

RESEARCH

Open Access



Gene syntaxes modulate gene expression and circuit behavior on plasmids

Yijie Deng^{1*}, Hannah E. Maurais¹, Kai Etheridge¹ and Rahul Sarpeshkar^{1,2*}

Abstract

Achieving consistent and predictable gene expression from plasmids remains challenging. While much attention has focused on intra-genetic elements like promoters and ribosomal binding sites, the spatial arrangement of genes within plasmids—referred to as gene syntax—also plays a crucial role in shaping gene expression dynamics. This study addresses the largely overlooked impact of gene syntaxes on gene expression variability and accuracy. Utilizing a dual-fluorescent protein system, we systematically investigated how different gene orientations and orders affect expression profiles including mean levels, relative expression ratios, and cell-to-cell variations. We found that arbitrary gene placement on a plasmid can cause significantly different expression means and ratios. Genes aligned in the same direction as a plasmid's origin of replication (Ori) typically exhibit higher expression levels; adjacent genes in the divergent orientation tend to suppress each other's expression; altering gene order without changing orientation can yield varied expression. Despite unchanged total cell-to-cell variation across different syntaxes, gene syntaxes can also influence intrinsic and extrinsic noise. Interestingly, cell-to-cell variation appears to depend on the reporter proteins, with RFP consistently showing higher variation than GFP. Moreover, the effects of gene syntax can propagate to downstream circuits, strongly affecting the performance of incoherent feedforward loops and contributing to unpredictable outcomes in genetic networks. Our findings reveal that gene syntaxes on plasmids modulate gene expression and circuit behavior, providing valuable insights for the rational design of plasmids and genetic circuits.

Keywords Plasmid design, Genetic context, Gene syntax, Genetic circuits, Gene noise, Incoherent feedforward loops

Introduction

Plasmids stand as indispensable tools in biology, facilitating a myriad of applications such as genome editing, recombinant protein production, metabolic engineering, and genetic circuits. Despite their longstanding utility after their discovery in the 1950s [1], achieving robust and predictable behaviors from designed plasmids remains a formidable challenge. Even with identical promoters and plasmid backbones, gene expression profiles can vary significantly depending on the placement and context of the gene within the plasmid. While intra-genetic elements such as promoters, ribosomal binding sites (RBS), and protein-coding sequences have traditionally received the most attention in plasmid design [2],

*Correspondence:

Yijie Deng

Yijie.deng@dartmouth.edu

Rahul Sarpeshkar

rahul.sarpeshkar@dartmouth.edu

¹Thayer School of Engineering, Dartmouth College, Hanover, NH 03755, USA

²Departments of Engineering, Microbiology & Immunology, Physics, and Molecular and Systems Biology, Dartmouth College, Hanover, NH 03755, USA



© The Author(s) 2025. **Open Access** This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

inter-genetic elements, including spatial arrangements of neighboring genes, also exert unexpected effects [2–5]. Thus, a comprehensive understanding of these mechanisms is paramount for the rational design of plasmids and genetic circuits.

Relative orientations, locations or orders of a gene, termed as gene syntaxes in this work, can significantly impact gene expression dynamics. During transcription, DNA supercoiling is generated by twisting of the DNA helix upon itself, causing under/over-wound DNA regions; it alters the binding of RNA polymerases (RNAPs) and thus the transcription of neighboring genes [3, 6]. The collisions between RNAPs can also lead to alterations in transcriptional profiles [7]. Additionally, the collisions between replication and transcription machineries, known as replication-transcription conflicts, underscore the importance of gene order and orientation relative to replication origins [8, 9]. Therefore, gene syntaxes play a crucial role in the modulation of both gene expression and DNA replication.

Achieving precise gene expression levels and accurate relative gene expression ratios is crucial in synthetic biology and its applications. For instance, metabolic engineering relies on the fine-tuning of enzyme expression levels and their relative ratios to optimize product yields [10, 11]. Gene reporter assays, which are essential for studying promoter activity, often include a reference reporter gene to account for global cellular changes induced by stimuli. The expression ratios between the reporter and reference genes provide reliable measurements of promoter activity. Therefore, improving the accuracy and precision of gene expression when designing plasmids is highly desirable.

Cell-to-cell variation (gene noise) is ubiquitous in biological systems, posing practical challenges in synthetic biology. Gene noise can be dissected into intrinsic and extrinsic noise. Intrinsic noise stems from inherent random molecular events during protein expression within each cell such as molecular binding, transcription, and translation; extrinsic noise, on the other hand, arises from cell-to-cell fluctuations in the cellular components and local environments, such as local nutrient availability and numbers of transcription and translation machineries in a cell [12]. Gene noise leads to phenotypic and functional heterogeneity, undermining the reliability and predictability of engineered biological systems [13]. For example, gene noise can cause cascading effects on the performance of downstream circuits [13, 14], compromising the functionality of synthetic genetic circuits and leading to unpredictable outcomes in applications ranging from gene reporter assays to metabolic engineering. Gene noise thus limits the precision and accuracy of many synthetic biology tools and applications [12, 15]. Therefore, understanding how gene syntaxes on plasmids

contribute to gene noise and expression dynamics is crucial for mitigating potential adverse effects and advancing the reliability of synthetic biological systems.

Despite the significance of gene syntaxes, their impacts on the accuracy and variability of gene expression on plasmids remains largely unexplored. There is a notable gap in systematic and quantitative investigations into how gene syntaxes influence gene expression profiles including the mean expression levels and the expression ratios between two neighboring genes. In this study, we systematically analyzed the influences of gene syntaxes on expression means, relative expression ratios, dynamics, and cell-to-cell variations of gene expression in plasmids. Leveraging a classic dual-fluorescent protein system [16], we examined both intrinsic and extrinsic noise across various plasmid constructs with distinct gene syntaxes. By elucidating the effects of gene syntaxes on plasmid-based gene expression profiles, our work aims to provide fundamental insights crucial for optimizing reporter gene assays and synthetic genetic circuits, thereby advancing prediction and precision in biological systems engineering.

Materials and methods

Strains, media, and growth conditions

The *E. coli* NEB 10-beta strain (cat#C3019H, New England BioLabs, Inc.) was used to construct all plasmids and tested for behaviors of reporters and circuits. This strain has been widely used in previous studies to test genetic circuits [17, 18], providing a well-established foundation for comparison. Some characteristics listed in the product description make NEB 10-beta an ideal choice for our experiments. First, NEB 10-beta can accommodate large plasmids, which is essential for testing complex genetic circuits that involve multiple genes and regulatory elements. Second, the strain includes the *recA1* mutation, which reduces homologous recombination of cloned DNA, particularly for constructs with repetitive sequences, ensuring plasmid stability during experiments. Third, its high transformation efficiency simplifies the cloning and transformation processes, enabling efficient construction and testing of plasmids. Lastly, the strain carries mutations in arabinose metabolism, making it particularly suitable for experiments involving arabinose as an inducer. This feature was leveraged to test the behavior of the incoherent feedforward loop (iFFL) circuit in our study.

A modified M63 medium was used to study gene expression, which consisted of $(\text{NH}_4)_2\text{SO}_4$ (2.0 g/L), KH_2PO_4 (13.6 g/L), FeSO_4 (0.5 mg/L), MgSO_4 (1 mM), thiamine (0.5 mg/L), yeast extract (0.3 g/L), tryptone (0.6 g/L), NaCl (0.6 g/L) with pH 7.0. To ensure cultures achieve true steady states during the log phase, a continuous culture system, Chi.Bio [19], was used to study gene

expression profiles from plasmids with different syntaxes. The circuit performance of the incoherent feedforward loop (iFFL) was evaluated in Luria-Bertani (LB) medium. This rich medium supported rapid bacterial growth and thereby reduced the time required to reach steady states for sampling and measurements. Importantly, LB medium contains no or negligible glucose that could otherwise suppress the activity of sugar-responsive promoters, including the P_{BAD} promoter used in our circuit. All cultures were grown in the Chi.Bio continuous culture system and maintained at a constant cell density at the middle log phase (OD_{600} of 0.5, $\sim 6 \times 10^7$ CFU/ml). For the iFFL experiments, the OD_{600} of cultures was kept at 0.3. Antibiotic concentrations used in the media were 50 μ g/ml for carbenicillin, 50 μ g/ml for kanamycin, and 35 μ g/ml for chloramphenicol.

Plasmid design and circuit construction

The single- and dual-reporter plasmids were designed so that all reporter genes shared identical promoter and ribosome binding site (RBS) sequences. To diminish any effect of upstream sequence on the transcription of reporter proteins, ~ 120 bp of the upstream DNA sequence (from the -35 site) were kept identical among all plasmid constructs. The potential transcription activity of the region, up to 400 bp (bp) upstream of the transcription start site (TSS), was further evaluated by using the Promoter Calculator from De Novo DNA [20]. The Promoter Calculator was developed using a statistical thermodynamic model combined with a machine learning approach. It calculates the binding free energy of RNA polymerase (RNAP)/sigma70 to a DNA sequence and quantifies the transcription initiation rate for each potential transcription start site (TSS) with high accuracy [20]. The predicted transcription rates from the Promoter Calculator were measured by arbitrary unit (a.u.) with 1000 a.u. representing a promoter activity of 0.06 RNAP/DNA/min [20]. As a cross-validation, potential cryptic promoters in upstream regions were also evaluated using the classical bacterial promoter prediction tool, BPROM [21]. Unless otherwise mentioned, in this work transcriptional read-through from upstream genes are negligible because all the reporter genes were followed by strong terminators identified by an accurate bacterial terminator prediction tool [22]. When necessary, an extra synthetic strong terminator (DT5) [23] was inserted upstream of a reporter gene to block any transcriptional read-through from upstream genes. For the pSC101 plasmid, the direction of Ori is opposite to the RepA gene [24, 25].

