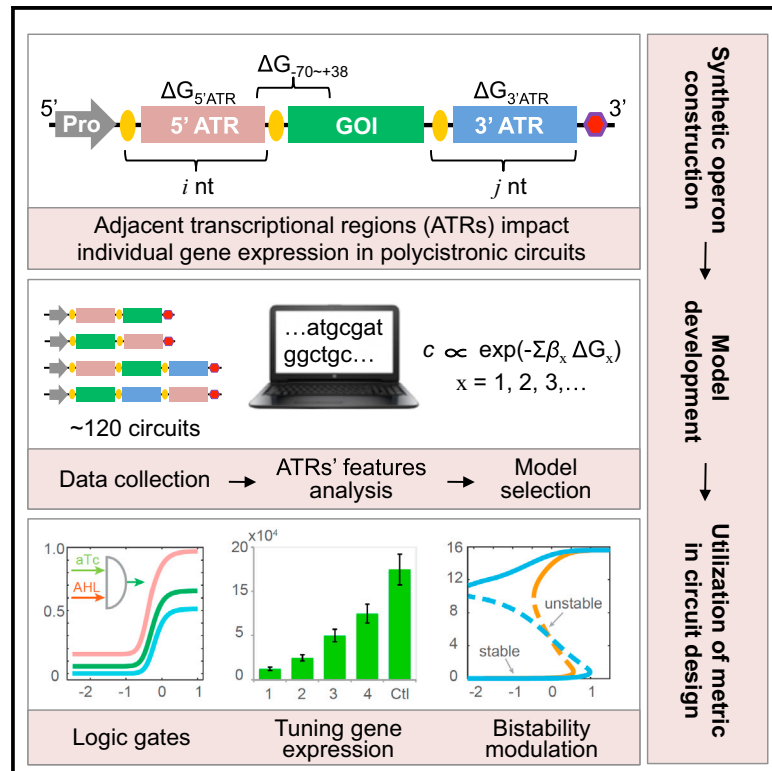**Article**

Design of Adjacent Transcriptional Regions to Tune Gene Expression and Facilitate Circuit Construction

Graphical Abstract

Authors

Fuqing Wu, Qi Zhang, Xiao Wang

Correspondence

xiaowang@asu.edu

In Brief

Wu et al. quantify the effect of adjacent transcriptional regions (ATR) on protein expression in ~120 synthetic polycistronic gene circuits. Data-driven analysis yields a new protein expression metric that strongly correlates with features of ATRs, including GC content, size, and stability of mRNA folding near ribosomal binding sites. This metric's utility is demonstrated in predicting each gene's relative expression levels in synthetic logic gates and circuit outputs and tuning gene expression and nonlinear dynamics of bistable gene networks.

Highlights

- Adjacent transcriptional regions (ATR) affect gene expression in polycistronic circuits

- ATRs with high GC content, small size, and low folding energy promote gene expression

- A thermodynamic model can predict protein expression levels based on ATR properties

- Model's utility is demonstrated in tuning gene expression and synthetic circuit design

CellPress

Design of Adjacent Transcriptional Regions to Tune Gene Expression and Facilitate Circuit Construction

Fuqing Wu,[1,2] Qi Zhang,[1,2] and Xiao Wang[1,3,*]
[1]School of Biological and Health Systems Engineering, Arizona State University, Tempe, AZ 85287, USA
[2]These authors contributed equally
[3]Lead Contact
*Correspondence: xiaowang@asu.edu
https://doi.org/10.1016/j.cels.2018.01.010

SUMMARY

Polycistronic architecture is common for synthetic gene circuits, however, it remains unknown how expression of one gene is affected by the presence of other genes/noncoding regions in the operon, termed adjacent transcriptional regions (ATR). Here, we constructed synthetic operons with a reporter gene flanked by different ATRs, and we found that ATRs with high GC content, small size, and low folding energy lead to high gene expression. Based on these results, we built a model of gene expression and generated a metric that takes into account ATRs. We used the metric to design and construct logic gates with low basal expression and high sensitivity and nonlinearity. Furthermore, we rationally designed synthetic 5′ATRs with different GC content and sizes to tune protein expression levels over a 300-fold range and used these to build synthetic toggle switches with varying basal expression and degrees of bistability. Our comprehensive model and gene expression metric could facilitate the future engineering of more complex synthetic gene circuits.

INTRODUCTION

Gene circuit engineering as one of the foundation technologies has helped start the burgeoning development of bacterial synthetic biology. Based on a large collection of well-characterized biological components, including promoters, ribosome binding sites, transcriptional factors, terminators, RNA elements, and other small modules, complex gene circuits with designed functions can be wired using established biological principles. Toggle switch and repressilator are two of the earliest examples of engineered gene circuits (Elowitz and Leibler, 2000; Gardner et al., 2000). Now synthetic biologists are paying increasing attention to develop innovative gene circuits for spatial pattern formation (Cao et al., 2016; Liu et al., 2011), drug development (Smanski et al., 2016; Wright, 2014), pathogen detection (Pardee et al., 2014, 2016), in vivo delivery (Din et al., 2016), and other biotechnological applications, including nitrogen fixation (Mus et al., 2016; Wu and Wang, 2015) and environmental bioremediation (Zhang and Nielsen, 2014).

Currently, circuit assembly has two main strategies: one is monocistronic construct, in which one promoter drives one gene expression and ensures each gene is being expressed independently; the other is polycistronic construct, in which one promoter transcribes multiple genes (operon) into a single mRNA but is translated into individual products (Figure 1A). Operon, a cluster of genes with functional associations under control of a single promoter, is a common type of genome organization in prokaryotic cells and is also widely found in eukaryotes and viruses (Rocha, 2008). This operon organization strategy, here mainly referring to the genes' order and position downstream of the promoter in an operon, ensures coordinated gene expression and regulation and enables bacteria cells to rapidly respond to environmental changes. In synthetic biology, this organization (synthetic operon) facilitates rapid construction of genetic cascades and decreases the number of biological components (such as the promoters and terminators) required for complex genetic circuits, and therefore is widely used in circuit engineering (Ma et al., 2016; Lee et al., 2016; Farasat et al., 2014; Cameron and Collins, 2014; Prindle et al., 2014; Yang et al., 2014; Litcofsky et al., 2012; Wu et al., 2017).

However, it remains unknown whether/how gene expression is affected by immediately adjacent genes in a polycistronic operon. Two previous reports have indicated that gene position and transcriptional distance can affect gene expression in a synthetic operon (Chizzolini et al., 2014; Lim et al., 2011). But little research has systematically studied the effects of adjacent genes in synthetic operons on the circuit's gene expression, dynamics, and functionality. This factor is more prominent for synthetic operons containing a cluster of genes and complex multi-layered genetic circuits. Deciphering the effects of adjacent transcriptional region (ATRs) on gene expression would advance our understanding of determinants of gene expression in synthetic circuits and accelerate circuit design and assembly. Such effect has been generally neglected during engineering of synthetic gene networks, leading to unexpected circuit performance or failure (Brophy and Voigt, 2016; Carr et al., 2017; Yeung et al., 2017). Hence, development of a predictive method to evaluate each gene's expression level in a circuit would be of great importance to circumvent the need for trial and error in circuit design and assembly.
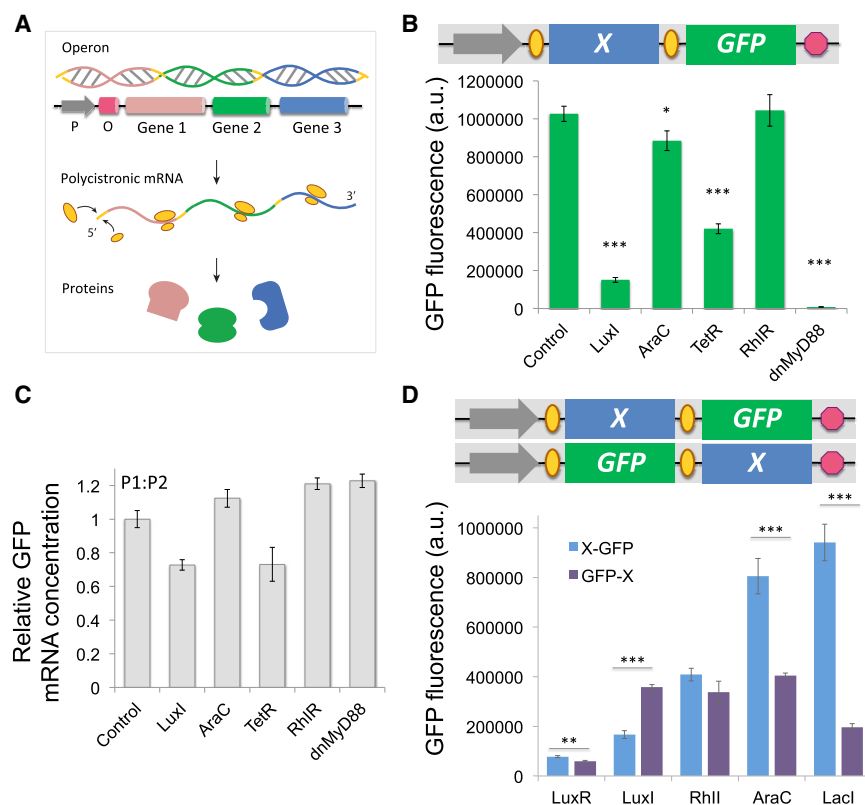
Figure 1. Protein Expression Is Significantly Influenced by Its Adjacent Genes and Position in Synthetic Operons

(A) Illustration of the operon structure and gene expression. The three structural genes are transcribed as a polycistronic mRNA but translated into individual proteins. P, promoter; O, operator. Yellow oval, ribosome.

(B) Top: Schematic representation of synthetic bicistronic gene circuits with gene X and GFP. Gray arrow, constitutive promoter; orange oval, ribosome binding site; red hexagon, transcriptional terminator. Bottom: Flow cytometry results show GFP expression is influenced by its 5'ATRs. X represents a gene name (i.e., $LuxI$, $AraC$, $TetR$, $RhlR$, and $dnMyD88$). "Control" is without X gene in the circuit. Rectangles with filled colors represent different genes. Data represent the mean ± SE of eight replicates.

(C) Relative GFP mRNA concentrations (normalized to 16S rRNA control) for the circuits in (B) determined by RT-qPCR. Primer pair P1:P2 was designed to amplify GFP gene from the sample cDNA.

(D) Top: Schematic representation of synthetic bicistronic gene circuits with gene X and GFP, but with switched positions in the circuit. Gene position in the operon affects GFP expression.

Data represent the mean ± SE of eight replicates. *$p < 0.05$, **$p < 0.001$, and ***$p < 0.0001$ by Student's t test.

To quantify the effects of ATRs on gene expression, here we systematically analyzed the effect of adjacent genes and noncoding regions on GFP expression levels through construction of ~120 synthetic gene circuits (operons) in *Escherichia coli*. Data-driven analysis yields a new protein expression metric that strongly correlates with the features of ATRs including GC content, size, and stability of mRNA folding near ribosomal binding sites (RBS). We demonstrated this metric's utility in evaluating relative expression levels of genes by incorporating it in the design and construction of logic gates with lower basal expression and higher sensitivity and nonlinearity. Furthermore, we designed synthetic 5'ATRs to tune protein expression levels over a 300-fold range. Finally, by combining ATR regulation and mathematical modeling, we illustrated the application of synthetic ATRs in quantitatively tuning nonlinear dynamics of bistable gene networks.

