the system, provides proper spacing for intergenic regions, and avoids inadvertent PAM sequences. In our experience, lack of promoter failure when randomizing this upstream sequence suggests that it is not as critical to CRISPRa efficacy as the other promoter components shown here. Once candidate new promoters are designed, they can be characterized using the same workflow shown in Figure 3 of this work. This overall process is similar to how the J5 and J6 promoters were designed following the initial design of the J3 promoter.



Supplementary Figure 17. Degree of experimental sampling influences correlation between observed and predicted LNT titer. Within the full experimental design space of 64 strains, ART predictions present the possibility of experimentally testing only a subset of the design space. The size of such subsets affects the accuracy of modeled production from the rest of the design space, presenting a tradeoff between experimental effort and prediction accuracy. Smaller subsets could result in more uncertainty in the model's predictions (higher MAE), while larger subsets could provide more accurate predictions (lower MAE) at the cost of greater experimental effort. Using the LNT pathway, this trend was apparent in subsets consisting of a small number of specifically-selected strains (curated, orange) but not in randomly-selected subsets (grey). Strategic selection of the subset strain identities can also result in lower prediction uncertainty than most randomly-selected subsets. Beyond the constant inclusion of high-high-high, med-med-med, low-low-low, and OT-OT-OT, strain selection for the curated subsets drew randomly from a set of only the 24 strains that have non-duplicated inputs, e.g. medium-high-low but not medium-medium-low, instead of drawing from the whole set of 64 strains. MAE values are derived from observed vs. predicted correlation plots such as Supplementary Figure 14. This approach potentially allows reasonably accurate model predictions with conveniently low experimental effort, although testing with other pathways should be undertaken. An optimal subset size depends on the complexity of the production landscape within the overall design space, and therefore on the complexity of the pathway itself. For the LNT pathway, for example, carefully choosing and experimentally testing only 22 (34%) of the 64 possible strains still results in reasonably low MAE, but subsets ≤25% result in higher MAE. We do not expect a particular subset size to be optimal for multiple different pathways, because production from different pathways will map differently onto the underlying design space. Different production landscapes will translate to different objective functions in the model, of variable complexity. More complex objective functions will require more training data to allow accurate predictions. Shading represents the standard deviation of MAE values from $n = 3$ different subsets of each subset size. Source data and strain selections are provided as a Source Data file.



Supplementary Figure 18. Representative flow cytometry distributions of fluorescent output strains. Histograms of representative cultures from the triple-fluorescence output combinatorial library (Figure 4b) are shown in GFP, BFP and RFP channels. The first row includes the J306 truncation series while J5 and J6 are highly expressed (20 bp spacers): strains 1 (4464±61 cells after gating), 17 (4528±127 cells), 33 (4624±146 cells), and 49 (4872±244 cells). The second row includes the J506 truncation series while J3 and J6 are highly expressed: strains 1, 5 (4507±244 cells), 9 (4181±682 cells), and 13 (3846±1150 cells). The third row includes the J606 truncation series while J3 and J5 are highly expressed: strains 1, 2 (4392±277 cells), 3 (4290±233 cells), and 4 (4477±228 cells). All rows include the strain with an empty output plasmid containing no genes encoding fluorescent products (the same strain used for baseline subtraction). Figure 4b and Supplementary Figure 11 were plotted using the median value of each replicate's flow cytometry distribution, baseline-subtracted using a strain lacking the genes encoding the fluorescent proteins (grey, 3734±580 cells).



Supplementary Figure 19. Fluorescence assay for quantification of biopterins. **a** Standard curve of fluorescence excitation at 340nm and emission at 440nm versus concentration of BH2 spiked into cultures of the strain harboring an empty output plasmid. BH2 was chosen as the basis of this standard curve because of previous reports of >80% of biopterins production being in that oxidation state[6]. **b** Fluorescence values of combinatorial scRNA library variants expressing the BH4 pathway. Because most of the product is in the BH2 oxidation state, we use the BH2 calibration from panel **a** to evaluate total production in each strain. The fluorescence data plotted here were converted into the concentration values shown in Figure 5. These values have not been baseline-subtracted, but are normalized across multiple experiments on different days using a subset of common strains within the library, indicated by arrows below the *x*-axis heatmap. Note that expression-dependent variations in production are readily apparent even when the data are plotted as raw fluorescence values without baseline subtraction. Values represent the average ± standard deviation calculated from n = 3 biologically independent samples. Source data are provided as a Source Data file.