We used a constitutive promoter modified from BBA_J23105 from the Anderson promoter collection to drive stable protein expression. The core sequence of the constitutive promoter used is 5'-tttacgctagctcagactctatctatgctagc-3'. A relatively weak RBS sequence was used for

all reporter proteins with a sequence of 5'-acgtcgactctc-gagtgagattgtgacggtagcgtattttgcgggacatacggaggaaccta-agggtta-3', which was designed from the RBS Calculator [26]. The backbone of plasmid pUC19 with its origin of replication, ColE1, was used to make the high-copy-number plasmids while the backbone of pSC101 was used to make the low-copy-number plasmids. The super-folder green fluorescent protein gene (sfGFP) [27] and/or a similarly bright fast-fold red fluorescent protein, mScarlet-I [28], were used for reporters. The sfGFP gene was amplified via PCR from plasmid DA313 [29] and the mScarlet-I gene was synthesized as a gene block by IDT Inc. To make an iFFL circuit, a high-copy-plasmid (DA396) carrying a P_{BAD} promoter driving TetR and a hybrid promoter P_{BAD} -tetO driving deGFP as a reporter [30] was constructed and co-transformed to NEB 10-beta cells with a low-copy-plasmid carrying AraC fused to mScarlet-I. The DNA sequences and detailed maps of plasmid constructs are provided in the Supplementary materials (Figures S1 and S2, and Table S1).

All plasmids were constructed by Gibson assembly [31] using NEBuilder[®] HiFi DNA Assembly kit (NEB, Inc.). PCR reactions were conducted using high-fidelity Q5 DNA polymerase (NEB, Inc.) and the products were purified prior to Gibson assembly. Gene blocks and primers were synthesized by IDT Inc. The assembled plasmids were transformed into chemically competent *E. coli* NEB 10-beta cells (NEB, Inc.) and selected for appropriate antibiotic resistance. Plasmids, PCR products, and DNA fragments from agarose gel were purified with Qiagen miniprep, PCR purification, and Gel extraction kits, respectively. The accuracy of all plasmid sequences was confirmed through Sanger DNA sequencing and/or whole plasmid sequencing performed by Azenta Life Sciences Inc.

Flow cytometry

The GFP and RFP signals were quantified using a CytoFLEX S Flow Cytometer (Beckman Coulter, Inc.). Bacterial samples were diluted into phosphate-buffered saline (PBS) and immediately analyzed via flow cytometry. Cells were gated based on forward scatter (FSC) and side scatter (SSC) with a gain setting of 500x for both channels. An FSC-H threshold of 20,000 (a.u.) was used to eliminate events due to instrument noise. Approximately 100,000 events were collected per sample for subsequent analyses. Green fluorescence was detected using the FITC channel, while the RFP signal was captured via the PE channel. Autofluorescence from bacterial cells lacking plasmids was measured and subtracted from each sample. To ensure the accuracy and precision of the flow cytometer across experiments, AccuCheck ERF Reference Particles (Invitrogen Inc.) were employed for monitoring and calibrating fluorescence measurements as

needed. To remove events due to the potential crossover of samples and initial unevenness of sample flow rates, we collected data after letting samples run steadily for 5 s. Compensation matrices were generated from bacteria carrying single reporters using CytExpert software (Beckman Coulter Inc.), and applied uniformly to all samples under identical conditions.

Data analysis

Unless otherwise mentioned, flow cytometry data analysis was conducted in RStudio (version 4.2) [32] using packages including flowCore [33] and flowStats [34]. Raw flow cytometry data were first passed through the quality control check by FlowAI package [35] to remove electronic background noise, possible cell clumps and non-cell particles. Subsequently, the extreme outliers were removed, i.e., the top and bottom 1% of total events were filtered out in FSC x SSC plotting. Events with fluorescence below the autofluorescence were also removed. Autofluorescence was measured under the same conditions in NEB 10-beta cells without plasmids. The cleaned data (normally >90000 single-cell measurements per sample) were used to analyze the mean, expression ratio, and variation of a reporter protein in a population. Since gene expression in bacteria can be well described by a log-normal distribution [36, 37], the means of the log-fluorescence (i.e., geometric means of fluorescence) were computed as a measure of the mean expression levels of a reporter gene. Cell-to-cell variations were assessed by the standard deviations of log-fluorescence that are equivalent to the coefficient of variation (CV) [36]. For the iFFL experiments, median fluorescence values were calculated using CytExpert (V2.4.0.28, Beckman Coulter, Inc.) as measures of GFP and RFP expression; the robust coefficient of variation (rCV) was analyzed in the software and used to indicate the cell-to-cell variations of a reporter gene [38].

Calculation of intrinsic and extrinsic noise

To calculate the intrinsic and extrinsic noise of gene expression, we employed the formulas described for the dual-fluorescent protein system [16] and the flow cytometry methodology previously developed [36, 37, 39]. Briefly, a narrow population centered around the medians of FSC and SSC was first gated on the FSC-SSC panel to remove cells with different morphologies using “ellipsoidGate” in flowCore package [33]. The narrow gating strategy is critical to minimize the extrinsic variations due to physiological and morphological differences among cells including cell sizes, growth rates and cellular components. This approach enables the isolation of a more uniform cell population, ensuring more accurate noise calculations [37, 39]. This narrow gating usually resulted in ~5000–7000 uniform single cells

with fluorescence approximate to normal distributions from which the expression levels of GFP and RFP were analyzed to obtain intrinsic and extrinsic noise values of the population. To ensure accuracy, the fluorescence signals of each protein were normalized to their respective mean signals before noise calculations using the formulas developed for noise calculations [16].

Statistics

All statistical analyses were performed in GraphPad Prism (Version 10.2.3). Analysis of variance (ANOVA) tests followed by Tukey’s Honest Significant Difference (HSD) post-hoc tests were conducted to compare the mean values across different groups with $P < 0.05$ indicating significant difference.

Results and discussion

Different gene syntaxes lead to varied gene expression profiles

To systematically investigate how gene orientations and orders modulate expression profiles, we designed seven plasmid constructs with varying gene syntaxes of two reporter genes. In these constructs, the reporter genes GFP and RFP were arranged in different orders and orientations relative to the plasmid’s origin of replication (ColE1) (Fig. 1). Both GFP and RFP were driven by identical promoters and ribosome binding site (RBS) sequences across all plasmids. Strong terminators were included downstream of each reporter gene to minimize potential transcriptional interactions or read-through effects. The use of a shared constitutive promoter ensured consistency by eliminating cell-to-cell variation of transcriptional factors that might otherwise complicate the comparison.

We also ensured that the upstream (5’-end) DNA sequences of the promoter identical for ~120 bp for all constructs. This strategy simplifies cloning of repeated sequences and is sufficient to accommodate potential UP elements (~20 bp) that may influence promoter activity [40]. The RNAP holoenzyme footprint spans approximately 35 bp [41], corresponding to the size of a promoter’s core region. By using an upstream sequence of 120 bp—over three times this footprint—we avoided potential promoter overlapping and minimized RNAP-RNAP clashes arising from transcription of the upstream gene. Additionally, when combined with the intergenic region sequence, this length (typically >200 bp) ensures consistent upstream transcriptional activity across a sufficient distance. The schematic plasmid maps are shown in Fig. 1, while detailed maps with annotations and full sequences are provided in the supplementary materials (Figures S1, S2, and Table S1).

We analyzed the expression means of the two reporter genes across the seven constructs. Our results show that,

Constructs	GFP expression (a.u.)	RFP expression (a.u.)	GFP/RFP
	21076±1850 a	15389±2342 ab	1.43±0.26 a
	14835 ±2179 c	11562±1960 c	1.33±0.15 a
	11583±704 d	10768±497 c	1.11±0.06 b
	17974±1382 b	14824±1970 ab	1.26±0.10 ab
	16828±1185 bc	13016±1331 bc	1.33±0.09 a
	19154±2075 ab	13574±2752 bc	1.46±0.18 a
	17240±1824 bc	16752±1509 a	1.06±0.05 b

Fig. 1 Impact of gene syntaxes on the population means and ratios of gene expression. The arrangement and orientation of GFP and RFP genes on a ColE1 plasmid backbone were systematically varied. Both fluorescent protein genes were constructed with identical cassettes, comprising constitutive promoters, RBS sequences, and upstream DNA regions. The mean expression levels of the reporter proteins and the GFP/RFP expression ratios were analyzed across different gene arrangements. GFP and RFP fluorescence were measured in separate channels using flow cytometry; thus, the GFP/RFP ratios represent the relative expression levels of GFP and RFP in the cells rather than their absolute expression levels. To minimize variations, GFP/RFP ratios were calculated at the single-cell level before averaging the population means of GFP/RFP. Data are presented as mean ± SD, based on eight replicates. Significant differences ($P < 0.05$) were analyzed by ANOVA followed by Tukey's HSD post-hoc tests, and are indicated by different letter(s) within each column. Values that do not share a common letter are significantly different from each other. For example, values with different letters (e.g., "a" vs. "bc") are significantly different while those sharing the same letter are not significantly different (e.g., "a" vs. "ab")

depending on their gene syntaxes, GFP expression levels varied by over 1.8-fold, and RFP expression levels varied by 1.56-fold between the highest and lowest constructs. This finding indicates that gene syntaxes can significantly alter gene expression in plasmids. The expression ratios of the two reporter proteins within the same cell also changed considerably, ranging from 1.06 to 1.46 (about a 1.4-fold difference). Despite identical promoters, RBS, and upstream sequences, the arbitrary placement of genes in a plasmid can lead to considerable variations in gene expression profiles.