RESULTS

Protein Expression Is Influenced by Adjacent Genes and Position

To examine whether protein expression is affected by its neighbors in a polycistronic setting, we first constructed a two-gene operon (gene X and GFP), which is driven by a constitutive promoter (Figure 1B). Flow cytometry results showed that for different X, GFP expression varies significantly. Specifically, circuits with $AraC$ and $RhlR$ as X showed a comparable level of GFP fluorescence with the control (without X gene), while the others ($LuxI$, $TetR$, and $dnMyD88$) showed high expression variations, ranging

from 6-fold to over 120-fold decrease compared with control (Figure 1B). Membrane protein dnMyD88 shows the most significant influence on its neighbor GFP expression. On the other hand, RT-qPCR measurements of transcripts of GFP showed much smaller variations of mRNA concentrations between different circuits, for P1:P2 (GFP N-terminal) or P3:P4 (GFP C-terminal) primer pairs (Figures 1C and S1A–S1C). So the variation of mRNA concentrations for each construct is insufficient to explain the fluorescence differences, which is in agreement with previous studies that protein and mRNA copy numbers in *E. coli* cells are uncorrelated (Lim et al., 2011; Taniguchi et al., 2010).

Next, we further investigated the influence of a gene's position on its expression. As shown in Figure 1D, higher GFP expression is observed when GFP is arranged distal to the promoter for the bicistronic constructs that X gene is $RhlI$, $AraC$, or $LacI$, while there are cases showing a similar level of GFP fluorescence ($LuxR$) or higher ($LuxI$) when GFP is arranged right downstream of the promoter. Results from tricistronic constructs also indicate that GFP expression is varied for different positions in the circuit and adjacent genes (Figures S1D–S1G). Moreover, for different Xs with the same position, GFP shows substantial variations, consistent with results shown in Figure 1B. Altogether, these results demonstrate that a gene's sequence and position in operons have an effect on the expression of adjacent genes.

Quantitative Characterizations of ATR Effects on Synthetic Operons

To quantify the impact of ATRs on protein expression, we designed and constructed ~80 circuits with different neighbor

protein-coding genes and varying sizes (*X* and *Y*) to cover a wide range of GFP gene position and neighbor features (GC content, size, and mRNA secondary structure). These genes are commonly used in synthetic biology, including transcriptional factors, quorum-sensing components, and other functional genes (Table S1). To ensure experimental consistency, all circuits were constructed using the same constitutive promoter, RBS, terminator, and expression vector.

First, GFP was arranged to the distal end of synthetic bicistronic and tricistronic operons, and the DNA sequence starting from the transcription start site after the promoter to the beginning of the RBS of GFP is denoted as 5′ATRs (Figure 2A). Log transformation was applied to the original data because of its large variability ranging from 21,000 to 1,900,000 (GFP fluorescence, arbitrary unit) and inconstant variance (STAR Methods). GFP expression increased with the total 5′ATRs GC content, while 5′ATR length had a negative effect on GFP expression. Sliding window analysis of 5′ATR GC content suggested that the GC content of the whole 5′ATR has the highest fitting efficiency (Figure S2A). We hypothesize that high GC content could increase total mRNA stability, while a long transcription process could decrease the probability of complete GFP transcription/translation and increase the probability of degradation. In addition, previous studies reported that RNA secondary structure near the RBS influences a gene's expression, so local folding energy from the −70-nt to +38-nt region around GFP's RBS (GFP's translation starting site is denoted as +1) was calculated. Consistent with previous reports (Kudla et al., 2009; Mao et al., 2014; Tuller et al., 2010), our analysis also shows that GFP expression is significantly correlated with folding energy around the RBS of GFP (Figure 2A and Table S2).

Next, GFP was placed in the middle of the operon, and the sequence between the stop codon of GFP and the transcriptional terminator is denoted as 3′ATR. We found that 5′ ATR GC content (positive impact) and local mRNA folding free energy (negative impact) have the most significant impacts on GFP expression, and 3′ATR GC content has a small contribution to GFP variations in this case (Figure 2B and Table S2). Finally, circuits with GFP engineered proximally to the promoter were also constructed and investigated to probe the relationship between GFP expression and its 3′ATR. Similarly, results show that 3′ATR GC content and size have a positive and negative relationship with GFP fluorescence, respectively (Figure 2C and Table S2). Sliding window analysis further revealed that the GC content of the first 100 nt of 3′ATR has the highest fitting efficiency, suggesting the rear 100 nt is important for GFP expression (Figure S2B).

Noncoding DNA sequences make up about 12% of the bacterial genome and play important roles in the regulation of gene expression and metabolism (Ahnert et al., 2008; Oliva et al., 2015). To investigate whether noncoding sequences would similarly affect adjacent gene expression in synthetic operons, we engineered 32 synthetic circuits with 32 genes (Table S1), which are placed immediately downstream of the promoter without RBS to greatly limit their translation (Figure S2C). Our results showed a strong relationship between GFP expression and noncoding 5′ATR GC content, size, and local mRNA folding energy (Figure 2D and Table S2). Higher GFP expression was observed for circuits with the same genes with RBS than those without

RBS (Figure S2D), suggesting the RBS of 5′ATR may be important for mRNA stabilization and expression efficacy.

Altogether, these results offer direct evidence that adjacent coding and noncoding DNA fragments affect gene expression in synthetic operons, and ATR GC content has a positive correlation with GFP expression while ATR size and local free energy are both negatively correlated.

Comprehensive Model of ATR Regulation
Our results revealed that gene expression in operons is affected by the sequence features of its adjacent genes and local mRNA secondary structures. The explicit mechanism of these effects remains elusive. We employed the same promoter, RBS, vector, and host cell for all the circuits to minimize the impact of transcription on protein expression variation. And there is a lack of complicated post-translational modifications in *E. coli*, so we believe that the ATR alters the secondary or tertiary structures of mRNA locally and/or globally, which perturbs the GFP mRNA translation and degradation process (Figure 2E). The GC content of 5′ and 3′ ATRs has a positive relationship with GFP expression (Figure 2). After the RBS is transcribed, ribosome and RNase competitively bind to mRNA (Chen et al., 2015a; Mackie, 2013). So we infer that a GC-rich 5′ and 3′ATR, which is likely to have a more stable secondary structure (Emory et al., 1992; Selinger et al., 2003), could help stabilize the GFP transcript and decrease the risk of degradation by RNase, and thus result in higher GFP expression. On the other hand, the 5′ and 3′ ATR sizes are negatively correlated with GFP expression (Figure 2). Longer ATR may lead to lower mRNA stability due to the increased probability of elongation pausing and degradation of RNase. Moreover, the local mRNA folding energy near GFP's RBS (nt −70 to +38) is believed to have an impact on the translation initiation of GFP (Kudla et al., 2009; Mao et al., 2014; Tuller et al., 2010). Overall, our statistics analysis revealed that 5′ATR GC content is the most important variable in the regression models for the *X-GFP* circuit (Figure 2A, partial R^2 = 0.44, Table S3) and *X-GFP-Y* (Figure 2B, partial R^2 = 0.51), whereas 3′ATR size has a bigger role in the model of *GFP-X* (Figure 2C, partial R^2 = 0.58, and Table S3). This result suggests that gene expression may be more easily modulated by the GC content of its 5′ATR and the size of 3′ATR.

To explore the possible mechanistic basis of ATR regulation and make quantitative predictions, we developed a comprehensive linear model integrating the three scenarios in Figures 2A–2C. The biophysical model was based on previous pioneer work characterizing the relationship between free energy changes and protein translation initiation (Salis et al., 2009; Serra and Turner, 1995; de Smit and van Duin, 1990; Xia et al., 1998). We next developed a comprehensive model to explore the possible mechanistic basis of ATR regulation. The model builds on measurements of sequence-dependent energetic changes during polycistronic mRNA folding and translation. The energetic changes correspond to the translation efficiency and protein abundance (*c*).

$$c \propto \exp\left(-\sum \beta_x \, \Delta G_{x_i}\right), \quad x = 1, 2, 3, \ldots$$

where ΔG is the energy term and β is the scaling coefficient (Salis et al., 2009). For a given gene in an operon, the size of 5′ and 3′
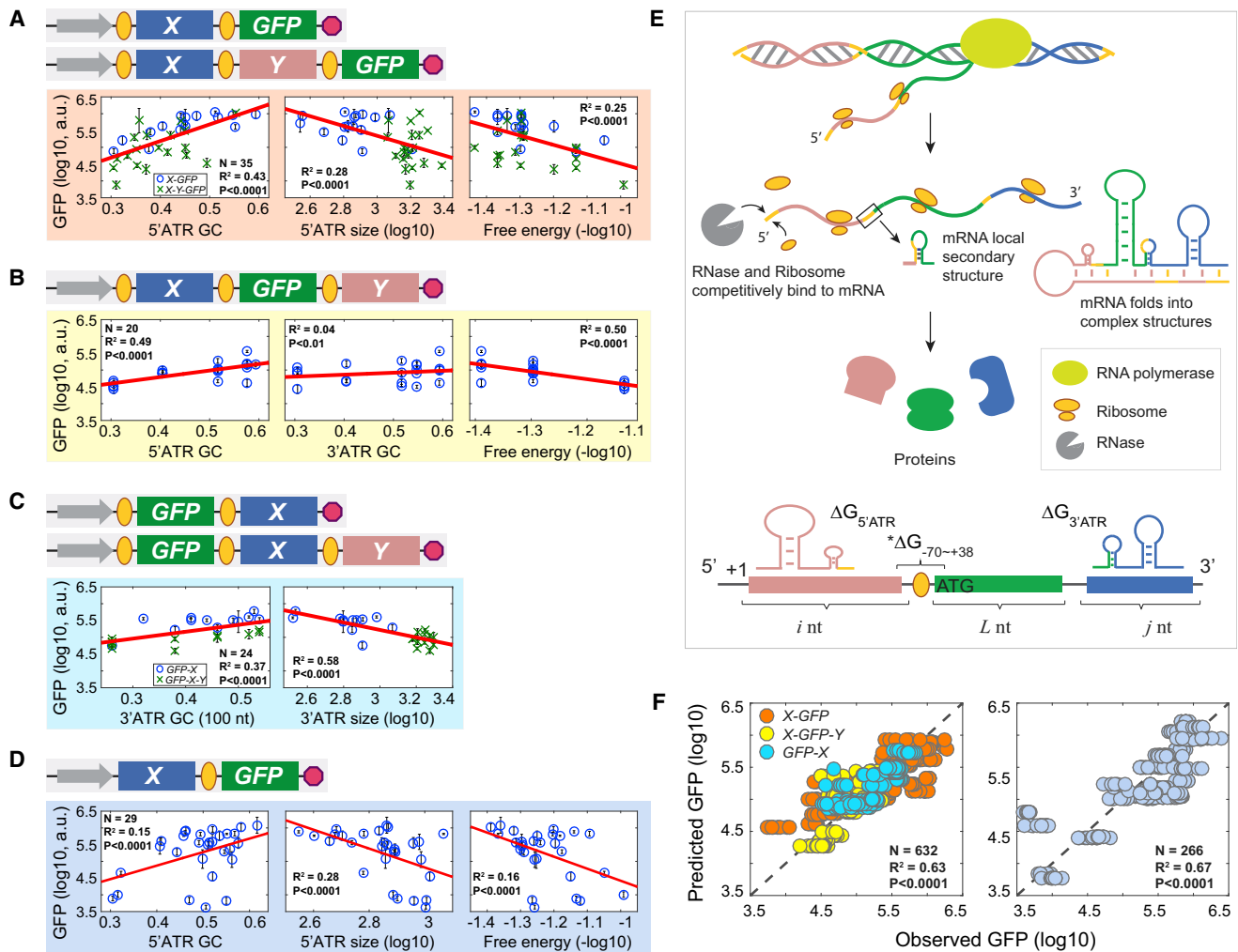
Figure 2. Quantitative Characterization of Adjacent Gene Regulation in Synthetic Operons

(A) Scenario 1: GFP is arranged distal to the promoter. Top: Schematic representation of synthetic polycistronic gene circuits *X-GFP*. *X* and *Y* represent different gene names. Bottom: GFP expression is significantly affected by its 5′ATRs' GC content, size, and local folding free energy. 35 genetic circuits with one or two genes placed in front of GFP, which are labeled with different symbols in the regression results. The red lines are the linear regression results from the data. Error bars are the SD of eight measurements performed in three different days. Coefficients with SE and equations can be found in Table S2.