Supplementary Table 1. *E. coli* strains.

Strain	Description	Genotype
MG1655	parent *E. coli* strain	F- λ- ilvG- rfb-50 rph-1
JM109[8]	*lacZ*-deficient	recA1, endA1, gyrA96, thi, hsdR17, supE44, relA1, λ-, Δ(lac-proAB), [F' traD36, proAB, lacIqZΔM15]

Supplementary Table 2. Select scRNA MFE structures (Supplementary Figure 4).

scRNA	RNA sequence and predicted MFE structure[a]
J306	UUGUGUCCAGAACGCUCCGUGUUUUAGAGCUAGAAAUAGCAAGUUAAAAUAAGGCUAGUCCGUUAUCAACUUGAAAAAGUGGCACAUGAGGAUCACCCAUGUGCUUUUUUU .....((.(((((((...)))))))..))((((....))))(((((...((((((.....)).)))))..))))).(((((.(((((((((.((....)))))))))).)))))
J506	AGCAGCAUGAGCAGCAUUGAGUUUUAGAGCUAGAAAUAGCAAGUUAAAAUAAGGCUAGUCCGUUAUCAACUUGAAAAAGUGGCACAUGAGGAUCACCCAUGUGCUUUUUUU ......(((.(((((.....(((((((.((((....))))...)))))))...)))..)).))).....((((....))))(((((((.((....)))))))))..... GUCGCAGUCGCGCGAGCACUGUUUUAGAGCUAGAAAUAGCAAGUUAAAAUAAGGCUAGUCCGUUAUCAACUUGAAAAAGUGGCACAUGAGGAUCACCCAUGUGCUUUUUUU
J606	..(((......)))(((..(((((((((.((((....))))...)))))))))..)))............((((....))))(((((((.((....)))))))))..... ACAAUUGUCGCAACAUGAUCGUUGGUGAGUUAGAAAUAACAAGUACCAAUAAGGCUAGUCCGUUAUCAACUUGAAAAAGUGGCACAUGAGGAUCACCCAUGUGCUUUUUUU
AA	.....(((((((.((((....))))...)))))))..................((((....))))(((((((.((....)))))))))..... UGUCUCUUGAUGCAAGAGCGGUUUUAGAGCUAGAAAUAGCAAGUUAAAAUAAGGCUAGUCCGUUAUCAACUUGAAAAAGUGGCACAUGAGGAUCACCCAUGUGCUUUUUUU
g3	((.((((((...))))))).........((((....))))(((((...((((((.....)).)))))..))))).(((((.(((((((((.((....))))))))).)))))
J106	AGGACGCCUUUGGUAACCGCGUUUUAGAGCUAGAAAUAGCAAGUUAAAAUAAGGCUAGUCCGUUAUCAACUUGAAAAAGUGGCACAUGAGGAUCACCCAUGUGCUUUUUUU .(((((((((...((((.((..(((.........)))..))..))))....))))))..))))........((((....))))(((((((.((....)))))))))..... AGAGUAGACAGAUUAGCAGUGUUUUAGAGCUAGAAAUAGCAAGUUAAAAUAAGGCUAGUCCGUUAUCAACUUGAAAAAGUGGCACAUGAGGAUCACCCAUGUGCUUUUUUU
K	......(((.(((((((..(((((((((.((((....))))...)))))))))..)))))))..))).....((((....))))(((((((.((....)))))))))..... GUGAUAAACGGACUAGCCUUGUUUUAGAGCUAGAAAUAGCAAGUUAAAAUAAGGCUAGUCCGUUAUCAACUUGAAAAAGUGGCACAUGAGGAUCACCCAUGUGCUUUUUUU
L	.((((((.((((((((((((((((((((.((((....))))...))))))))))))))))))))))))((((....))))(((((((.((....)))))))))..... GAUGAUAGCGUGACUAGCUGGUUUUAGAGCUAGAAAUAGCAAGUUAAAAUAAGGCUAGUCCGUUAUCAACUUGAAAAAGUGGCACAUGAGGAUCACCCAUGUGCUUUUUUU
W	..(((((((((.(((((((((.(((((((.((((....))))...)))))))..)))))))))))))))))((((....))))(((((((.((....)))))))))..... GUUUUAGAGCUAGAAAUAGCAAGUUAAAAU