The expression ratio between a reporter and a reference gene is often used as a measurement of the activity of a promoter. Our finding that gene syntaxes can cause significant differences in relative gene expression ratios

indicates that gene syntaxes should be carefully considered when designing plasmids to ensure the accuracy of gene reporter assays and other synthetic-biology applications. It is important to note that the GFP/RFP ratio in our study represents the relative fluorescence signal between these two reporter proteins, indicating the relative expression levels rather than the absolute ratio of protein abundance.

Gene orientation relative to plasmid Ori affects expression profiles

We sought to identify potential mechanisms by which gene syntaxes affect gene expression on plasmids. It has been noticed that the collisions between replication and transcription machineries can influence gene expression

and DNA replication [8, 9]. It is expected that the frequent collisions between plasmid replication and gene transcription might interfere with gene expression on the plasmid. The direction of Ori determines the synthesis direction of the leading DNA strand. The continuous synthesis of the leading strand can hinder the transcription of a gene that faces against the Ori, resulting in head-on conflicts. On the other hand, if the orientation of gene expression follows the same direction as Ori, termed codirectional orientation, the conflicts might be less severe than the head-on orientation [9]. Therefore, we hypothesized that codirectional-orientated genes on a plasmid will have higher expression than the head-on orientated genes, given the same promoter, RBS, and other factors.

To test this hypothesis, we compared gene expression between a GFP gene oriented either codirectionally with or head-on against the plasmid Ori (Fig. 2A). We found that the codirectional GFP gene expressed at significantly higher levels than the head-on GFP gene (Fig. 2B). This

higher expression could be due to the transcription from cryptic promoters in the upstream intergenic region and/or transcriptional read-through from upstream coding region. To rule out these possibilities, we first used the Salis Promoter Calculator [20], an accurate promoter prediction model, to assess promoter activity within the intergenic sequence up to 400 bp before the transcription start site (TSS) of the GFP gene. Our analysis revealed that although the codirectional GFP (DA469) exhibited lower upstream transcriptional activity compared to the head-on GFP gene (DA477) (Fig. 2D), the codirectional GFP still demonstrated higher expression levels than the head-on construct (Fig. 2B). Additionally, transcriptional read-through from upstream genes was dismissed for two reasons. First, the antibiotic resistance gene, AmpR, located upstream of Ori and GFP, exhibited low promoter activity (Figure S3A); second, using a reliable terminator prediction tool, we identified at least two terminators before the GFP gene that effectively blocked any read-through upstream (Figure S3B) [23]. These results

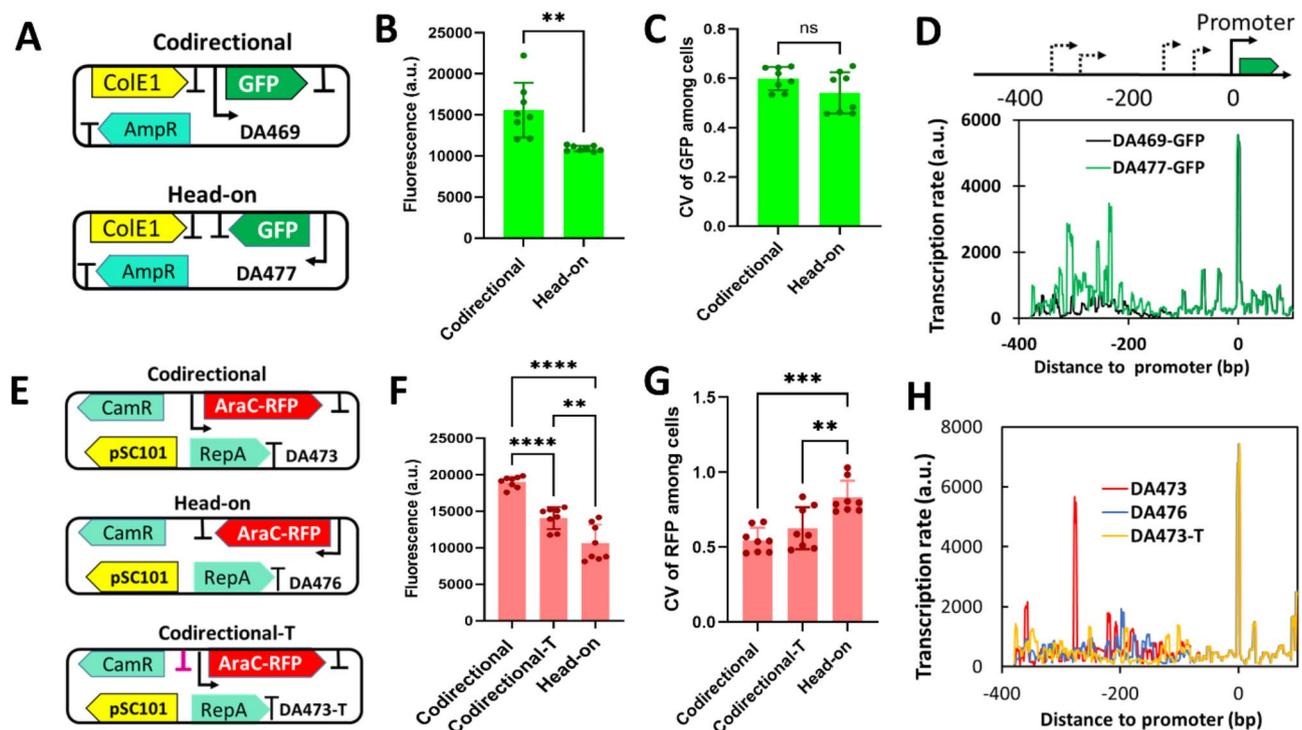


Fig. 2 Gene orientation relative to the plasmid origin of replication affects gene expression profiles. **(A)** The designs of two plasmids with opposite gene orientations of GFP on a high-copy-number plasmid with ColE1 as Ori. Codirectional orientation means the transcriptional direction of a gene (e.g., GFP) is the same as the Ori of the plasmids while the head-on gene orientation indicates the opposite transcriptional direction to the plasmid Ori. The orientation of a gene is indicated on a plasmid map as either clockwise or counterclockwise. **(B)** The mean GFP expression between codirectional and head-on gene orientation. **(C)** Cell-to-cell variation of GFP expression among the two gene orientations. **(D)** Contribution of the upstream region to the transcription of GFP gene. Distance to promoter is the number of base pairs away from the transcription start site (TSS), which is indicated by the 0 on the axis; the negative sign denotes upstream of the promoter. The high peaks indicate potential promoters. **(E)** The plasmid design of gene orientations relative to pSC101 Ori on a low-copy-number plasmid backbone. Codirectional-T means that a terminator was inserted upstream of the reporter gene (AraC-RFP fusion protein) to control for the upstream transcriptional activity. **(F)** The mean AraC-RFP expression among three constructs. **(G)** Cell-to-cell variation of gene expression among the three constructs. **(H)** Contribution of upstream DNA sequences to the transcription of AraC-RFP. Note that transcription rates in **(D, H)** were evaluated using the Promoter Calculator, a reliable computational tool for identifying promoters and predicting their activity. Values are presented as mean \pm SD ($n=8$). **** $p < 0.0001$; *** $p < 0.0001$; ** $p < 0.01$; * $p < 0.05$; ns, not significant

confirm that the higher GFP expression is largely due to the codirectional orientation of the gene rather than the effects of the upstream genetic context or transcriptional read-through.

To further validate the effect of gene orientation on gene expression, we compared codirectional and head-on gene expression in different plasmid backbones with pSC101 Ori. We designed the pSC101 plasmids carrying an AraC-RFP fusion protein as a reporter driven by a different constitutive promoter. Since the codirectional AraC-RFP gene (DA473) has higher transcriptional activity at around -280 site than the head-on reporter gene (DA476) (Fig. 2H), we constructed another plasmid (DA473-T) by inserting a strong terminator [23] upstream of the reporter gene which eliminates any upstream transcriptional activity or read-through (Fig. 2H). We then compared the expression levels among the three plasmids and found that the codirectional reporter gene (Codirectional-T) still contributed to higher expression levels (Fig. 2D), even when the upstream transcriptional activity was controlled to be similar to that of DA476 (Fig. 2H). In addition to higher expression levels, codirectional orientation seems to lower the cell-to-cell variations (CV) compared to the head-on oriented gene in the pSC101 plasmid, but no significant difference of the CV was observed in the ColE1 plasmid (Fig. 2C). Our finding that head-on orientation results in lower expression levels indicates that the collisions between DNA replication and transcription inhibit

gene expression on plasmids, which is consistent with previous studies [3, 9].