(B) Scenario 2: GFP is placed in the middle of the three-gene operons (*X-GFP-Y*). GFP expression is significantly correlated with its 5′ and 3′ ATR GC content and local folding free energy. 20 circuits with different *X* and *Y* gene combinations were constructed.

(C) Scenario 3: GFP is placed proximal to promoter (*GFP-X*). GFP expression is significantly affected by its 3′ATR GC content and size. 24 circuits with different 3′ATRs were constructed, and different symbols are used to indicate bi- or tricistronic constructs in the regression results.

(D) Investigation of noncoding ATR regulation on GFP expression. *X* gene would not be expressed owing to a lack of RBS. GFP expression is significantly correlated with 5′ATR GC content, size, and local folding free energy. 29 circuits with different *X* genes were constructed.

(E) A comprehensive model for ATR regulation on protein expression. Top: Co-transcriptional translation and degradation. After RBS is transcribed, RNase and ribosome competitively bind to mRNA to initiate translation or degradation. Generally, gene expression is influenced by overall stability and local secondary structure. Bottom: Illustration of the five variables in the model: $\Delta G_{5'ATR}$, $\Delta G_{3'ATR\_100}$, $\Delta G_{-70\sim+38}$, and transcriptional sizes (i, j). -70 and $+38$ correspond to the position of the start codon (AUG) of the gene of interest.

(F) Left: Experimentally observed GFP expressions are plotted against the GFP values predicted by the coding ATR model with the five statistically significant energetic terms and fitted coefficients. If the model predicted values and experimentally observed values agreed perfectly ($R^2 = 100\%$), all the data points would fall on the dotted diagonal line of the squares. N is the total measurements for the 79 circuits. Dots with different colors indicate the data source from the three scenarios in (A–C). Right: Experimentally measured GFP fluorescence is plotted against the GFP expression predicted by the noncoding ATR model with the three statistically significant energetic terms ($\Delta G_{5'ATR}$, i, and $\Delta G_{-70\sim+38}$).

ATRs is denoted as *i* nt and *j* nt, respectively (Figure 2E). The minimum free energy of the local GFP mRNA secondary structure around the RBS is $\Delta G_{-70\sim+38}$. The entire folding energy for 5′ATR is $\Delta G_{5'ATR}$. The GC content of the first 100-nt 3′ATR has

the highest fitting efficacy for GFP expression (Figures 2C and S2B), and it is known that GC content is correlated with the thermodynamic parameter ΔG (Seffens and Digby, 1999; Trotta, 2014), so we only calculated the free energy of the first 100 nt

of 3′ATR ($\Delta G_{3'ATR\_100}$). Thus, the sum of the energy changes can be quantified to assess the abundance of a given gene expression:

$$-\sum \beta_x \Delta G_x = \beta_0 + \beta_1 \cdot \Delta G_{5'ATR} + \beta_2 \cdot \Delta G_{3'ATR\_100} + \beta_3 \cdot i \cdot Gm$$
$$+ \beta_4 \cdot j \cdot Gm + \beta_3 \cdot \Delta G_{-70\sim+38}$$

The folding energy of $\Delta G_{5'ATR}$, $\Delta G_{3'ATR\_100}$, and $\Delta G_{-70\sim+38}$ is totally sequence dependent, and Gm is an average energy cost for synthesizing a nucleotide, which here for simplicity we assume is a constant. Although all the five variables are contained in the model, some variables may be unnecessary for a specific gene organization in a circuit. For example, in the noncoding ATR cases with $X\text{-}GFP$ organization (Figure 2D), the j and $\Delta G_{3'ATR\_100}$ terms are constant values, owing to a lack of 3′ATRs.

The comprehensive model combined the three different scenarios with GFP placed at different positions in a polycistronic gene circuit (Figures 2A–2C). To verify whether the five variables are necessary for the best prediction of the model, we performed stepwise regression to test the significance of each variable through adding or removing one of the variables step by step (the significance level for variable entry or stay is 0.05). From the sequence of generated models, the selected model is chosen based on the lowest Akaike information criterion. Our results indicated that all five variables are necessary for the coding ATR model integrating the three scenarios in Figures 2A–2C (Summary of Stepwise Selection in Table S3), and the comprehensive model explains 63% of GFP variations (Figure 2F, left). The noncoding ATR model with the three statistically significant variables $\Delta G_{5'ATR}$, i, and $\Delta G_{-70\sim+38}$ (Summary of Stepwise Selection in Table S3) explains 67% of GFP variations (Figure 2F, right). With the comprehensive model, we can evaluate the influence of the adjacent transcriptional sequences on the expression of a certain gene in the operon, which provides a guide for circuit design and optimization during circuit engineering.

Protein Expression Metric Guided Logic Circuit Design

To illustrate how the metric could be used to guide circuit design, synthetic AND logic gate was designed and tested. The gate is composed of a hybrid promoter *pLux/tet*, which has one LuxR-AHL and one TetR binding site. GFP is the output. Maximized GFP expression is achieved in presence of two inputs AHL and aTc (Figure 3A), where AHL binds with LuxR protein to activate *pLux/tet* transcription and aTc can block TetR repression to *pLux/tet*. LuxR and TetR are constitutively expressed from the same promoter.

There are two possible ways to assemble this circuit, one is LuxR-TetR (LT) combination, and the other is TetR-LuxR (TL). The GC content of *LuxR* (30.3% GC, 781 bp) is lower than *TetR* (40.4% GC, 685 bp). So in AND-gate LT, TetR expression is lowered by its 5′-low-GC-content neighbor while the impact of LuxR to TetR expression in logic TL is minor because the size of 3′ATR is a more significant factor than GC content. We then calculated the equation for each circuit design and fed it into our model; the results indicate that LuxR expression in TL decreases by 4.4% compared with gate LT, however, TetR expression increases by 93.6% in circuit TL (Table S2). Therefore, we infer that the basal GFP expression in circuit LT would

be greater than in TL, whereas TL would harbor more dynamic responses with induction of aTc because of higher TetR expression. An ordinary differential equation (ODE) model was then developed to simulate GFP expression based on the normalized LuxR and TetR production rate changes in the LT and TL gates (STAR Methods). By tuning the relative production rates of LuxR and TetR according to the comprehensive regression model, we can predict GFP dynamics under induction of AHL and aTc (Figures 3B and S3A, solid lines). It can be seen that, after normalization, experimental dose-response results, shown as colored circles, are consistent with ODE model predictions for all aTc concentrations. Basal expression of *pLux/tet* in circuit LT is significantly higher than in circuit TL (Figures 3B and S3A, data points with error bar). Moreover, the maximum GFP fluorescence is also higher in circuit LT, owing to decreased LuxR expression in gate TL. In addition, the sensitivity to AHL (concentration for half-maximal activation of GFP, $K_{0.5}$) is improved 2.4- to 64.5-fold in circuit TL compared with LT for different concentrations of aTc. And the nonlinearity (Hill coefficient) is generally increased 2- to 5-fold with high concentrations of aTc induction. These data are in accordance with the model calculations that TetR expression is relatively increased in circuit TL than in LT, which suppresses the basal expression of *pLux/tet* and improves the sensitivity and nonlinearity of the promoter to AHL and aTc.

To further validate the metric's utility, another two AND-gate gene circuits (LI and IL) with the position of the genes switched (*LuxR* and *LacI*) were designed (Figure 3C). Hybrid promoter *pLux/lac* was used to indicate the relative concentrations of LuxR and LacI produced from the operon. *LacI* (53.3%, 1,153 bp) has a high GC content, which may increase LuxR expression. Our model calculations showed that LuxR expression increases by 74.3% and LacI increases by 38.1% in circuit IL compared with LI (Table S2). Since promoter *pLux/lac* has two LacI-binding sites (one is in the region between −35 and −10, and the other is downstream of −10 element), so the overall LacI inhibition efficiency is increased ~76.2% considering the importance of spacing between the −35 and −10 elements to RNA polymerase binding. Therefore, the basal GFP expression of logic IL would be lowered compared with LI. The ODE model also indicates higher GFP expression in gate LI (Figures 3D and S3B, solid lines). Experimental results confirmed that the basal expression for circuit LI is ~54-fold higher than IL, and GFP expression under each induction is higher in gate LI, which is consistent with the ODE model results (Figures 3D and S3B).

Taken together, the two sets of AND logic gates are an example of applying our comprehensive model-based tool to evaluate each gene's relative expression level in synthetic AND-gate gene circuits, and verify that ATRs' features and local mRNA stability changes in an operon-based gene network affect protein expression and circuit performance, including basal level, sensitivity, and nonlinearity. Furthermore, the tool could serve as a much-needed quantitative guide for rational design and optimization of gene expression for large genetic circuits.

Tuning Gene Expression with Synthetic 5′ ATRs

In general, the minimum free energy of RNA folding has a negative correlation with GC content (Seffens and Digby, 1999; Trotta, 2014). Next, we sought to use synthetic noncoding
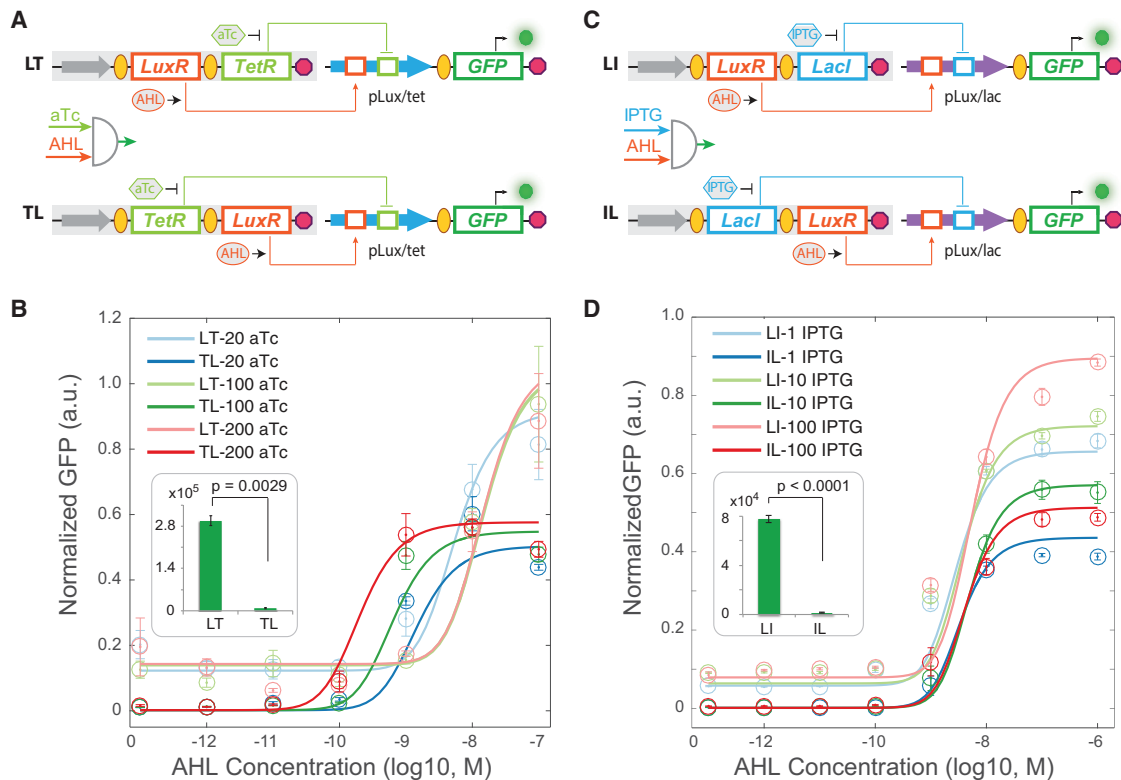
Figure 3. Model-Guided Circuit Design for Synthetic Logic Gates

(A) Two designs for *pLux/tet*–AND logic gate. A constitutive promoter (gray arrow) drives LuxR (orange rectangle) and TetR (green rectangle) expression. *pLux/tet* is highly activated in the presence of both AHL and aTc. LT and TL represent the order of LuxR and TetR positions in the operon. LuxR can bind with AHL (gray oval) to activate *pLux/tet* promoter (blue arrow), while aTc (green hexagon) can block TetR inhibition to *pLux/tet* promoter. Lines with arrowheads indicate activation, and lines with T bars indicate inhibition.