[a] MFE structures predicted by the RNAfold WebServer and output in dot-bracket notation.

Supplementary Table 3. scRNAs with orthogonal dCas9-binding handle sequences.

scRNA	RNA sequence of handle	Figure
Wild-type	GUUUUAGAGCUAGAAAUAGCAAGUUAAAAU	All except Supplementary Fig 5
AA – EE	GUUGGUGAGUUAGAAAUAACAAGUACCAAU	Supplementary Fig 5

Supplementary Table 4. Select kinetic and thermodynamic parameters of scRNA truncations.

scRNA	Folding barrier	Net binding energy
J306-20	10	-32.3
J306-19	10	-31.4
J306-18	10	-29.8
J306-17	10	-27.1
J306-14	4.9	-21.7
J306-11	4.9	-14.2
J506-20	6.9	-33.3
J506-19	6.9	-31.7
J506-18	6.9	-28.3
J506-17	6.9	-25.7
J506-14	6.6	-18.4
J506-11	5.4	-14
J606-20	9.3	-33.6
J606-19	9.3	-30.9
J606-18	8.9	-29.4
J606-17	8.5	-27.4
J606-14	8.5	-19.8
J606-11	7.1	-14.2

Supplementary Table 5. Sequences of key CRISPRa elements: promoters and scRNA array.