In addition to studying the effects of gene orientation on the expression of single-gene reporters, we extended our investigation to more complex plasmid constructs containing dual-reporter genes. For each pair of plasmid constructs, the GFP or RFP gene was placed at the same location but in opposite directions on the parent plasmids (Fig. 3A and D). Transcriptional read-through was diminished by the presence of strong terminators after each reporter gene (Figure S1). Similar to the findings of single reporter genes, the results show that codirectional genes usually express more proteins than head-on genes. Codirectional GFP had higher expression levels than its head-on counterparts for both pairs (Fig. 3A and D), even when the upstream transcription activity of codirectional GFP is lower (Fig. 3A and B) or similar to the head-on GFP (Fig. 3E). Codirectional RFP also expressed more than the head-on RFP in the DA464 and DA463 pair. However, it is interesting to note that while comparison of the GFP pair (Fig. 3A) clearly showed significantly higher expression in the codirectional orientation compared to the head-on orientation, the RFP pair placed at the same location as GFP (Fig. 3D) only exhibited a slight increase in codirectional expression, which was not statistically significant. This discrepancy may arise from several factors. First, when the RFP gene was switched in orientation on DA465, its upstream region showed higher transcriptional activity from potential cryptic

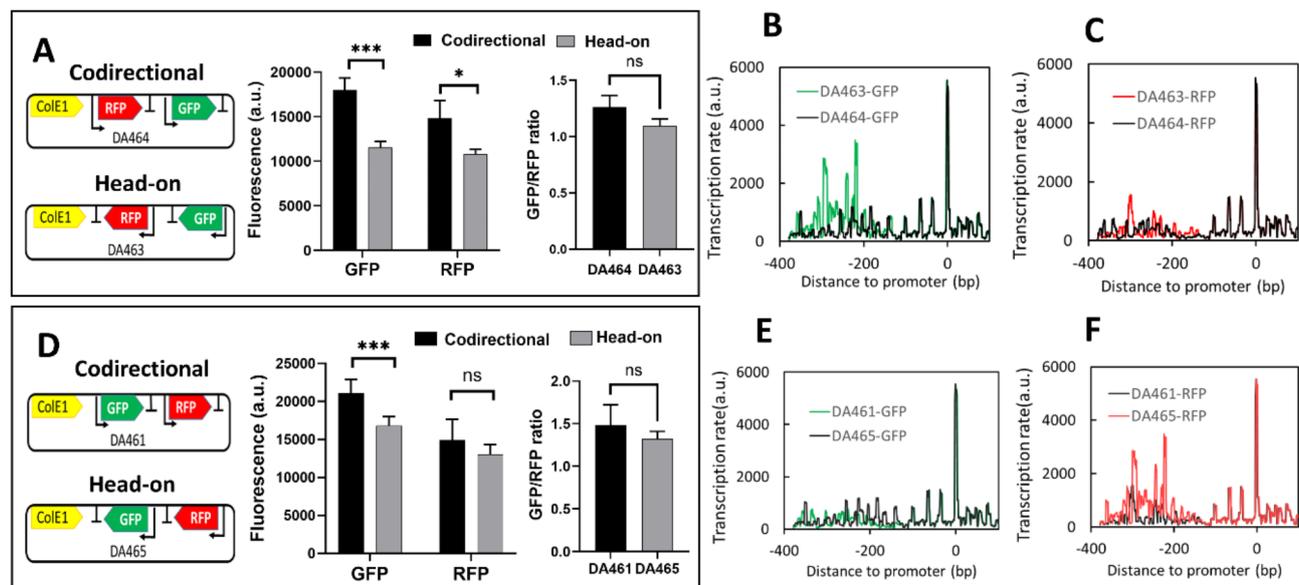


Fig. 3 Gene orientation relative to the plasmid Ori affects gene expression profiles. **(A, D)** In each plasmid pair, two reporter genes, placed at the same location on the corresponding plasmid, face the same direction either both codirectional to or both opposite to the plasmid Ori (ColE1). The mean levels and ratios of gene expression were compared. **(B, C)** Contribution of upstream DNA sequences to the transcription of GFP **(B)** or RFP **(C)** on plasmids DA464 and DA463, respectively. **(E, F)** Contribution of upstream DNA sequences to the transcription of GFP **(E)** and RFP **(F)** on plasmids DA461 and DA465. Note that transcription rates were evaluated using the Promoter Calculator, a reliable computational tool for identifying promoters and predicting their activity. Values are presented as mean \pm SD ($n=8$). *** $p < 0.0001$; ** $p < 0.01$; * $p < 0.05$; ns, not significant

promoters (indicated by more red peaks) (Fig. 3F). This increased activity may offset the inhibitory effect of head-on transcription-replication collisions. Additionally, factors such as experimental variability, statistical power, and protein-specific characteristics may influence expression. For example, RFP consistently exhibited higher variation (CV) than GFP (as discussed later), which reduced statistical power when comparing RFP expression levels. Therefore, the increased RFP expression in the codirectional orientation did not reach statistical significance. Interestingly, the consistent observation of enhanced expression for both GFP (Fig. 3A) and RFP (Fig. 3D) in the codirectional orientation at the same plasmid location suggests that the effect of gene orientation may be more pronounced when genes are placed immediately adjacent to the Ori.

Although gene orientation results in differential gene expression, the expression ratios of GFP to RFP appear independent of gene orientations as long as the two reporter genes face the same direction. The ratios of GFP/RFP remained largely unchanged between the codirectional and head-on orientation in each plasmid pair (Fig. 3A and D). The results are reasonable because as long as both reporter genes face the same direction, both of their expression levels either increase or decrease in a similar scale, depending on the relative orientation to the Ori of a plasmid, thus resulting in similar ratios.

To confirm that gene orientation also influences gene expression in other plasmid backbones, we built two plasmids where GFP and RFP were either codirectional or opposite to the direction of pSC101 Ori of a low-copy-number plasmid (Figure S4). Again, we found that genes codirectional to Ori expressed more than genes facing against Ori (Figure S4B). Since the upstream transcriptional activity is similar between the plasmid constructs (Figure S4D, E) and no apparent transcriptional read-through occurred due to the presence of weak promoter and efficient terminators (Figures S2 and S3), the higher expression levels can primarily be attributed to the codirectional orientation of the genes.

It should be noted that we utilized Salis's Promoter Calculator to predict upstream transcription over a 400-bp region, encompassing the entire intergenic region and, if present, a partial upstream coding region. This robust and accurate computational tool has been extensively validated in prior studies [42–45]. Its high reliability has been particularly demonstrated with over 5,000 designed promoters, utilizing precise transcription rate measurements in controlled *in vitro* transcription reactions [20]. In our study, we experimentally confirmed the tool's accuracy (Fig. 2E, F, G, H). For example, DA473 exhibited higher AraC-RFP expression and upstream transcriptional activity compared to DA476. When a strong terminator (DT5) [23] was inserted in DA473T, the

upstream transcriptional activity was effectively blocked; as predicted by Promoter Calculator, this resulted in lower transcription activity levels, comparable to those of DA476 (Fig. 2H), and a corresponding reduction in reporter expression. To further validate these predictions, we employed BPROM [21], an independent bacterial promoter prediction tool, which yielded similar patterns of predicted promoter scores (Figures S5–S7), providing cross-validation for the Promoter Calculator's results. The high similarity of upstream DNA sequences in our constructs yielded consistent transcription patterns, further minimizing prediction errors and reinforcing the robustness of our conclusions. Additionally, we observed consistent results across different plasmid backbones (Figs. 2 and 3, Figure S4), supporting our conclusion that gene orientation aligned with the plasmid origin of replication (Ori) enhances expression compared to genes positioned opposite the Ori. Overall, these results confirm the reliability of the Promoter Calculator in predicting upstream transcriptional activity and support its application in our study.

It is currently unclear how the arrangements of antibiotic resistance genes affect the expression of reporter genes, though their impact is expected to be small in our study. First, the expression of resistance genes is generally low, as their primary role is to confer resistance rather than produce high levels of protein. For example, the promoter activity of resistance genes such as AmpR, CamR or KanR in the backbones of our constructs is much weaker (Figure S3) than that of our reporter genes. Second, when a resistance gene is positioned upstream of a reporter gene, efficient terminators are included between the two (refer to plasmid maps in Figures S1 and S2). These terminators effectively minimize transcriptional read-through into downstream genes, further reducing the potential impact of resistance markers on reporter gene expression. The native plasmid backbones used in our study already incorporate these terminators as part of their design. By building our plasmids and circuits on these commonly used backbones, we ensured consistency across constructs, providing robust controls for comparing differences within groups.

It is worth noting that the CamR gene is positioned between the reporter gene (AraC-RFP) and the pSC101 origin (Fig. 2E). However, due to the reasons outlined above, its effect on reporter gene expression appears negligible. As further validation, we tested dual gene reporters in another pSC101 backbone where KanR, along with its promoter and terminator, is positioned between sfGFP and the pSC101 origin (Figure S4). Consistent with our findings, genes aligned with the Ori exhibited higher expression levels than those oriented oppositely.

While exploring the potential syntax effect of resistance genes is an interesting avenue for future work, we believe

its impact in our current experiments is minimal. Our study specifically aimed to systematically analyze plasmid gene syntaxes and their influence on gene expression, noise, and circuit behaviors using the commonly adopted standard native backbones.

Relative orientations of two neighboring genes affect expression levels

We next investigated whether the relative orientations of two neighboring genes affect gene expression. There are three possible orientations for two adjacent genes: convergent, divergent, and tandem (Fig. 4A). To examine their effects on gene expression profiles, we constructed plasmids for each orientation. Our results show that the tandem orientation exhibited the highest expression levels for both GFP and RFP, while the divergent orientation generally resulted in the lowest expression levels for both genes (Fig. 4B). This finding aligns with previous reports indicating that divergent orientation inhibited gene expression [3]. Interestingly, while relative orientations influenced expression levels, it did not affect the GFP/RFP ratios (Fig. 4C).