(B) Dose-response curves for different concentrations of AHL and aTc. The solid lines are from ODE model simulations based on the calculated relative changes of LuxR and TetR concentrations in LT and TL from our linear comprehensive model. Data points with error bars are experimental results, showing good match with model predictions. The inset diagram is the basal expression of GFP for design of LT and TL. Color curves are inductions with different aTc concentrations (20 ng/mL, 100 ng/mL, and 200 ng/mL).

(C) Two designs for *pLux/lac*–AND logic gate. A constitutive promoter drives LuxR and LacI expression. *pLux/lac* (purple arrow) is highly activated in the presence of both AHL (N-(β-ketocaproyl)-L-homoserine lactone) and IPTG (isopropyl β-D-1-thiogalactopyranoside, blue hexagon). LuxR can bind with AHL to activate *pLux/lac* promoter, while IPTG can block LacI inhibition to *pLux/lac* promoter. LI and IL represent the order of LuxR and LacI positions in the operon.

(D) Dose-response curves for different concentrations of AHL and IPTG. The solid lines are model simulations based on the calculated relative changes of LuxR and LacI concentrations in LI and IL from our linear comprehensive model. Experimental results (data points with error bar) show good match with model predictions. Color curves are inductions with different IPTG concentrations (1 μM, 10 μM, and 100 μM). Inset diagram is the basal expression of GFP for design of LI and IL. Data represent the mean ± SE of three replicates. p values were calculated using Student's t test.

DNA fragments, with the same size but varying GC content or the same GC content but varying sizes, to fine-tune gene expression in synthetic circuits. We first synthesized six short DNA fragments (with a constant size of 200 bp) with varying GC content from 28% to 53%, which were inserted downstream of the *LuxR* gene but upstream of *GFP* in the two-gene operon (*Promoter-LuxR-Synthetic fragment-GFP*). According to our model, synthetic fragments with varying GC content could tune GFP expression (Figure S3C).

Experimental results show that GFP expression is continuously increased for synthetic fragments with increasing GC content from 28% to 53% (Figure 4A). Low-GC-content fragments downregulated GFP expression about 25-fold. Microscopy results further confirmed flow cytometry data and visualized a gradual increase of fluorescence intensity with increasing GC content ATRs (Figure 4B). Using this strategy, we further synthe-

sized 13 DNA fragments as 5′ATRs with varying GC content but with a constant size (200 bp), and placed downstream of the promoter (Figure 4C). Results indicate that synthetic short DNA sequences have a substantial impact on GFP expression: low-GC-content ATRs largely decrease expression of neighbors (up to 366-fold) and exhibit a gradually increasing pattern from 28% to 48%, while high GC content (48%–67%) ATRs drive GFP expression to a level comparable with the control (without synthetic fragments). It is possible that GFP achieves its maximum expression when the upstream ATR mRNA piece has a relatively stable structure.

To further verify the role of ATR regulation, we varied the size of 5′ATR through shortening and adding a common sequence (Egbert and Klavins, 2012). Using S44 (GC, 44%; size, 200 bp) in Figure 4C as the seed sequence, we shortened it to 100 bp and 50 bp, and lengthened it from 400 bp (combined with two
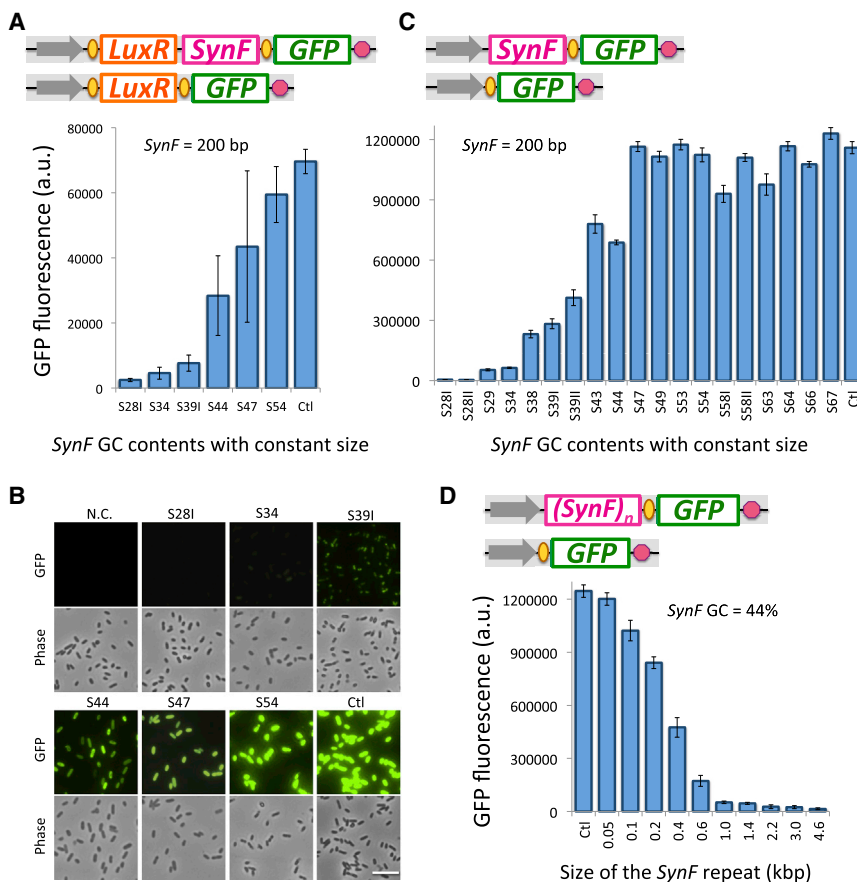
Figure 4. Tuning Gene Expression with Synthetic 5′ Adjacent Transcriptional Regions

(A) Synthetic 5′ATRs (*SynF*) to tune GFP expression for circuit *CP-LuxR-GFP*. 200 bp ATRs were inserted between LuxR and GFP genes to tune GFP expression, and the control (Ctl) was constructed without an ATR insert. Flow cytometry results indicate that GFP fluorescence increases with gradually increasing 5′ATR GC content from 28% to 54%.

(B) Microscopy results for GFP fluorescence for the constructs in (A). Scale bar, 5 μm. Magnification, 40×.

(C) Synthetic 5′ATRs (*SynF*) with different GC content to tune GFP expression for circuit CP-GFP. All the *SynF* are the same size (200 bp) and are inserted upstream of GFP gene (top). Flow cytometry results of GFP fluorescence for 5′ATRs with GC content from 28% to 67% (bottom).

(D) Circuits with different sizes of 5′ATR (through shortening and adding a common sequence S44; GC, 44%; size, 200 bp) were constructed to tune GFP expression. Flow cytometry results show that GFP fluorescence intensity gradually decreases with increasing size of 5′ATR.

Error bars are mean ± SD of at least ten measurements performed on three different days.

pieces of S44) to 4,600 bp (combined with 23 pieces of S44), and all ten fragments have the same GC content (44%, Figure 4D). Model analysis and flow cytometry results show that GFP fluorescence intensity gradually decreases with increasing 5′ATR sizes (Figures S3C and 4D). We also used the data to further refine our comprehensive noncoding model and found three variables $\Delta G_{5'ATR}$, *i*, and $\Delta G_{-70\sim+38}$ are still required for the best fitting efficacy and explains 60.1% of GFP variations (Figure S3D and Table S4). The refined model further expands the variables' range (GC, 28% to 67%; size, 50 to 4,600 bp) and could provide more accurate predictions. Taken together, we demonstrate that synthetic noncoding 5′ATRs with designed GC content and sizes can be used to accurately tune gene expression and achieve expression levels spanning more than 300-fold.

Using Synthetic ATRs to Tune Toggle Switches
Finally, we illustrated the application of synthetic ATRs to modulate the nonlinear bistable potential of synthetic toggle switches. As illustrated in Figure 5A, LacI protein could inhibit TetR by binding the *pLac* promoter, while TetR could bind *pTet* to block LacI expression, forming a mutually inhibitory network. Here, we positioned 200 bp synthetic ATRs with 28% and 67% GC content upstream of RBS-TetR module to tune TetR production (T_S28 and T_S67). According to our analysis above, low-GC-content 5′ATR can downregulate TetR expression, while high GC content can keep TetR at a high level.

Flow cytometry was employed to analyze the initial states of the toggle switches with ATR insertions. As shown in Figure 5B, T_WT initially shows bimodal distribution, GFP-ON and GFP-OFF populations, resulting from gene expression noise in a relatively balanced system. In contrast, both T_S28 and T_S67 exhibited unimodal distributions. Synthetic ATR S28 decreased TetR expression leading to higher LacI and GFP expression, whereas the fragment with 67% GC content showed a lower GFP expression than T_S28 and slightly lower than the high GFP population cells in T_WT (Figure 5B). The results indicate that the synthetic ATRs can tune the initial steady states of toggle switches and modulate the population from bimodal to unimodal distributions.

To achieve a quantitative understanding of the ATR's regulation on bistability, we performed bifurcation analysis from the same mathematical model as the classical toggle switch (Gardner et al., 2000). We found that the production rate of TetR has a considerable effect on bistability and the bistable region. A small production rate, corresponding to low-GC ATR, has a small bistable region, whereas an increase in the production rate leads to a larger bistable region (Figure 5C). Experimentally, hysteresis of the three toggles was tested to verify the model analysis. The results indicate that all three toggles exhibited hysteresis, and T_WT harbors the broadest bistable region (Figures 5D–5F). Moreover, consistent with model analysis, the bistable regions gradually decreased from T_WT to T_S67 to T_S28. Collectively, these results validate a novel strategy of using synthetic ATRs to tune the initial steady states and bistability of gene networks. Furthermore, this example demonstrates the feasibility of bridging ATR regulation with mathematical modeling to quantitatively understand and tune gene network dynamics.
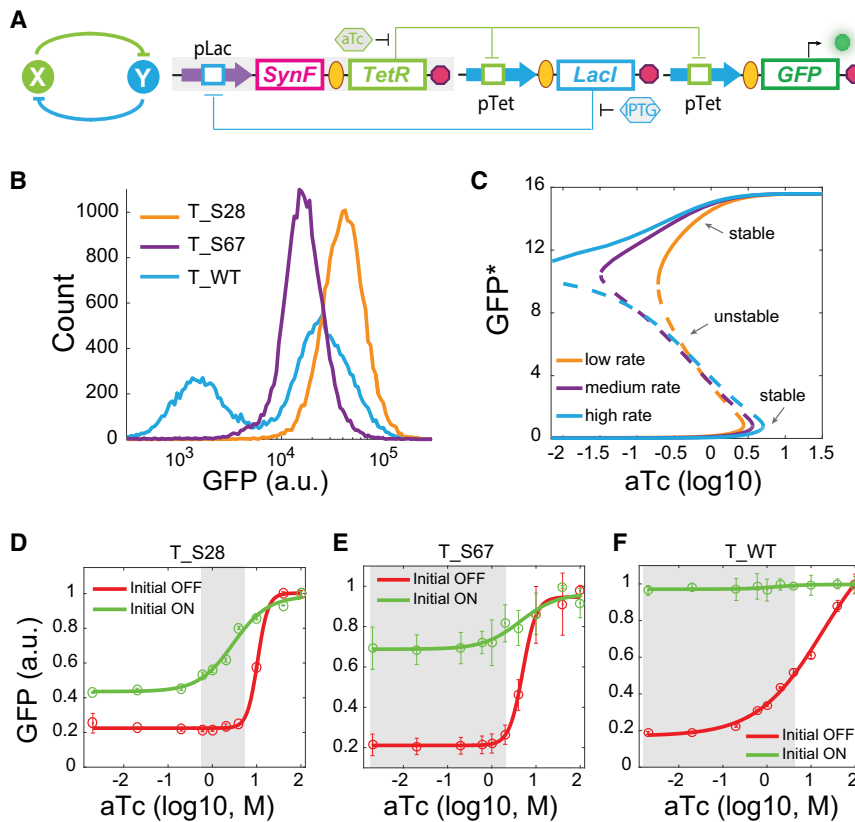
Figure 5. Using Synthetic ATRs to Modulate Bistability of Toggle Switches

(A) Left: Abstract diagram of toggle switch topology, where X and Y mutually inhibit each other. Right: Molecular implementation of the toggle switch. LacI inhibits TetR by binding the *pLac* promoter, while TetR binds *pTet* to block LacI expression, forming a mutually inhibitory network. Inducers IPTG and aTc (hexagon) can relieve LacI and TetR inhibition, respectively. GFP serves as the readout of the *pTet* promoter. Synthetic ATRs (*SynF*) were arranged right upstream of the *TetR* gene.