Element	Figures	Components	Sequence
J3 promoter[1]	2-6, S1, S3-S5, S7-S12, S17, S18	Upstream region, PAM site, J306 target site, UP-element (J3), BBa_J23117 promoter	AGCATTTGCGATCATTCACGCAGCGCTTATTCAGTTGCTCACTGCG ATGTCATAATCATCGCTACGAGCTGTGAAAGATGCATAAAGCTCGT ACGACGCGTTCGCTCGTCTCCTCACTTCTCCTACGGAGCGTTCTGG ACACAACGTCGTCTTGAAGTTGCGATTATAGATTGACAGCTAGCTC AGTCCTAGGGATTGTGCTAGC
J5 promoter	3-6, S7, S8, S10-S12, S17, S18	Upstream region, PAM site, J506 target site, UP-element (#25)[1], BBa_J23117 promoter	TATACATCGCATCACTACACTATTGATTATCATTGTGTACGTAACG AGCTTGCACAACGTGAAGTTCTTCGAGCACTTCAGCTCGCAACGTA AATGACAGTTGCTGTTAAGTGACGTGAATCCTTCAATGCTGCTCAT GCTGCTGTCGTAAATAAGTAAGTCACTCCCACTTGACAGCTAGCTC AGTCCTAGGGATTGTGCTAGC
J6 promoter	3-6, S7, S8, S10-S12, S17, S18	Upstream region, PAM site, J606 target site, UP-element (#24)[1], BBa_J23117 promoter	CTGCACGAGTTCGCTGTCGAGACAAGTCTCTTAGCGACGTATTACG AAGATCACATAGTCAGATGAAGCTATAGAGCACGACGCTAACGATT ACGTCACGCTTGACACAACAGTTTCGCTACCTAGTGCTCGCGCGAC TGCGACGTTGTCCTTCTAGTCGCCCATGACTCTTGACAGCTAGCTC AGTCCTAGGGATTGTGCTAGC
scRNA array example: strain #1	4-6, S8, S10-12, S17, S18	BBa_J23105 (SpeI) promoter, J306 spacer sequence, dCas9 handle, MS2 hairpin, rrnB terminator, BBa_J23105 (SpeI) promoter, J506 spacer sequence, dCas9 handle, MS2 hairpin, ECK120033736 terminator, BBa_J23105 (SpeI) promoter, J606 spacer sequence, dCas9 handle, MS2 hairpin, ECK120033736 terminator	TAAGTTGTTACTAGATTTACGGCTAGCTCAGTCCTAGGTACTATAC TAGTTTGTGTCCAGAACGCTCCGTGTTTTAGAGCTAGAAATAGCAA GTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACATG AGGATCACCCATGTGCTTTTTTTGAAGCTTGGGCCCGAACAAAAAC TCATCTCAGAAGAGGATCTGAATAGCGCCGTCGACCATCATCATCA TCATCATTGAGTTTAAACGGTCTCCAGCTTGGCTGTTTTGGCGGAT GAGAGAAGATTTTCAGCCTGATACAGATTAAATCAGAACGCAGAAG CGGTCTGATAAAACAGAATTTGCCTGGCGGCAGTAGCGCGGTGGTC CCACCTGACCCCATGCCGAACTCAGAAGTGAAACGCCGTAGCGCCG ATGGTAGTGTGGGGTCTCCCCATGCGAGAGTAGGGAACTGCCAGGC ATCAAATAAAACGAAAGGCTCAGTCGAAAGACTGGGCCTTTCGTTT TATCTGTTGTTTGTCGGTGAACTGGATCCTTACAGATCATTTACGG CTAGCTCAGTCCTAGGTACTATACTAGTAGCAGCATGAGCAGCATT GAGTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTT ATCAACTTGAAAAAGTGGCACATGAGGATCACCCATGTGCTTTTTT TAACGCATGAGAAAGCCCCCGGAAGATCACCTTCCGGGGGCTTTTT TATTGCGCAGATCATTTACGGCTAGCTCAGTCCTAGGTACTATACT AGTGTCGCAGTCGCGCGAGCACTGTTTTAGAGCTAGAAATAGCAAG TTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACATGA GGATCACCCATGTGCTTTTTTTAACGCATGAGAAAGCCCCCGGAAG ATCACCTTCCGGGGGCTTTTTTATTGCGCAGATCTCCTTTGAGT

Supplementary Table 6. Terminator sequences.

Terminator	Sequence
TrrnB	GAAGCTTGGGCCCGAACAAAAACTCatctcagaagaggatctgaatagcgccgtcgaccatcatcatcatcatca ttgagtttaaacggtctccagcttggctgttttggcggatgagagaagattttcagcctgatacagattaaatca gaacgcagaagcggtctgataaaacagaatttgcctggcggcagtagcgcggtggtcccacctgaccccatgccg aactcagaagtgaaacgccgtagcgccgatggtagtgtggggtctccccatgcgagagtagggaactgccaggca tcaaataaaacgaaaggctcagtcgaaagactGGGCCTTTCGTTTTATCTGTTGTTTGTCGGTGAACT
BBa_B0015	ccaggcatcaaataaaacgaaaggctcagtcgaaagactgggcctttcgttttatctgttgtttgtcggtgaacg ctctctactagagtcacactggctcaccttcgggtgggcctttctgcgtttata
BBa_B1002	cgcaaaaaaccccgcttcggcggggttttttcgc
ECK120033736[9]	aacgcatgagAAAGCCCCCGGAAGATCACCTTCCGGGGGCTTTtttattgcgc
ECK120033737[9]	ggaaacacagAAAAAAGCCCGCACCTGACAGTGCGGGCTTTTTTTTTcgaccaaagg
ECK120010818[9]	GTCAGTTTCACCTGTTTTACGTAAAAACCCGCTTCGGCGGGTTTTTACTTTTGG
ECK120015440[9]	tccggcaattAAAAAAGCGGCTAACCACGCCGCTTTTTTtacgtctgca

Supplementary Method 1. Energetic parameters describing scRNA folding.