DNA supercoiling is an important factor mediating the effects of gene context on expression [3–5]. During transcription, RNAPs generate DNA supercoiling by twisting the DNA double helix, creating both underwound and overwound regions [46, 47]. Particularly, when the upstream of a promoter is bound by a DNA-binding protein, such as a transcriptional repressor, it impedes the free movement of supercoils, leading to the accumulation of DNA supercoiling near the promoter [3–5]. This supercoiled promoter region interferes with RNAP binding, thereby reducing transcription [3–5]. Prior studies have indicated that DNA supercoiling plays a dominant role in mediating the inhibitory effects of divergent-oriented genes in the presence of transcriptional repressors. For instance, experiments

using topoisomerases like TopoI and gyrase, which relax supercoiling, nearly abolished these inhibitory effects [3, 4]. Our results imply that, even in the absence of transcriptional factors, DNA supercoiling may still interfere with transcription and potentially suppress the expression of divergent-oriented genes (Fig. 4). With two adjacent promoters in a divergent arrangement, transcription of one gene by RNAPs generates DNA supercoiling that represses the transcription initiation of the other. This reciprocal interference likely accounts for the observed lower expression levels in the divergent orientation.

While DNA supercoiling appears to be an important mechanism, other factors may also contribute to the different expression observed among the three orientations. These include orientation relative to the plasmid's Ori, cryptic transcriptional activity of intergenic DNA regions, and transcriptional read-through. For GFP, the lower expression in the divergent orientation may also be partly due to head-on collisions between transcription and plasmid replication as discussed earlier. However, RFP in the divergent orientation, in the absence of head-on collisions, did not result in higher expression compared to the other orientations, thereby excluding head-on collision as a major factor. Cryptic transcriptional activity in intergenic regions was similar for GFP among the three orientations (Figure S3A), suggesting it does not significantly contribute to the observed differences. Similarly, the intergenic transcription activity for divergent-oriented RFP was comparable to the tandem orientation (Figure S3B), further ruling out cryptic activity as a main factor and suggesting the DNA-supercoiling effect. Transcriptional read-through is unlikely to explain the results for GFP, given the presence of efficient upstream terminators (e.g., for tandem-oriented GFP) and weak upstream promoter activity such as that of the AmpR gene (Figure S1). Likewise, for RFP, terminators are present upstream of the tandem-oriented gene, ruling

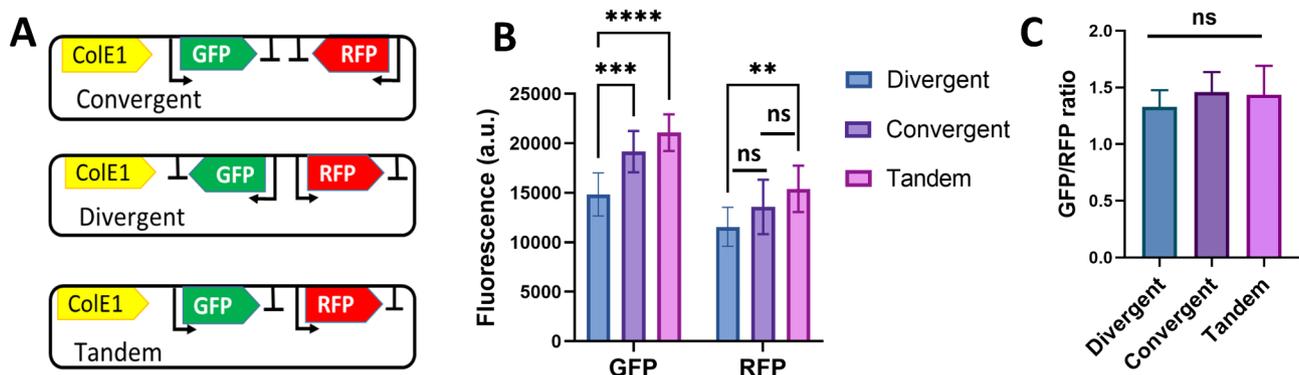


Fig. 4 Gene expression profiles of the convergent, divergent, and tandem orientations. **(A)** The designs of plasmids carrying two adjacent reporter genes with either convergent, divergent, or tandem orientation. **(B)** The mean expression levels of GFP and RFP were compared among the three relative orientations. **(C)** The expression ratios of GFP and RFP were compared among the three relative orientations. Values are presented as mean \pm SD ($n = 8$). **** $p < 0.0001$; *** $p < 0.0001$; ** $p < 0.01$; ns, not significant

out read-through as a significant factor for the higher expression in this orientation. Therefore, DNA supercoiling-mediated transcriptional interference likely remains the dominant explanation for reduced expression in the divergent orientation, although direct experimental validation will be necessary in future studies to confirm this conclusion.

The relative orientation between a reporter gene and its adjacent antibiotic resistance gene (AmpR) on the plasmid is another potential factor. However, given the weak transcription of AmpR in the plasmid used (Figure S4A), transcriptional interference between AmpR and the reporter genes is likely minimal compared to interference between the two reporter genes.

Collectively, our findings demonstrate that placing two neighboring genes in divergent orientation inhibits gene expression on plasmids. While identifying the exact mechanism by using tools such as topoisomerases and single-molecule techniques [48] is beyond the scope of this study, our results suggest a practical design guideline for plasmid construction: avoiding divergent placement of neighboring genes can minimize transcriptional interference. This could be a valuable rule for optimizing plasmid design and improving the predictability of gene expression.

Gene order can affect gene expression on plasmids

The order or location of genes within a genome has been shown to influence gene expression significantly [49–51]. In molecular cloning, determining the placement of a gene is a critical consideration during plasmid design. To explore whether gene order (or location) on a plasmid affects gene expression profiles, we compared three pairs of plasmid constructs with GFP and RFP placed in different orders (Fig. 5). In each pair, the reporter genes were oriented in the same direction. We found that the order of the GFP gene on the plasmid altered its expression in all three pairs of plasmids tested. Additionally, two of the three pairs showed differential expression of RFP depending on its order on the plasmid. Several mechanisms could contribute to these results. First, changing gene orders may alter the relative orientations of adjacent genes. For example, genes that were originally divergent may become convergent when the gene order is changed, leading to variations in gene expression (Fig. 5A), as we observed in the previous section. Second, changing the gene order may create a different genetic context, including variations in intergenic regions, which could influence transcriptional activity and subsequently alter gene expression (Figure S9). While these two mechanisms are likely contributors, other factors may also interact to influence gene expression. Disentangling all potential factors was beyond the scope of this study. We also noticed that varying gene order could change GFP/RFP

ratios, as two of three pairs showed differences in GFP/RFP ratios (Fig. 5). Our results clearly demonstrate that gene order or location significantly affects gene expression on plasmids, an important factor to consider when designing plasmids that carry multiple genes such as operons. This finding highlights the importance of keeping gene order consistent to increase the predictability of genetic designs.

Effect of gene syntaxes on the variability of gene expression

Cell-to-cell variation (gene noise) is common in biological systems and poses significant challenges in synthetic biology by contributing to phenotypic and functional heterogeneity, which undermines the reliability of engineered systems [13, 14]. We therefore examined the effect of gene syntaxes on cell-to-cell variation in a bacterial population. The coefficient of variation (CV) was used as a measure of cell-to-cell variation and was calculated from the standard deviation of the logarithmic transformed data, which is equivalent to CV. We did not find significant variation in GFP expression or RFP expression across seven plasmids with different gene syntaxes (Fig. 6A, B). However, we found that the variations of GFP are consistently lower than the variations of RFP, regardless of different gene syntaxes.

These results suggest that different proteins can exhibit different expression variability within a population, even when driven by identical promoters and RBS sequences on the same plasmid. Notably, RFP consistently shows greater variability than GFP, although the reasons for this remain unclear. According to FPbase (<https://www.fpbase.org/>), sfGFP and mScarlet-I have comparable brightness values (54.15 for sfGFP and 56.16 for mScarlet-I), indicating that the observed differences in signal intensities are primarily due to instrument settings and/or the characteristics of the RFP detection channel on the flow cytometer used rather than the differences in protein brightness. Although higher RFP variability could partially stem from lower RFP signals compared to GFP, it likely also arises from biological factors such as protein folding dynamics, sensitivity to cellular pH, and local oxygen concentration, which may be protein-specific features. Constructs like DA463 and DA474, which showed comparable GFP and RFP signals (GFP/RFP close to 1.0, Fig. 1), still demonstrated consistently higher CV for RFP than GFP (Fig. 6C). These observations indicate that the elevated variability in RFP expression is primarily due to inherent biological differences between the two fluorescent proteins. For example, RFP may be more sensitive to heterogeneity in cellular conditions, such as intracellular oxygen fluctuations or pH variations within the bacterial population. Additionally, differences in protein folding dynamics, including the distinct maturation times