(B) Initial steady states for the three toggles. Toggle without ATR insertion (T_WT) shows bimodal distribution (GFP-OFF and GFP-ON), while T_S28 (ATR with 28% GC content, 200 bp) shows higher GFP expression and T_S67 (ATR with 67% GC content, 200 bp) shows lower GFP expression than the GFP-ON population of T_WT.

(C) Bifurcation analysis for GFP (LacI) expression with different TetR production rates under induction of varying concentrations of aTc. A low production rate for TetR, corresponding to T_S28, has the smallest bistable region, while a high rate (corresponding to T_WT) has the broadest bistable region. Solid lines represent stable steady-state solutions and dotted lines are unstable steady-state solutions. GFP* is the computed GFP abundance from the model.

(D–F) Hysteresis results for toggles (D) T_S28, (E) T_S67, and (F) T_WT under induction of varying concentrations of aTc. Red lines indicate the initial OFF cells with basal GFP expression, while green lines indicate the initial ON cells with high GFP expression. Data represent the mean ± SD of three replicates. The gray area is the presumed bistable region for each circuit.

DISCUSSION

Circuit engineering is the first step for synthetic biologists to achieve designed functionalities with synthetic gene circuits. A successful synthetic gene circuit depends on full characterization of the biological components and the interactions that emerge between modules when assembled into a complete gene network (Bennett and Hasty, 2009; Brophy and Voigt, 2014; Nielsen et al., 2016; Wu et al., 2017). Development of a reliable tool to predict protein expression in the circuit has wide applications in biotechnology. For example, RBS Calculator is a well-developed design tool to predict and control translation initiation and protein expression in bacteria (Farasat et al., 2014; Salis et al., 2009).

Here, we systematically investigated how adjacent transcriptional regions affect protein expression in synthetic operon-based gene circuits. Through placing the GFP at different positions (proximal, middle, and distal) to the promoter, we developed a new protein expression metric that takes into account the features of adjacent transcriptional regions, including GC content, size, and stability of mRNA folding near RBS (Figure 2). The metric was established from about 120 gene circuits, which to our knowledge represents one of the largest databases of operon-based synthetic gene circuits in one study so far. This metric explains 63% and 67% of GFP variations in the coding

ATR and noncoding ATR polycistronic gene circuits, respectively. Moreover, our experimental results also demonstrated the metric's predictions of gene expression changes and induced nonlinear dynamic responses in different genetic contexts (Figures 3, 4, and 5), suggesting the model's utility in guiding circuit design. Most ATRs in the circuits were 500–2,000 bp, and the maximum is 2,422 bp, which may undermine the contribution of ATR size to GFP variation. Moreover, because of the limitation in sample size and available gene resources, the collected data are not perfectly normally distributed, especially for circuits with GFP in the middle (*X-GFP-Y*), which may compromise the robustness of the model.

Consistent with previous results that gene position in operons can affect gene expression (Chizzolini et al., 2014; Lim et al., 2011), our results further demonstrate that gene position (corresponding to change of ATR) significantly altered gene network dynamics, including basal expression, system sensitivity, and nonlinearity, which has profound impacts for nonlinear dynamic systems. Such an adjacent gene regulation effect has been generally neglected during construction of synthetic gene networks.

Although it is relatively well established that gene expression is influenced by the local context, holistic understanding of architectural rules governing polycistronic gene circuits remains largely unexplored. Compared with previous gene expression

tuning strategies or insulation strategies, such as RBS Calculator, bicistronic design with translation of a short leader peptide, or a designed DNA sequence surrounding the start codon (mostly less than 100 bp) (Farasat et al., 2014; Ferreira et al., 2013; Li et al., 2017; Mutalik et al., 2013; Salis et al., 2009), our work places more emphasis on whether and how polycistronic operon organization (*X-GFP*, *X-GFP-Y*, and *GFP-X*) and different adjacent genes (size ranging from 313 to 2,362 bp, and GC content ranging from 30.3% to 60.4%) affect protein expression in operon-based gene circuits. Furthermore, we validated that the usage of designed synthetic DNA fragments with either different GC content (28%–67%) or size (50–4,600 bp) as 5′ATRs tuned gene expression and modulated bistable regions of genetic toggle switches. The synthetic ATRs have a wide variable interval, therefore making them potentially applicable to a broad range of scientific and engineering tasks. Such a gene expression tuning strategy also avoids the production of unwanted peptides and hence reduces potential metabolic burden. We also observed that circuits having different ATRs have an impact on the time that cells reach stationary phase with similar optical density (Figure S4B), suggesting that ATRs could be used as a means to "program" the metabolic load and fitness of a cell simultaneously.

Our results show that the context dependency of gene expression is not just limited to the RBS region but also includes characteristics of the whole operon. This "global" effect in polycistronic operons could be quantified by a biophysical model, which explains nearly two-thirds of protein expression variations across all the circuits with different configurations. The quantitative relationship between adjacent transcriptional regions and gene expression regulation in polycistronic circuits helps to evaluate each gene's relative expression levels in a circuit and predict circuit outputs, which would save experimentalist's time and resources to screen and test combinations of modules, and thus should greatly facilitate optimization of circuit design and accelerate the engineering of complex gene networks.

A central goal of synthetic biology is to develop genetic circuits to program cell behaviors in a predictable way. With the increasing complexity of integrated multi-layer circuits, organization of specific bio-components and circuitry structure design become extremely important for functionality (Chen et al., 2015b; Nielsen et al., 2016; Wu et al., 2014). The tool we provide here could serve as a much-needed quantitative guide for rational design and optimization of gene expression for large genetic circuits.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- CONTACT FOR REAGENT AND RESOURCE SHARING
- EXPERIMENTAL MODEL AND SUBJECT DETAILS
  - ○ Strains, Media, and Culture Conditions
- METHODS DETAILS
  - ○ Plasmid Construction
  - ○ Minimum Free Energy Calculation
  - ○ RT-qPCR
  - ○ Flow Cytometry Measurements
  - ○ Hysteresis Experiment
  - ○ Sample Preparation and Microscopy
  - ○ Growth Curve Assay
- QUANTIFICATION AND STATISTICAL ANALYSIS
  - ○ Statistical Analysis and Comprehensive Model Development
  - ○ Deterministic Model Construction and Prediction for the Logic Gate
  - ○ Bifurcation Analysis for the Synthetic Toggle Switches
- DATA AND SOFTWARE AVAILABILITY

AUTHOR CONTRIBUTIONS

F.W. and X.W. designed the research. F.W. and Q.Z. performed the experiments. F.W., Q.Z., and X.W. analyzed the data and wrote the manuscript.

REFERENCES

Ahnert, S.E., Fink, T.M., and Zinovyev, A. (2008). How much non-coding DNA do eukaryotes require? J. Theor. Biol. *252*, 587–592.

Bennett, M.R., and Hasty, J. (2009). Overpowering the component problem. Nat. Biotechnol. *27*, 450–451.

Brophy, J.A., and Voigt, C.A. (2016). Antisense transcription as a tool to tune gene expression. Mol. Syst. Biol. *12*, 854.

Brophy, J.A.N., and Voigt, C.A. (2014). Principles of genetic circuit design. Nat. Methods *11*, 508–520.

Cameron, D.E., and Collins, J.J. (2014). Tunable protein degradation in bacteria. Nat. Biotechnol. *32*, 1276–1281.

Cao, Y., Ryser, M.D., Payne, S., Li, B., Rao, C.V., and You, L. (2016). Collective space-sensing coordinates pattern scaling in engineered bacteria. Cell *165*, 620–630.

Carr, S.B., Beal, J., and Densmore, D.M. (2017). Reducing DNA context dependence in bacterial promoters. PLoS One *12*, e0176013.

Chen, H., Shiroguchi, K., Ge, H., and Xie, X.S. (2015a). Genome-wide study of mRNA degradation and transcript elongation in *Escherichia coli*. Mol. Syst. Biol. *11*, 781.

Chen, Y., Kim, J.K., Hirning, A.J., Josić, K., and Bennett, M.R. (2015b). Emergent genetic oscillations in a synthetic microbial consortium. Science *349*, 986–989.

Chizzolini, F., Forlin, M., Cecchi, D., and Mansy, S.S. (2014). Gene position more strongly influences cell-free protein expression from operons than T7 transcriptional promoter strength. ACS Synth. Biol. *3*, 363–371.

Din, M.O., Danino, T., Prindle, A., Skalak, M., Selimkhanov, J., Allen, K., Julio, E., Atolia, E., Tsimring, L.S., Bhatia, S.N., et al. (2016). Synchronized cycles of bacterial lysis for in vivo delivery. Nature *536*, 81–85.

Egbert, R.G., and Klavins, E. (2012). Fine-tuning gene networks using simple sequence repeats. Proc. Natl. Acad. Sci. USA *109*, 16817–16822.

Elowitz, M.B., and Leibler, S. (2000). A synthetic oscillatory network of transcriptional regulators. Nature *403*, 335–338.

Emory, S.A., Bouvet, P., and Belasco, J.G. (1992). A 5'-terminal stem-loop structure can stabilize mRNA in *Escherichia coli*. Genes Dev. *6*, 135–148.

Espah Borujeni, A., and Salis, H.M. (2016). Translation initiation is controlled by RNA folding kinetics via a ribosome drafting mechanism. J. Am. Chem. Soc. *138*, 7016–7023.

Farasat, I., Kushwaha, M., Collens, J., Easterbrook, M., Guido, M., and Salis, H.M. (2014). Efficient search, mapping, and optimization of multi-protein genetic systems in diverse bacteria. Mol. Syst. Biol. *10*, 731.

Ferreira, J.P., Overton, K.W., and Wang, C.L. (2013). Tuning gene expression with synthetic upstream open reading frames. Proc. Natl. Acad. Sci. USA *110*, 11284–11289.

Gardner, T.S., Cantor, C.R., and Collins, J.J. (2000). Construction of a genetic toggle switch in *Escherichia coli*. Nature *403*, 339–342.

Kudla, G., Murray, A.W., Tollervey, D., and Plotkin, J.B. (2009). Coding-sequence determinants of gene expression in *Escherichia coli*. Science *324*, 255–258.

Lee, J.W., Gyorgy, A., Cameron, D.E., Pyenson, N., Choi, K.R., Way, J.C., Silver, P.A., Vecchio, D.D., and Collins, J.J. (2016). Creating single-copy genetic circuits. Mol. Cell *63*, 329–336.

Li, J., Liang, Q., Song, W., and Marchisio, M.A. (2017). Nucleotides upstream of the Kozak sequence strongly influence gene expression in the yeast *S. cerevisiae*. J. Biol. Eng. *11*, 25.