Folding Barrier (FB) can be further divided into two separate barrier heights: the Handle Barrier and the Spacer Barrier (Supplementary Figure 2b). The Handle Barrier is defined as the barrier height for the conversion from the minimum free energy (MFE) structure to the most stable structure where the dCas9-binding handle is correctly folded. The Spacer Barrier represents the barrier height for the conversion from the most stable structure where the handle is correctly folded to the active structure. Within our dataset the Spacer Barrier explains substantially more of the variation in observed CRISPRa levels, suggesting that the ability of the dCas9-scRNA complex to bind to the DNA is a more important determinant of CRISPRa in our system than the dCas9 binding to an scRNA.

Overall, the kinetic parameter Folding Barrier has the most explanatory power of CRISPRa activity. To test the performance of Folding Barrier for predicting CRISPR-activated expression, we expanded our set of unique synthetic promoters for CRISPRa. We constructed 24 new CRISPRa promoters with target sites with Folding Barriers between about 5 kcal/mol and 35 kcal/mol. We chose these values in order to interrogate whether parameter values at the extremes of that range would enable us to identify highly active (or inactive) scRNAs that would be less likely to be identified by chance. Similar to the previous set where the scRNA target sites were randomly selected, we observed dramatic differences in CRISPRa activity with each promoter, varying by more than 40-fold (Figure 2d). The Folding Barrier parameter explained the majority of the variation in CRISPR-activated expression we observed, with a Spearman rank correlation coefficient (r_S) of 0.8 (Supplementary Figure 3). Interestingly, there appears to be a threshold—roughly 10 kcal/mol—below which reducing the Folding Barrier does not further increase CRISPRa levels.

To investigate if CRISPRa activity could be improved by further minimizing the Folding Barrier, we designed five additional scRNAs with Folding Barrier values lower than 3.5 kcal/mol. In order to design these highly-unstructured scRNAs, we had to mutate the dCas9-binding handle sequence to increase its folding stability: our computational predictions suggested that the stability of the wild-type handle was insufficient. For three of the five scRNAs, the MFE structure was the same as the active structure, implying that no further structural rearrangement was needed to bind to dCas9. CRISPR-activated gene expression using these scRNAs was generally high enough to indicate adequate activation, but was similar to the J306 (FB = 10 kcal/mol) level or slightly less (Supplementary Figure 6), suggesting a possible CRISPRa expression plateau in our system. This result further supports the idea of Folding Barrier's predictive power being most useful as a threshold rather than as a quantitative prediction. Below this threshold, other parameters are likely to have greater quantitative predictive power. These AA-EE scRNAs are not pictured in Figure 2d, but they further support the sigmoidal fit proposed there.

In addition to the kinetic parameters, we computed a set of metrics that analyze scRNA folding by relying solely on thermodynamic parameters. Net Binding Energy (NBE) is obtained by first calculating the ΔΔG between the MFE structure of a scRNA and its active conformation (Folding Energy), then adding the ΔG of the spacer binding the DNA target (Binding Energy), modeled here as an RNA-RNA duplex according to the Vienna tool RNAduplex (Supplementary

Figure 2a). Net Binding Energy, then, is a composite metric combining the energy penalty associated with gRNA refolding with the energy of the spacer binding to its target.

The abilities of Folding Energy and Net Binding Energy to predict CRISPRa function were quite similar in our set of scRNAs (Supplementary Figure 3), suggesting that including Binding Energy adds little or no predictive value. Indeed, Binding Energy's correlation with CRISPRa activity is quite low. Still, Folding Energy has a high enough correlation to provide useful predictions, especially as a supplement to Folding Barrier below its 10 kcal/mol threshold.