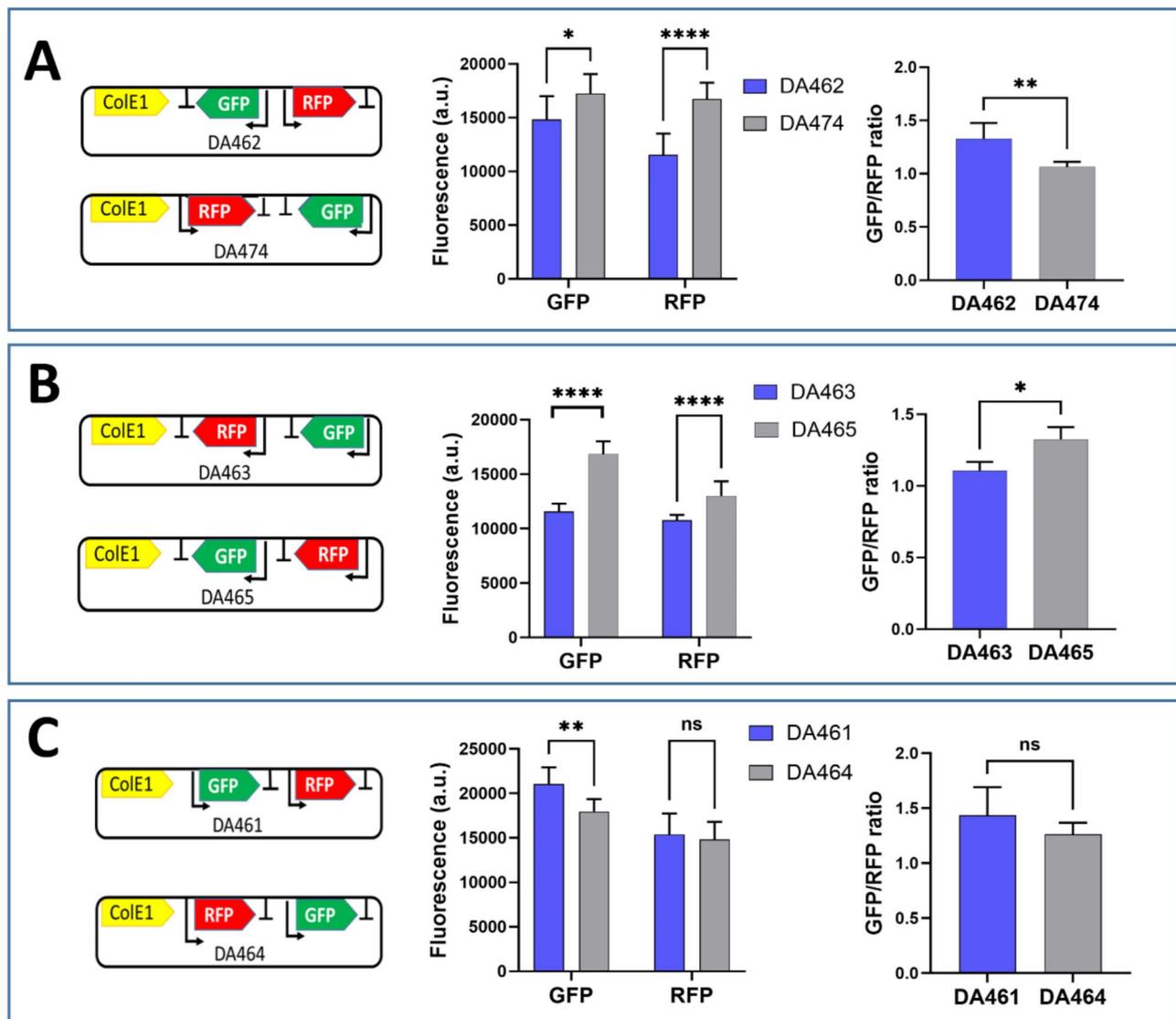


Fig. 5 Effect of gene order on gene expression profiles. (A, B, C) Three pairs of plasmids carrying GFP and RFP with the same promoter and RBS sequences. The order of the two reporter genes is switched on the same plasmid for each pair while their orientations to plasmid Ori remain the same. The expression means and expression ratios were compared within each pair of plasmids. Values are presented as mean \pm SD ($n=8$). **** $p < 0.0001$; *** $p < 0.0001$; ** $p < 0.01$; * $p < 0.05$; ns, not significant

between sfGFP and RFP (mScarlet-I) [52], could also contribute to the observed cell-to-cell variations.

Gene noise encompasses intrinsic noise (η_{int}^2) and extrinsic noise (η_{ext}^2). Taking advantage of the dual-reporter system [16] and flow cytometry methodology developed previously [36, 37, 39], we dissected the total noise into intrinsic and extrinsic noise from plasmids with varied gene syntaxes. Our results indicate that all plasmid constructs exhibited relatively low noise levels, irrespective of their gene syntaxes. Intrinsic noise was consistently below 0.1, while extrinsic noise remained under 0.17 (close to the extrinsic noise limit) [53–55]. Notably, genes arranged in different orientations and orders displayed significant variations in both intrinsic

and extrinsic noise (Fig. 6D, E). Despite these differences, the total noise remained largely unchanged across the seven gene syntaxes tested, with the exception of DA462, which exhibited slightly higher total noise compared to DA474.

While it may be anticipated that head-on orientation (e.g., DA465) (Fig. 1) would result in higher intrinsic noise compared to codirectional constructs (e.g., DA464), our data showed a slightly higher noise level in the codirectional construct. This counterintuitive result likely arises from several factors. First, the intrinsic noise levels calculated in our study were small (< 0.1), making the observed differences marginal and potentially influenced by experimental variability. Second, variations in

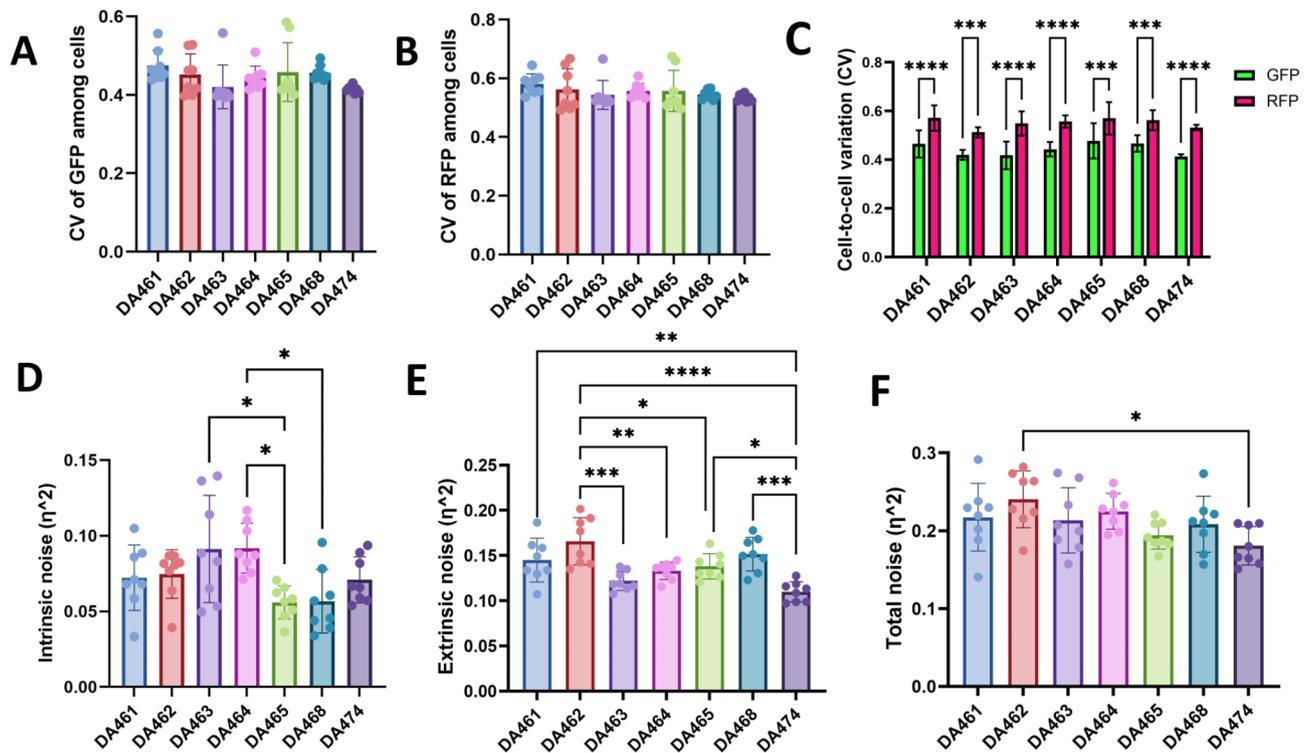


Fig. 6 Cell-to-cell variations of gene expression among varied gene syntaxes. (A) Cell-to-cell variations of GFP expression among varied gene syntaxes in bacterial cells. The coefficient of variation (CV) was used as a measure of cell-to-cell variations among a bacterial population. The seven plasmid constructs with different gene syntaxes are the same as in Fig. 1. (B) Cell-to-cell variations of RFP expression among varied gene syntaxes in bacterial cells. (C) Pairwise comparison of gene noise between GFP and RFP from the same plasmid. (D) The comparison of intrinsic gene noise across different gene syntaxes. (E) The comparison of extrinsic gene noise across different gene syntaxes. (F) The comparison of total gene expression noise across different gene syntaxes. **** $p < 0.0001$; *** $p < 0.0001$; ** $p < 0.01$; * $p < 0.05$

factors such as promoter and ribosome binding site (RBS) strength could significantly influence noise intensity [56, 57], thereby affecting the sensitivity of noise detection. Future investigations with systematically varied promoter and RBS strengths are needed to resolve these discrepancies and provide a clearer understanding of the relationship between gene orientation and noise.

The impact of gene syntaxes on intrinsic or extrinsic noise has yet to be fully understood. It is plausible that the altered transcription of reporter genes, driven by variations in gene syntaxes, contributes to the observed changes in intrinsic noise, as transcription is recognized as a major source of this type of noise [58]. Furthermore, the different expression levels resulting from various gene syntaxes could lead to fluctuations in cellular components and plasmid copy numbers within host cells, thereby generating differential extrinsic noise among different plasmids [58, 59].