Lim, H.N., Lee, Y., and Hussein, R. (2011). Fundamental relationship between operon organization and gene expression. Proc. Natl. Acad. Sci. USA *108*, 10626–10631.

Litcofsky, K.D., Afeyan, R.B., Krom, R.J., Khalil, A.S., and Collins, J.J. (2012). Iterative plug-and-play methodology for constructing and modifying synthetic gene networks. Nat. Methods *9*, 1077–1080.

Liu, C., Fu, X., Liu, L., Ren, X., Chau, C.K., Li, S., Xiang, L., Zeng, H., Chen, G., Tang, L.H., et al. (2011). Sequential establishment of stripe patterns in an expanding cell population. Science *334*, 238–241.

Ma, K.C., Perli, S.D., and Lu, T.K. (2016). Foundations and emerging paradigms for computing in living cells. J. Mol. Biol. *428*, 893–915.

Mackie, G.A. (2013). RNase E: at the interface of bacterial RNA processing and decay. Nat. Rev. Microbiol. *11*, 45–57.

Mao, Y., Liu, H., Liu, Y., and Tao, S. (2014). Deciphering the rules by which dynamics of mRNA secondary structure affect translation efficiency in *Saccharomyces cerevisiae*. Nucleic Acids Res. *42*, 4813–4822.

Mus, F., Crook, M.B., Garcia, K., Costas, A.G., Geddes, B.A., Kouri, E.-D., Paramasivan, P., Ryu, M.-H., Oldroyd, G.E.D., Poole, P.S., et al. (2016). Symbiotic nitrogen fixation and challenges to extending it to non-legumes. Appl. Environ. Microbiol. https://doi.org/10.1128/AEM.01055-16.

Mutalik, V.K., Guimaraes, J.C., Cambray, G., Lam, C., Christoffersen, M.J., Mai, Q.A., Tran, A.B., Paull, M., Keasling, J.D., Arkin, A.P., et al. (2013). Precise and reliable gene expression via standard transcription and translation initiation elements. Nat. Methods *10*, 354–360.

Nielsen, A.A.K., Der, B.S., Shin, J., Vaidyanathan, P., Paralanov, V., Strychalski, E.A., Ross, D., Densmore, D., and Voigt, C.A. (2016). Genetic circuit design automation. Science *352*, aac7341.

Oliva, G., Sahr, T., and Buchrieser, C. (2015). Small RNAs, 5' UTR elements and RNA-binding proteins in intracellular bacteria: impact on metabolism and virulence. FEMS Microbiol. Rev. *39*, 331–349.

Pardee, K., Green, A.A., Ferrante, T., Cameron, D.E., DaleyKeyser, A., Yin, P., and Collins, J.J. (2014). Paper-based synthetic gene networks. Cell *159*, 940–954.

Pardee, K., Green, A.A., Takahashi, M.K., Braff, D., Lambert, G., Lee, J.W., Ferrante, T., Ma, D., Donghia, N., Fan, M., et al. (2016). Rapid, low-cost detection of zika virus using programmable biomolecular components. Cell *165*, 1255–1266.

Prindle, A., Selimkhanov, J., Li, H., Razinkov, I., Tsimring, L.S., and Hasty, J. (2014). Rapid and tunable post-translational coupling of genetic circuits. Nature *508*, 387–391.

Rocha, E.P. (2008). The organization of the bacterial genome. Annu. Rev. Genet. *42*, 211–233.

Salis, H.M., Mirsky, E.A., and Voigt, C.A. (2009). Automated design of synthetic ribosome binding sites to precisely control protein expression. Nat. Biotechnol. *27*, 946–950.

Seffens, W., and Digby, D. (1999). mRNAs have greater negative folding free energies than shuffled or codon choice randomized sequences. Nucleic Acids Res. *27*, 1578–1584.

Selinger, D.W., Saxena, R.M., Cheung, K.J., Church, G.M., and Rosenow, C. (2003). Global RNA half-life analysis in *Escherichia coli* reveals positional patterns of transcript degradation. Genome Res. *13*, 216–223.

Serra, M.J., and Turner, D.H. (1995). Predicting thermodynamic properties of RNA. Methods Enzymol. *259*, 242–261.

Smanski, M.J., Zhou, H., Claesen, J., Shen, B., Fischbach, M.A., and Voigt, C.A. (2016). Synthetic biology to access and expand nature's chemical diversity. Nat. Rev. Microbiol. *14*, 135–149.

de Smit, M.H., and van Duin, J. (1990). Secondary structure of the ribosome binding site determines translational efficiency: a quantitative analysis. Proc. Natl. Acad. Sci. USA *87*, 7668–7672.

Taniguchi, Y., Choi, P.J., Li, G.W., Chen, H., Babu, M., Hearn, J., Emili, A., and Xie, X.S. (2010). Quantifying *E. coli* proteome and transcriptome with single-molecule sensitivity in single cells. Science *329*, 533–538.

Trotta, E. (2014). On the normalization of the minimum free energy of RNAs by sequence length. PLoS One *9*, e113380.

Tuller, T., Waldman, Y.Y., Kupiec, M., and Ruppin, E. (2010). Translation efficiency is determined by both codon bias and folding energy. Proc. Natl. Acad. Sci. USA *107*, 3645–3650.

Wright, G. (2014). Perspective: synthetic biology revives antibiotics. Nature *509*, S13.

Wu, F., and Wang, X. (2015). Applications of synthetic gene networks. Sci. Prog. *98*, 244–252.

Wu, F., Menn, D.J., and Wang, X. (2014). Quorum-sensing crosstalk-driven synthetic circuits: from unimodality to trimodality. Chem. Biol. *21*, 1629–1638.

Wu, F., Su, R.Q., Lai, Y.C., and Wang, X. (2017). Engineering of a synthetic quadrastable gene network to approach Waddington landscape and cell fate determination. Elife *6*, e23702.

Xia, T., SantaLucia, J., Burkard, M.E., Kierzek, R., Schroeder, S.J., Jiao, X., Cox, C., and Turner, D.H. (1998). Thermodynamic parameters for an expanded nearest-neighbor model for formation of RNA duplexes with Watson-Crick base pairs. Biochemistry *37*, 14719–14735.

Yang, L., Nielsen, A.A.K., Fernandez-Rodriguez, J., McClune, C.J., Laub, M.T., Lu, T.K., and Voigt, C.A. (2014). Permanent genetic memory with >1-byte capacity. Nat. Methods *11*, 1261–1266.

Yeung, E., Dy, A.J., Martin, K.B., Ng, A.H., Del Vecchio, D., Beck, J.L., Collins, J.J., and Murray, R.M. (2017). Biophysical constraints arising from compositional context in synthetic gene networks. Cell Syst. *5*, 11–24.e12.

Zadeh, J.N., Steenberg, C.D., Bois, J.S., Wolfe, B.R., Pierce, M.B., Khan, A.R., Dirks, R.M., and Pierce, N.A. (2011). NUPACK: analysis and design of nucleic acid systems. J. Comput. Chem. *32*, 170–173.

Zhang, W., and Nielsen, D.R. (2014). Synthetic biology applications in industrial microbiology. Front. Microbiol. *5*, 451.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Bacterial and Virus Strains		
Escherichia coli DH10B	Invitrogen	Cat# 18297010
Escherichia coli MG1655 strain	James Collins laboratory (Litcofsky et al., 2012)	N/A
Escherichia coli MG1655 strain with *lacI-/-*	James Collins laboratory (Litcofsky et al., 2012)	N/A
Chemicals, Peptides, and Recombinant Proteins		
AHL (N-(β-Ketocaproyl)-L-homoserine lactone)	Sigma-Aldrich	Cat# 143537-62-6
Isopropyl β-D-1-thiogalactopyranoside	Sigma-Aldrich	Cat# 367-93-1
Anhydrotetracycline hydrochloride	Sigma-Aldrich	Cat# 13803-65-1
Ampicillin sodium salt	Sigma-Aldrich	Cat# A9518
Oligonucleotides		
Primers for gene amplification (See Table S6)	This paper	N/A
gBlocks for synthetic ATRs (See Table S6)	This paper	N/A
Recombinant DNA		
Plasmids with Biobricks parts (See Table S1)	This paper	N/A
Plasmids with amplified transcriptional factors (See Table S1)	This paper	N/A
Plasmids with synthetic ATRs (See Table S6)	This paper	N/A
Software and Algorithms		
MATLAB version R2014b	MathWorks	https://www.mathworks.com/
SAS 9.4	SAS Institute	https://www.sas.com/
XLSTAT version 2017.4	Addinsoft	https://www.xlstat.com/
XPP AUTO 8.0	University of Pittsburgh	http://www.math.pitt.edu/~bard/xpp/xpp.html
Deposited Data		
Data generating the comprehensive biophysical models and code for the model simulation and statistical analysis (See Data S1)	This paper	N/A
DNA sequence data for all the synthetic circuits (See Table S6)	This paper	N/A

CONTACT FOR REAGENT AND RESOURCE SHARING

Further information and requests for resources and reagents should be directed to and will be fulfilled by the Lead Contact, Xiao Wang (xiaowang@asu.edu).

EXPERIMENTAL MODEL AND SUBJECT DETAILS

Strains, Media, and Culture Conditions

All cloning experiments and fluorescent measurements were performed in *Escherichia coli* DH10B (Invitrogen). Synthetic toggle switches (T_S28, T_S67 and T_WT) were tested in *E. coli* K-12 MG1655 strain with *lacI-/-* (Litcofsky et al., 2012). Cells were cultured in liquid or solid Luria-Bertani (LB) broth medium with 100 μg/ml ampicillin at 37°C. Chemicals AHL (N-(β-Ketocaproyl)-L-homoserine lactone), IPTG (isopropyl β-D-1-thiogalactopyranoside), and aTc (anhydrotetracycline) were dissolved in ddH2O and diluted into indicated working concentrations. Cultures were shaken in 5 mL and/or 15 mL tubes at 220 rotations per minute (r.p.m).

METHODS DETAILS

Plasmid Construction

Most genes are obtained from iGEM Registry (http://parts.igem.org/Main_Page). These genes are often used in synthetic biology projects, including transcription factors, quorum-sensing components, and other functional genes (Table S1). Plasmids

were constructed using standard molecular biology techniques and all genetic circuits were assembled based on standardized BioBrick methods. As an example, construct *Promoter-TetR-GFP* is composed of five BioBrick standard biological parts: BBa_J23104 (constitutive promoter, CP), BBa_B0034 (ribosome binding site, RBS), BBa_C0040 (*tetR*), BBa_E0040 (green fluorescent protein, GFP) and BBa_B0015 (transcriptional terminator). To produce RBS-TetR module, plasmid containing *TetR* was digested by *XbaI* and *PstI* as the insert fragment while RBS vector was cut by *SpeI* and *PstI*. Both fragment and vector were separated on 1% TAE agarose gel electrophoresis and purified using PureLink gel extraction Kit (Invitrogen). Purified fragment and vector were then ligated by T4 DNA ligase (New England Biolabs, NEB). The ligation products were further transformed into *E. coli* DH10B and plated on LB agar plate with 100 μg/ml ampicillin for screening. Finally, plasmids extracted by GenElute HP MiniPrep Kit (SIGMA-ALDRICH) were confirmed through gel electrophoresis (digested by *EcoRI* and *PstI*) and DNA Sequencing (Biodesign sequencing Lab, ASU). Similar steps were carried out for subsequent rounds of cloning to assemble the whole construct. All the circuits' DNA sequences are provided in the Table S6.

Also, 17 transcriptional factors with varying GC content and sizes used in Figure 2D were amplified from *E. coli* genome with designed primers (Table S1). Synthetic sequences were randomly generated with the same length (200 bp) but various GC contents (28%-67%). Sequences with ribosome binding site-features (AGGAGG) were redesigned to exclude its translation potential. All synthetic sequences and primers were synthesized as custom DNA oligos or gBlocks gene fragments from Integrated DNA Technologies (IDT). In order to express consistently in the cell, all constructs were finally subcloned into pSB1A3 vector prior to the test.