To aid in identifying optimal scRNAs while avoiding defective ones, we additionally considered a combined screen using a kinetic parameter, Folding Barrier, and an energetic one, Net Binding Energy. The strategy was to first screen candidate scRNAs for those possessing sub-threshold FB values ≤10 kcal/mol, effectively screening against defective scRNAs that are kinetically trapped and unable to reach their active state. Remaining candidates would then be ranked based on their NBE value, enriching for optimal guides that possess ideal thermodynamic and kinetic properties (as in Supplementary Figure 5). We ultimately adopted the simpler use of only the FB screening for this work, which still represents an improvement over all of the common gRNA screening metrics we analyzed (Supplementary Figure 3), but the combined screening metric FB-NBE remains an effective strategy to simultaneously avoid defective guides and enrich for optimal guides.

Surprisingly, the handle fraction, which is the fraction of the population of scRNAs expected to have the Cas9-binding handle correctly folded, demonstrates no correlation with CRISPRa activity in our system (Supplementary Figure 3).

Supplementary Method 2. Orthogonal CRISPRa promoter design.

The synthetic CRISPRa promoters described in this study (J3, J5, and J6) are 170 bp in length and contain a PAM site for scRNA targeting at -81 bp to the TSS on the non-template strand, where maximum CRISPRa activity was observed in our previous study[10]. Each synthetic CRISPRa promoter contains a unique 20 bp scRNA target sequence and a unique 26 bp sequence, called UP-element, between the target site and the minimal promoter (Supplementary Figure 16). These sequences were previously characterized to give high CRISPR-activated expression (scRNA target/spacer sequences B and E and UP-elements #24 and #25[1]). Sequences upstream of the -81 bp target site were randomized and verified to lack additional PAM sites. The BBa_J23117 minimal promoter was used for all three synthetic CRISPRa promoters, as it exhibited the highest dynamic range from the tested set of minimal promoters[1]. The experiments in Figure 2 and Supplementary Figures 1, 3, 5, and 6 were performed using a series of J3 promoter variants where the J306 target site is replaced with unique target sites (g1 through EE, Supplementary Data 1).

Supplementary references

1. Fontana, J. *et al.* Effective CRISPRa-mediated control of gene expression in bacteria must overcome strict target site requirements. *Nature Communications* **11**, 1–11 (2020).
2. Doench, J. G. *et al.* Optimized sgRNA design to maximize activity and minimize off-target effects of CRISPR-Cas9. *Nat Biotechnol* **34**, 184–191 (2016).
3. Moreno-Mateos, M. A. *et al.* CRISPRscan: designing highly efficient sgRNAs for CRISPR-Cas9 targeting in vivo. *Nat Methods* **12**, 982–988 (2015).
4. Concordet, J.-P. & Haeussler, M. CRISPOR: intuitive guide selection for CRISPR/Cas9 genome editing experiments and screens. *Nucleic Acids Research* **46**, W242–W245 (2018).
5. Gruber, A. R., Lorenz, R., Bernhart, S. H., Neubock, R. & Hofacker, I. L. The Vienna RNA websuite. *Nucleic Acids Research* **36**, W70–W74 (2008).
6. Kiattisewee, C. *et al.* Portable bacterial CRISPR transcriptional activation enables metabolic engineering in *Pseudomonas putida*. *Metabolic Engineering* **66**, 283–295 (2021).
7. Alba Burbano, D. *et al.* Engineering activatable promoters for scalable and multi-input CRISPRa/i circuits. *Proc. Natl. Acad. Sci. U.S.A.* **120**, e2220358120 (2023).
8. Yanisch-Perron, C., Vieira, J. & Messing, J. Improved M13 phage cloning vectors and host strains: nucleotide sequences of the M13mpl8 and pUC19 vectors. *Gene* **33**, 103–119 (1985).
9. Chen, Y.-J. *et al.* Characterization of 582 natural and synthetic terminators and quantification of their design constraints. *Nature methods* **10**, 659 (2013).
10. Dong, C., Fontana, J., Patel, A., Carothers, J. M. & Zalatan, J. G. Synthetic CRISPR-Cas gene activators for transcriptional reprogramming in bacteria. *Nature Communications* **9**, (2018).