Nonetheless, numerous factors could interplay, adding layers of complexity to these explanations. It is worth noting that the genetic background of a host strain can affect gene noise, such as mRNA stability, availability of free RNAPs and sigma factors, or protein synthesis efficiency [56, 57, 60]. For instance, mutations in certain

genes such as *recA*, which is responsible for rescuing stalled replication forks, have been shown to increase expression noise, likely due to variations in DNA copy number across different chromosomal regions [16]. Similarly, genetic changes that affect the distribution of plasmid copy number could further amplify noise. Our constructs provide a valuable framework for investigating gene expression noise across different strains and for identifying genes that may contribute to expression variability in future studies.

Effect of gene syntaxes on the behavior of incoherent feedforward loops

Changes in gene expression can propagate through genetic networks, potentially altering downstream circuit performance. These cascading effects of gene syntaxes on the functionality of genetic circuits are a significant concern in synthetic biology. We explored the impact of gene syntaxes on the behavior of genetic circuits by designing an incoherent feedforward loop (iFFL) as a concrete example. In the iFFL circuit, the transcriptional factor AraC-RFP, constitutively expressed from a low-copy-number plasmid, activates the expression of both GFP and TetR genes on a high-copy-number plasmid; TetR

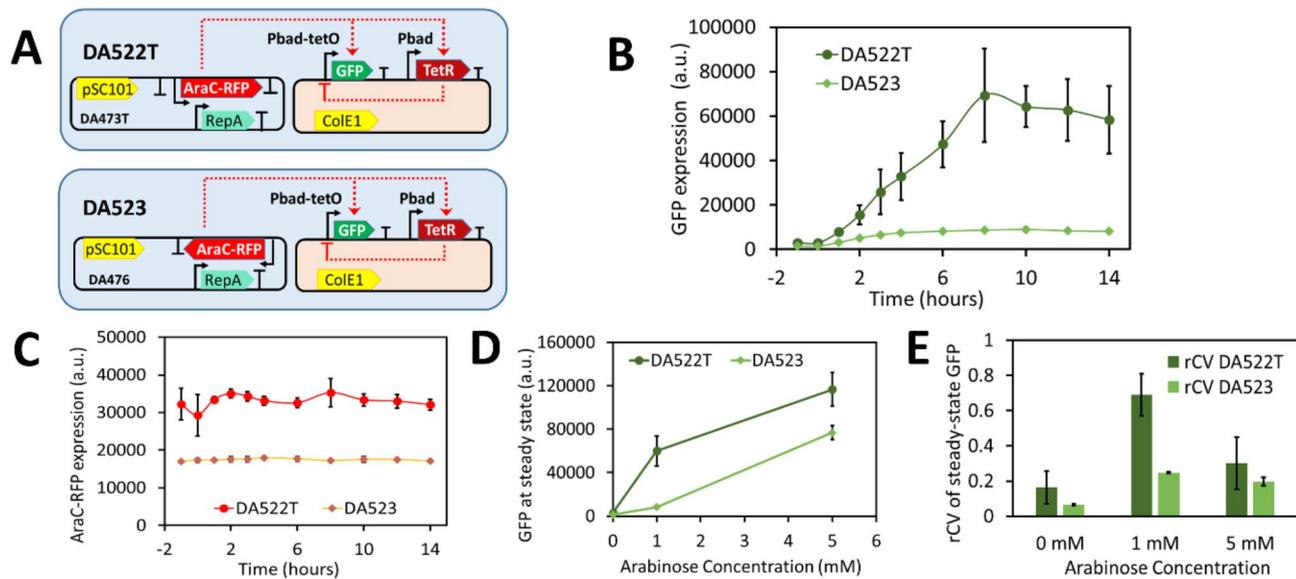


Fig. 7 The cascading effect of gene syntaxes on the performance of incoherent feedforward loops. **(A)** Two versions of iFFL differ in the orientation of AraC-RFP on the pSC101 plasmid. **(B)** Dynamics of GFP expression upon the addition of arabinose in a continuous culture system. Anhydrotetracycline (aTc) was added at a constant concentration of 10 ng/ml for both circuits in the beginning of experiments. Arabinose was supplemented to a final concentration of 1 mM after the culture (OD_{600}) reached a steady state. The GFP expression is measured by the median fluorescence intensity (MFI) of GFP. **(C)** The comparison of AraC-RFP expression between two versions of the iFFL. **(D)** The steady-state GFP expression upon the induction by arabinose. **(E)** The cell-to-cell variations of GFP expression at different arabinose concentrations, which were measured by the robust coefficient of variation (rCV)

then represses GFP expression (Fig. 7A, B), creating a characteristic bump curve for GFP expression. We built a pair of iFFL circuits (DA522T and DA523) using DA473-T and DA476 (Fig. 2E) that differ in the orientation of AraC-RFP on the pSC101 plasmid. In one circuit (DA522T), AraC-RFP is codirectional to the pSC101 Ori, while in the other (DA523), it is head-on. DA473-T was selected over DA473 because it includes a terminator upstream of AraC-RFP, reducing the influence of upstream transcriptional activity observed previously (Fig. 2H). The comparable upstream transcriptional activity of DA473-T and DA476 enabled us to isolate and specifically examine the effect of gene orientation on iFFL circuit behavior.

Our results demonstrate that gene orientation relative to the plasmid Ori can significantly alter the behavior of iFFL circuits (Fig. 7). In the DA522T circuit, where the gene is codirectional with the Ori, GFP expression rapidly increased upon the addition of the inducer arabinose, peaking around 8 h post-induction before gradually declining to a steady state after approximately 14 h. In contrast, the DA523 circuit, with a head-on orientation, exhibited a much lower peak in GFP expression before reaching a steady state (Fig. 7B). The different behaviors between the two iFFL circuits might be attributed to the varied expression of the transcriptional factor AraC-RFP (Fig. 7C), which is influenced by codirectional and head-on orientations, consistent with our earlier findings (Fig. 2E, F, G). These variations in AraC-RFP expression

contributed to downstream effects, including distinct GFP dynamics (Fig. 7B) and circuit responses to arabinose, as evidenced by differences in steady-state GFP levels and cell-to-cell variability (Fig. 7D and E). Our findings indicate that the effects of gene orientation can propagate through downstream circuits, critically influencing the performance of iFFL circuits, and highlight that gene syntaxes can contribute to unpredictable outcomes in genetic networks.

Conclusion

Our work demonstrates that different gene orientations and orders significantly impact expression means, relative expression ratios, and cell-to-cell variations. Although this study did not explore the detailed molecular mechanisms behind the observed effects of gene syntaxes—an area for future research—this study highlights the importance of gene syntaxes as a key factor in plasmid design. Genes aligned in the same direction as the plasmid origin (Ori) typically exhibit higher expression levels. The relative orientations of adjacent genes also modulate gene expression; for instance, divergent orientation can suppress the expression of both neighboring genes likely due to DNA supercoiling generated during transcription. Furthermore, altering gene orders on a plasmid without changing their transcriptional orientations to plasmid Ori can lead to varied expression profiles, presumptively caused by changes in upstream DNA context and relative gene orientations. Gene syntaxes also influence intrinsic

and extrinsic gene noise, despite unchanged total cell-to-cell variations across different syntaxes. Notably, cell-to-cell variation appears to depend on the reporter protein, with RFP consistently exhibiting higher variation than GFP. Ultimately, gene syntaxes can affect the behavior of genetic circuits, such as incoherent feedforward loops (iFFLs), leading to different dynamics and cell-to-cell variations. These findings highlight the importance of considering gene arrangement in plasmid design. Our results provide insights for the rational design of plasmids and engineering of genetic circuits to achieve enhanced reliability and precision of synthetic biological systems.

Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s13036-025-00493-0>.

Supplementary Material 1

Acknowledgements

No applicable.

Author contributions

Y.D. and R.S. conceived the project. Y.D. designed the experiments. H.E.M., Y.D., and K.E. performed the experiments. Y.D. and H.E.M. analyzed the data. All authors edited, reviewed and approved the paper.

Funding

This work was supported by funding from NSF Award 2240264 (CCF Division) and grants from Dartmouth College.

Data availability

All data used in this study are available in the main manuscript and the Supplementary material files; raw data are available upon request.

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

All authors have read and approved the submission of the manuscript.

Competing interests

The authors declare no competing interests.