Minimum Free Energy Calculation

All minimum free energy (MFE) of mRNAs were computed on Nucleic Acid Package (NUPACK) web server (Zadeh et al., 2011). Specifically, we chose Serra and Turner parameter set as the RNA energy parameter and set 37°C, 1.0 M Na+ and 0 M Mg^{2+} to be the prediction algorithm (Serra and Turner, 1995). $\Delta G_{5'ATR}$ and $\Delta G_{3'ATR\_100}$ were calculated from sequence including ATR (with or without RBS), and the two scar sequences introduced during cloning process. $\Delta G_{-70\sim+38}$ is obtained from 70 nt upstream sequence and 38 nt downstream around ATG (+1) codon of GFP gene.

RT-qPCR

Total RNA was extracted from three individual cell cultures (1.5 mL exponentially growing cell cultures, fresh cultures) for each construct in Figure 1B using Trizol (Invitron). DNase I (NEB) was used to remove traces of genomic DNA and then the total RNA was further purified using purelink RNA Mini Kit (Life technologies), and the eluted total RNA was quantified using BioTek's Synergy H1multi-mode Reader. cDNA was synthesized from RNA using an iScript cDNA synthesis kit and random primers (Bio-Rad). The reaction volume is 20 μL and ~1 μg RNA were used for reaction. Concentrations of cDNA are then quantified by qPCR using iTaq Universal SYBR Green Supermix (Bio-Rad) with the iQ5 Real-Time PCR detection system (Bio-Rad). Prokaryotic 16S rRNA was employed as endogenous control. Primers (IDT) used for amplifying 16S rRNA: 5'-*GAATGCCACGGTGAATACGTT*-3' (*rrnB*, forward, starting at the 1361st nucleotide), and 5'-*CACAAAGTGGTAAGCGCCCT*-3' (*rrnB*, reverse, starting at the 1475th nucleotide) (Lim et al., 2011). Two pairs of primers were designed to amplify GFP are P1: 5'-*CAGTGGAGAGGGTGAAGGTGA*-3' (forward, starting at the 87th nucleotide); and P2: 5'-*CCTGTACATAACCTTCGGGCAT*-3' (reverse, starting at the 283th nucleotide); P3: 5'-*AGACACGTGCTGAAGTCAAG*-3' (forward, starting at the 320th nucleotide); and P4: 5'-*TCTGCTAGTTGAACGCTTCCAT*-3' (reverse, starting at the 539th nucleotide). qPCR result is analyzed using Bio-rad CFX Manager software version 3.1. Each sample was performed with two replicates for both 16S rRNA and GFP cDNAs, and gene expression was normalized to 16S rRNA. Delta C_t values were calculated ($C_t^{target} - C_t^{16S}$) and compared with the biological control (*Constitutive promoter-RBS-GFP*) to calculate the relative GFP mRNA concentrations. The minimum information for publication of quantitative real-time PCR (MIQE) is also provided in Table S5.

Flow Cytometry Measurements

All confirmed constructs were re-transformed into DH10B strain. Single colonies were picked and cultured in 4 mL LB medium (100 μg/ml ampicillin) for 24 hr at 37°C for testing. Flow cytometry measurements were performed using Accuri C6 flow cytometer (Becton Dickinson) and all samples were analyzed at twelve hour and twenty-four hour time points, and the two time points showed similar GFP expression pattern (Figure S4A). GFP excitation: 488 nm, and emission: 530 ± 15 nm. All data were collected in a log mode. Data files were further analyzed by MATLAB (MathWorks). All the fluorescence data are collected by flow cytometry unless specified, and the fluorescence was not normalized against cell density because we measured the fluorescence of single cells, instead of the population, so the fluorescence value is not directly correlated with population density. 20,000 individual cells were analyzed for each sample at a slow flow rate.

Hysteresis Experiment

All synthetic toggle switch plasmids (T_S28, T_S67 and T_WT) were transformed into K-12 MG1655 strain with *lacI*-/-, and cells cultured overnight in LB medium. T_WT plasmid has been used in previous study (Wu et al., 2017). We prepared two pre-cultures with two initially different stable steady states, i.e., low GFP state (OFF) without inductions and high GFP state (ON) induced with enough aTc. The two cells were then inoculated into media containing an aTc concentration range so that cells with different initial conditions were grown in identical conditions. Specifically, for OFF-ON experiment, samples were diluted evenly into 5 ml polypropylene round-bottom tubes (Falcon) and induced with different amounts of aTc. Fluorescence was then measured at 6, 8 and 21 hr time points to monitor the fluorescence level. In our experiment, we found the intensity of fluorescence became stable after ~8 hr

induction. For the ON-OFF experiment, cells were induced with 40 ng/ml aTc initially to prepare the initial ON cells and fluorescence was measured at 8 hr to ensure they were fully induced. The initial ON cells were then collected by low-speed centrifugation, washed once to remove the inducer, resuspended with LB medium, diluted and transferred into fresh medium with various aTc concentrations at 1:100 ratio. Flow cytometry measurement was performed for each sample after 6, 10 and 18 hr culturing, respectively. Data shown in Figure 5 are 18 hr results.

Sample Preparation and Microscopy
Single colonies were picked and grew at 37°C in liquid LB medium. After 24 hours, 1 mL cells were collected and spun down at 2500 g for 5 min, washed with 1x phosphate buffer solution (PBS), and resuspended by 200 μL 1xPBS. 5 μL of concentrated cell solution was placed on glass microscope slides and images were captured with a Nikon Ti-Eclipse inverted microscope (magnification 40x). GFP was visualized with an excitation at 472 nm and emission at 520/35 nm using a Semrock band-pass filter. The exposure time for each sample is kept the same.

Growth Curve Assay
Ten different gene circuits with different fluorescence expression levels (high, medium and low) are selected to test their growth rates under the same condition. Single colonies harboring circuit plasmid were picked up and diluted into 4 mL LB medium, from which 300 μL were transferred into 96-well sterile plate. A negative control with only LB medium was also prepared. Optical density (OD600) and fluorescence (excitation: 485 nm; emission: 530 nm) were measured every 30 minutes by plate reader (BioTek) over 20 hours with shaking platform and temperature control (37°C). Three random colonies were picked up and triplicate wells were measured for each sample. Our results indicated that gene expression in the circuit influenced the time of cells going to the exponential phase, but all the samples went to stationary phase with similar OD value after ∼12 hr (Figure S4B). For stable protein expression, we chose the 24 hr data point in this study unless specified.

QUANTIFICATION AND STATISTICAL ANALYSIS

Statistical Analysis and Comprehensive Model Development
To investigate the correlation between GFP expression and sequence characteristics in different circuits with different genes and organizations, we performed multiple linear regression analysis using the classical statistical software SAS 9.4. Here, we mainly focused on five different independent variables including 5'- and 3'-ATR GC content (variable is computed as a percentage), 5'- and 3'- size (variable is computed as segment length), and $\Delta G_{-70\sim+38}$, all of which can be computed from the DNA sequence in each circuit. The dependent variable is GFP fluorescence measured by flow cytometry, which was transformed to log scale during analysis. Eight data points collected in three days were used for regression analysis in Figures 1 and 2, and twelve data points were collected in three different days for the 17 transcriptional factors insertion as non-coding ATRs, and all of the collected data points are imported to SAS for analysis (Data S1).

   All the information of the five variables is calculated from the specific DNA sequence. The 5'ATR includes the sequence from the scar right after the promoter to the scar right before the RBS of GFP. And the 3'ATR includes the sequence from the scar right after the GFP to the scar right before the terminator. The scar sequence is generated from the molecular cloning using biobrick modules, and the size is 6 or 8 nucleotides. GC content and size of ATRs are calculated using the web server Endmemo (http://www.endmemo.com/index.php). $\Delta G_{-70\sim+38}$ were computed using NUPACK web tool (http://www.nupack.org). Since the ΔG are negative values, log transformations were performed to the absolute value of ΔG, and then set to negative value. To build a comprehensive model for all the scenarios in Figures 2A–2C (GFP-X, X-GFP-Y, and X-GFP), we introduced dummy values for some variables in some regression analyses for analytical convenience. For example, construct GFP-X (Figure 2C) has no varied 5'ATR (only a RBS and scar sequence), and its GC content value is set to 0.04 instead of 0. Similarly, $\Delta G_{5'ATR}$ is set to −0.05 for constructs without 5'ATR, and $\Delta G_{3'ATR\_100}$ is set to −0.00001 for constructs without 3'ATR. These dummy values don't significantly influence model fitting efficiency.

   We first use scatter plot to display the relationship between GFP and each of the variables we are interested, without any data transformation. As shown in Figure S5, the data has a large variability ranging from 21,000 to 1900,000 (arbitrary unit), and the fit without transformation is weakly linear and heteroscedastic. It would be problematic to use linear data for regression because of the inconstant variance from the data. However, the log is a variance stabilizing transformation, and it clearly reduced changes in variability of the data along the x-axis (Figures 2A–2D). Furthermore, transformed data conforms to a nearly normal distribution (Figure S4C), more easily enabling us to perform multiple regression analysis to find a quantitative estimation of the relationship between GFP and the other three or five variables together.

   To explore possible mechanistic basis of ATR regulation, we developed a comprehensive linear model based on the sequence-dependent energetic changes during the polycistronic mRNA folding and translation and the costs of protein biosynthesis. The biophysical model was based on previous pioneer work characterizing the relationship between free energy changes and protein translation initiation (Espah Borujeni and Salis, 2016; Salis et al., 2009; de Smit and van Duin, 1990). We calculated the free energies for 5'ATR and the first 100 nucleotides of 3'ATR ($\Delta G_{5'ATR}$ and $\Delta G_{3'ATR\_100}$) using NUPACK. Since all the energy terms are negative values, absolute values were first acquired for each of them and then set to negative values for data analysis. The constant Gm is set to 1, and for cases of non-coding ATRs, the coefficients for j and $\Delta G_{3'ATR\_100}$ are set to 0, owing to a lack of 3'ATRs.

To find the linear comprehensive coding-ATR model having the best prediction of dependent variable from the independent variables, we performed stepwise regression with the five variables: $\Delta G_{5'ATR}$, $\Delta G_{3'ATR\_100}$, 5'ATR size, 3'ATR size and $\Delta G_{-70\sim+38}$. Stepwise regression is an automated tool for model selection through adding the most significant variable or removing the least significant variable as needed for each step (the significance level for variable entry or stay is 0.05). From the sequence of generated models, the selected model is chosen based on the lowest Akaike information criterion. Results showed that all the five variables are statistically significant for the best prediction of GFP expression in the comprehensive coding-ATR model, and explains 63% of GFP variations (Table S3). It is necessary to note that the negative correlation between protein abundance (*c*) and the sum of energetic terms ($\Sigma\beta_x\Delta G_x$) in the equation is already reflected in the coefficients of each term.

The fitting diagnostics indicated that there is no apparent trend for the residuals, and the data is roughly normally distributed, and the variables in the model explain most variation in the response variable from the residual-fit result (Figure S4C). The predicted value by observed GFP plot (Predicted Value - logGFP) reveals a reasonably successful model for explaining the variation in GFP for most of the circuits (Figure 2F, left panel and Figure S4C). The predicted responses (logGFP value) are calculated according to the generated linear regression model, with the corresponding inputs from each circuit. And a plot of predicted GFP against experimentally observed GFP values are then generated to evaluate and visualize the model-fitting efficacy (Figure 2F). If the model predicted values and observed values agreed perfectly (R^2 = 100%), all the data points would fall on the dotted diagonal line of the squares. However, several outliers in the combined model are also observed and some observations with high leverages might also be overly influencing the fit result (Figure S4D). Of the outliers, most of them are corresponding to specific circuits, such as outliers 217~224 corresponding to the tricistronic circuit (*promoter-luxR-appY-GFP,* has 8 data points). Observations with high leverages such as 505~512 are corresponding to the circuit *promoter-GFP-Zif23_GCN4*. Moreover, some outliers are also high-leverage observations. Given the data sample size (N = 632), the original data collection, and the overall data-fitting efficacy, we here didn't exclude the outliers or data with very high leverages (although that would improve the model-fitting efficacy).

Similar analysis was also applied to the data with non-coding ATR, and results showed that 5'ATR size and folding energy $\Delta G_{5'ATR}$, local mRNA folding energy $\Delta G_{-70\sim+38}$ are crucial for the best prediction of GFP expression in the comprehensive non-coding ATR model (Table S3). Furthermore, the three variables together explain two-thirds of GFP variations in those synthetic circuits (Figure 2F, right panel). The model and coefficients were also validated by another statistical software XLSTAT (version 2017.4). Based on the comprehensive non-coding model (Figure 2F), we then employed XLSTAT to predict GFP expression (mean and standard deviation) in circuits regulated with synthetic ATRs having either different GC contents or different in Figures 4C and 4D. Model predicted GFP (Figure S3C) has a similar expression trend with experimental results (Figures 4C and 4D).

We also performed k-fold cross validation to further assess our model performance (k = 10). The entire dataset was randomly partitioned into a training dataset, a validation dataset and a testing dataset. The model was built based on the training dataset (50% of the original data) and then validated on the other 25% dataset, and finally was used to assess the performance on the testing dataset (25% of the original data, Table S3). The selection method is stepwise, selection criterion is Schwarz' Bayesian Criterion (SBC) and stop criterion is Akaike's Information Criterion (AIC). We performed 10 times of the 10-fold validation, and found that the coefficients for each variable and intercept as well as R^2 are very close to the above comprehensive model (Table S3). Moreover, the standard deviation for the square root of mean squared error (RMSE) from the 10 repeats of 10-fold validation is very small (0.0064 for coding ATR, and 0.0128 for non-coding ATR), suggesting the model we built has a decent prediction accuracy and consistency.

In summary, we have demonstrated that the coding and non-coding adjacent transcription regions have remarkable effects on regulating GFP expressions in synthetic operon-based gene circuits (Figure 2). Furthermore, we can use a general biophysical model with sequence-dependent energetic changes to quantify the ATR regulation on gene expression. In this study, we mainly investigated five factors involved in ATR regulation: 5' and 3' ATRs free energies $\Delta G_{5'ATR}$ and $\Delta G_{3'ATR\_100}$, transcriptional sizes and the mRNA folding energy near the GFP starting codon. It is possible that there are some other unknown or uncharacterized factors influencing GFP expression, such as the codon degeneracy for the coding ATRs. Furthermore, there may have some special local secondary or higher structures in some ATRs, which may impact the degradation or translation of GFP.

Deterministic Model Construction and Prediction for the Logic Gate

In the four logic gates, GFP expression depends on the relative concentrations of activator (LuxR) and repressor (TetR or LacI) produced from a constitutive promoter. AHL binds with LuxR protein to activate *pLux/tet* transcription and aTc can block TetR repression to *pLux/tet*. Since the two sets of logic gates (LT/TL and LI/IL) are constructed similarly and described by the same deterministic equations, we here only explain the technical details for the gate LT. The model was built based on our previous work (Wu et al., 2014). From the biochemical reactions depicted in Figure 3A, we derived the following ordinary differential equations for intercellular concentrations of LuxR (U), TetR (R) and GFP (G):

$$\frac{dU}{dt} = (k_0 + \alpha_1) - d_1 \cdot U \qquad \text{(Equation 1)}$$

$$\frac{dR}{dt} = (k_0 + \alpha_2) - d_2 \cdot R \qquad \text{(Equation 2)}$$

$$\frac{dG}{dt} = \left(c_1 + \frac{K_1\,C}{C + K_n} \right) \cdot \frac{1}{K_r^{nt} + (R \cdot F)^{nt}} - d_3 \cdot G \qquad \text{(Equation 3)}$$

and

$$f = \frac{AHL^{ni}}{AHL^{ni} + K_i^{ni}} \qquad \text{(Equation 4)}$$

$$C = \frac{(f \cdot U)^2}{K_d} \qquad \text{(Equation 5)}$$

$$F = \frac{1}{K_t^{nr} + ATC^{nr}} \qquad \text{(Equation 6)}$$

The first two equations describe the concentrations of LuxR and TetR, both of which are driven by a constitutive promoter at a constant level (k_0). α_1 and α_2 are constants used to describe the relative changes of LuxR and TetR production, owing to the position changes in the And-gate circuit. d_1 and d_2 are the degradation rates for the LuxR and TetR protein, respectively. The third equation describes the concentration of GFP, which is determined by the relative concentrations of LuxR and TetR. LuxR binds to AHL molecules and forms the active LuxR monomers in the form of (LuxR-AHL), when the AHL concentration reaches a certain threshold (quorum-sensing mechanism). So the fraction of LuxR monomers (f) bound by AHL can be described by Equation 4, where ni is the binding cooperativity (Hill coefficient) between LuxR and AHL, and K_i represents the dissociation constant between LuxR and AHL. LuxR needs to form a dimer to bind the promoter and activate transcription, so the concentration of the functional LuxR dimer (C) that binds to the hybrid promoter $pLux/tet$ and activates its transcription can be described by Equation 5, where K_d is the is the dissociation constant for dimerization. Thus, GFP expression driven by LuxR and inducer AHL is represented by the first part of Equation 3. c_1 is the basal mRNA expression without LuxR protein; K_1 is the production rate; and K_n is the dissociation constant between C and $pLux/tet$ promoter. TetR protein can bind and inhibit GFP transcription, and the inhibition can be repressed by inducer aTc. So high GFP expression is achieved in presence of high doses of aTc, and vice versa (Equation 6). The second part of Equation 3 describes TetR inhibition to GFP expression, under induction of aTc. And the third part of Equation 3 is the degradation of GFP.

The three ordinary differential equations were used to model the two sets of AND-gate circuits: LT and TL, LI and IL. For each of the two sets, most parameters should be the same except α_1, α_2, c_1, and K_i. Based on the parameter used in our previous studies (Wu et al., 2014), we used the following parameters in our simulations: $k_0 = 1.0$, $d_1 = 0.2$, $d_2 = 0.2$, $d_3 = 0.2$, $c_1 = 0.002$ (for TL) or 0.08 (for LT), $K_1 = 1.7$, $K_n = 4.4$, $K_d = 13$, $K_t = 400$, $K_r = 3.2$, $ni = 1.2$, $nt = 2$, $ni = 1.2$, $nr = 2$. For circuits LI and IL, $c_1 = 0.002$ (for IL) or 0.05 (for LI), $K_t = 1000$, and the other parameters are the same.

From our comprehensive linear model, we calculated that LT has more LuxR than TetR production (Table S2), so the basal expression c_1 is set to a bigger value in LT model. K_i has little effect on the shape of the GFP dynamic curves, but determines the AHL concentration producing half conversion of LuxR monomers into LuxR-HSL complexes (half GFP activation). So the K_i value in the model is acquired from the experimental data. Through changing the relative expression of LuxR and TetR (i.e. α_1 and α_2), we can modulate GFP production dynamics (Figure S3E). To predict the GFP responses in circuit TL with AHL and aTc inductions, we use the parameter α_1 and α_2 in LT as a control to tune the parameter α_1 and α_2 in TL. According to the linear model calculations, the production rate for LuxR in LT and TL almost doesn't change, but production rate of TetR in TL increases by $\sim$93% (Table S2). For example, we set the production rates for LuxR and TetR in circuit LT to 1.0 ($k_0 + \alpha_1$) and 0.6 ($k_0 + \alpha_2$), respectively. So in the circuit TL, the two rates should be 1.0 ($k_0 + \alpha_1$) and 1.15 ($k_0 + \alpha_2$) based on calculations. For different doses of aTc induction, we allowed $\sim$10% parameter variations for α_1 and α_2. We found that the model simulations have a good match with our experimental data. The parameters for α_1 and α_2 in TL and LT under different doses of aTc are listed below:

Circuit	aTc (0 ng/ml)	aTc (0.2 ng/ml)	aTc (2 ng/ml)	aTc (20 ng/ml)	aTc (100 ng/ml)	aTc (200 ng/ml)
LT	$\alpha_1 = 0$ $\alpha_2 = -0.3$	$\alpha_1 = 0$ $\alpha_2 = -0.4$	$\alpha_1 = 0$ $\alpha_2 = -0.38$	$\alpha_1 = 0$ $\alpha_2 = -0.3$	$\alpha_1 = 0$ $\alpha_2 = -0.35$	$\alpha_1 = 0$ $\alpha_2 = -0.25$
TL	$\alpha_1 = 0.1$ $\alpha_2 = 0.1$	$\alpha_1 = 0.1$ $\alpha_2 = 0.05$	$\alpha_1 = 0.1$ $\alpha_2 = 0$	$\alpha_1 = 0.1$ $\alpha_2 = 0.1$	$\alpha_1 = 0.1$ $\alpha_2 = 0.1$	$\alpha_1 = 0.1$ $\alpha_2 = 0.25$

Compared to circuit LI, the production rate for LuxR in IL increases by $\sim$74%, and $\sim$38% for LacI (the overall inhibition efficiency may increase by $\sim$76%, Table S2). For example, we set the production rates for LuxR and LacI in circuit LI to 1.0 ($k_0 + \alpha_1$) and 0.8 ($k_0 + \alpha_2$), respectively. So in the circuit IL, the two rates should be 1.74 ($k_0 + \alpha_1$) and $\sim$1.41 ($k_0 + \alpha_2$) based on

calculations. For different doses of IPTG induction, we allowed ~10% parameter variations for α_1 and α_2. And the parameters for α_1 and α_2 in LI and IL under different doses of IPTG are listed below:

Circuit	IPTG (0 μM)	IPTG (0.1 μM)	IPTG (1 μM)	IPTG (10 μM)	IPTG (100 μM)	IPTG (200 μM)	IPTG (400 μM)
LI	$\alpha_1 = 0$ $\alpha_2 = -0.1$	$\alpha_1 = 0$ $\alpha_2 = -0.12$	$\alpha_1 = 0$ $\alpha_2 = -0.1$	$\alpha_1 = 0$ $\alpha_2 = -0.15$	$\alpha_1 = 0$ $\alpha_2 = -0.25$	$\alpha_1 = 0$ $\alpha_2 = -0.28$	$\alpha_1 = 0$ $\alpha_2 = -0.2$
IL	$\alpha_1 = 0.57$ $\alpha_2 = 0.57$	$\alpha_1 = 0.6$ $\alpha_2 = 0.5$	$\alpha_1 = 0.69$ $\alpha_2 = 0.5$	$\alpha_1 = 0.82$ $\alpha_2 = 0.35$	$\alpha_1 = 0.7$ $\alpha_2 = 0.4$	$\alpha_1 = 0.87$ $\alpha_2 = 0.5$	$\alpha_1 = 0.57$ $\alpha_2 = 0.56$

Bifurcation Analysis for the Synthetic Toggle Switches

For the toggle switch model in Figure 5, we used the same mathematical model and most parameters in the *Gardner et al* paper (Gardner et al., 2000). Here we think the synthetic ATRs mainly influenced the TetR production rate, with low rate corresponding to T_S28 ($\alpha_1 = 400$, $\beta = 2.7$), medium rate corresponding to T_S67 ($\alpha_1 = 600$, $\beta = 3.0$), and high rate corresponding to T_WT ($\alpha_1 = 1000$, $\beta = 3.245$). All the other parameters are set the same as in *Gardner et al* paper. Bifurcation analyses are performed using XPP-AUTO software (www.math.pitt.edu).

DATA AND SOFTWARE AVAILABILITY

Experimental data for the comprehensive biophysical model development, MATLAB code for the simulation of the logic gates, and SAS code for statistical analysis can be found in the Data S1.