Received: 12 November 2024 / Accepted: 3 March 2025

Published online: 27 March 2025

References

- Helinski DR. A brief history of plasmids. *EcoSal Plus*. 2022;10:eESP-0028-202. <https://doi.org/10.1128/ecosalplus.ESP-0028-2021%0A>
- Cardinale S, Arkin AP. Contextualizing context for synthetic biology - identifying causes of failure of synthetic biological systems. *Biotechnol J*. 2012;7:856–66.
- Yeung E, Dy AJ, Martin KB, Ng AH, Del Vecchio D, Beck JL et al. Biophysical constraints arising from compositional context in synthetic gene networks. *Cell Syst*. 2017;5:11–24.e12. <https://doi.org/10.1016/j.cels.2017.06.001>
- Boulas I, Bruno L, Rimsky S, Espeli O, Junier I, Rivoire O. Assessing in vivo the impact of gene context on transcription through DNA supercoiling. *Nucleic Acids Res*. 2023;51:9509–21. <https://academic.oup.com/nar/article/51/18/9509/7257932>
- Johnstone CP, Galloway KE. Supercoiling-mediated feedback rapidly couples and tunes transcription. *Cell Rep*. 2022;41:111492. <https://doi.org/10.1016/j.celrep.2022.111492>
- Stone A, Youssef A, Rijal S, Zhang R, Tian X-J. Context-dependent redesign of robust synthetic gene circuits. *Trends Biotechnol*. 2024;42:895–909. <https://linkinghub.elsevier.com/retrieve/pii/S0167779924000039>
- Wang L, Watters JW, Ju X, Lu G, Liu S. Head-on and co-directional RNA polymerase collisions orchestrate bidirectional transcription termination. *Mol Cell*. 2023;83:1153–1164.e4. <https://doi.org/10.1016/j.molcel.2023.02.017>
- Mirkin EV, Mirkin SM. Mechanisms of transcription-replication collisions in bacteria. *Mol Cell Biol*. 2005;25:888–95. <https://www.tandfonline.com/doi/full/10.1128/MCB.25.3.888-895.2005>
- Lang KS, Merrikh H. The clash of macromolecular titans: Replication-transcription conflicts in bacteria. *Annu Rev Microbiol*. 2018;72:71–88. <https://www.annualreviews.org/doi/10.1146/annurev-micro-090817-062514>
- Lim HG, Noh MH, Jeong JH, Park S, Jung GY. Optimum rebalancing of the 3-Hydroxypropionic acid production pathway from glycerol in *Escherichia coli*. *ACS Synth Biol*. 2016;5:1247–55. <https://doi.org/10.1021/acssynbio.5b00303>
- Jung J, Lim JH, Kim SY, Im D-K, Seok JY, Lee S-JV et al. Precise precursor rebalancing for isoprenoids production by fine control of *gapA* expression in *Escherichia coli*. *Metab Eng*. 2016;38:401–8. <https://linkinghub.elsevier.com/retrieve/pii/S1096717616301586>
- Raj A, van Oudenaarden A. Nature. Nurture, or chance: stochastic gene expression and its consequences. *Cell*. 2008;135:216–26.
- Samoilov M, Plyasunov S, Arkin AP. Stochastic amplification and signaling in enzymatic futile cycles through noise-induced bistability with oscillations. *Proc Natl Acad Sci*. 2005;102:2310–5. <https://pnas.org/doi/full/10.1073/pnas.0406841102>
- Oyarzún DA, Lugagne JB, Stan GBV. Noise propagation in synthetic gene circuits for metabolic control. *ACS Synth Biol*. 2015;4:116–25.
- Bandiera L, Furini S, Giordano E. Phenotypic variability in synthetic biology applications: dealing with noise in microbial gene expression. *Front Microbiol*. 2016;7:1–7.
- Elowitz MB, Levine AJ, Siggia ED, Swain PS. Stochastic gene expression in a single cell. *Science*. 2002;297:1183–6. <https://www.science.org/doi/10.1126/science.1070919>
- Andrews LB, Nielsen AAK, Voigt CA. Cellular checkpoint control using programmable sequential logic. *Science*. 2018;361. <https://doi.org/10.1126/science.aap8987>
- Zeng J, Banerjee A, Kim J, Deng Y, Chapman TW, Daniel R et al. A novel bioelectronic reporter system in living cells tested with a synthetic biological comparator. *Sci Rep*. 2019;9:7275. <http://www.nature.com/articles/s41598-019-43771-w>
- Steel H, Habgood R, Kelly CL, Papachristodoulou A. In situ characterisation and manipulation of biological systems with Chi.Bio. Baden T, editor. *PLOS Biol*. 2020;18:e3000794. Available from: <https://doi.org/10.1371/journal.pbio.1000794>
- LaFleur TL, Hossain A, Salis HM. Automated model-predictive design of synthetic promoters to control transcriptional profiles in bacteria. *Nat Commun*. 2022;13:5159. <https://www.nature.com/articles/s41467-022-32829-5>
- Solovyeve VSA. Automatic annotation of microbial genomes and metagenomic sequences. In: Roboert W, Li, editors. *Metagenomics and its applications in agriculture, biomedicine and environmental studies*. Hauppauge: Nova Science Publishers; 2011. pp. 61–78.
- Feng C-Q, Zhang Z-Y, Zhu X-J, Lin Y, Chen W, Tang H et al. iTerm-PseKNC: a sequence-based tool for predicting bacterial transcriptional terminators. Hancock J, editor. *Bioinformatics*. 2019;35:1469–77. <https://doi.org/10.1093/bioinformatics/bty827>
- Park Y, Borujeni AE, Gorochowski TE, Shin J, Voigt CA. Precision design of stable genetic circuits carried in highly-insulated *E. coli* genomic landing pads. *Mol Syst Biol*. 2020;16. <https://www.embopress.org/doi/full/https://doi.org/10.15252/msb.20209584>
- Cabello F, Timmis K, Cohen SN. Replication control in a composite plasmid constructed by in vitro linkage of two distinct replicons. *Nature*. 1976;259:285–90. <https://www.nature.com/articles/259285a0>
- Manen D, Caro L. The replication of plasmid pSC101. *Mol Microbiol*. 1991;5:233–7. <https://onlinelibrary.wiley.com/doi/10.1111/j.1365-2958.1991.tb02103.x>

26. Espah Borujeni A, Channarasappa AS, Salis HM. Translation rate is controlled by coupled trade-offs between site accessibility, selective RNA unfolding and sliding at upstream standby sites. *Nucleic Acids Res.* 2014;42:2646–59. <https://doi.org/10.1093/nar/gkt1139>
27. Pédelacq J-D, Cabantous S, Tran T, Terwilliger TC, Waldo GS. Engineering and characterization of a superfolder green fluorescent protein. *Nat Biotechnol.* 2006;24:79–88. <https://www.nature.com/articles/nbt1172>
28. Bindels DS, Haarbosch L, Van Weeren L, Postma M, Wiese KE, Mastop M, et al. mScarlet: A bright monomeric red fluorescent protein for cellular imaging. *Nat Methods.* 2016;14:53–6.
29. Deng Y, Beahm DR, Ran X, Riley TG, Sarpeshkar R. Rapid modeling of experimental molecular kinetics with simple electronic circuits instead of with complex differential equations. *Front Bioeng Biotechnol.* 2022;10:1–21. <https://www.frontiersin.org/articles/10.3389/fbioe.2022.947508/full>
30. Guo S, Murray RM. Construction of incoherent feedforward loop circuits in a cell-free system and in cells. *ACS Synth Biol.* 2019;8:606–10. <https://doi.org/10.1021/acssynbio.8b00493>
31. Gibson DG, Young L, Chuang R-Y, Venter JC, Hutchison CA, Smith HO. Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nat Methods.* 2009;6:343–5. <http://www.nature.com/nmeth/journal/v6/n5/pdf/nmeth.1318.pdf>
32. Allaire J, RStudio. Integrated development environment for R. *J Wildl Manage.* 2011;75:1753–66. <https://doi.org/10.1002/jwmg.232>
33. Hahne F, LeMeur N, Brinkman RR, Ellis B, Haaland P, Sarkar D et al. FlowCore: A bioconductor package for high throughput flow cytometry. *BMC Bioinformatics.* 2009;10.
34. Hahne F, Gopalakrishnan N, Khodabakhshi AH, Wong C-J, Lee K. flowStats K. Statistical methods for the analysis of flow cytometry data. 2021. <http://www.github.com/RGLab/flowStats>
35. Monaco G, Chen H, Poidinger M, Chen J, De Magalhães JP, Larbi A. FlowAI: automatic and interactive anomaly discerning tools for flow cytometry data. *Bioinformatics.* 2016;32:2473–80.
36. Urchueguía A, Galbusera L, Chauvin D, Bellement G, Julou T, van Nimwegen E. Genome-wide gene expression noise in *Escherichia coli* is condition-dependent and determined by propagation of noise through the regulatory network. *PLoS Biol.* 2021;19:1–22.
37. Silander OK, Nikolic N, Zaslaver A, Bren A, Kikoin I, Alon U et al. A genome-wide analysis of promoter-mediated phenotypic noise in *Escherichia coli*. *PLoS Genet.* 2012;8.
38. Mundt M, Anders A, Murray SM, Sourjik V. A system for gene expression noise control in yeast. *ACS Synth Biol.* 2018;7:2618–26.
39. Newman JRS, Ghaemmaghami S, Ihmels J, Breslow DK, Noble M, DeRisi JL, et al. Single-cell proteomic analysis of *S. cerevisiae* reveals the architecture of biological noise. *Nature.* 2006;441:840–6.
40. Estrem ST, Ross W, Gaal T, Chen ZWS, Niu W, Ebright RH et al. Bacterial promoter architecture: subsite structure of UP elements and interactions with the carboxy-terminal domain of the RNA polymerase alpha subunit. *Genes Dev.* 1999;13:2134–47. <https://doi.org/10.1101/gad.13.16.2134>
41. Ruff EF, Drennan AC, Capp MW, Poulos MA, Artsimovitch I, Record MT. E. coli RNA Polymerase determinants of open complex lifetime and structure. *J Mol Biol.* 2015;427:2435–50. <https://linkinghub.elsevier.com/retrieve/pii/S002228361500340X>
42. Nguyen E, Poli M, Durrant MG, Kang B, Katrekar D, Li DB et al. Sequence modeling and design from molecular to genome scale with Evo. *Science.* 2024;386. <https://doi.org/10.1126/science.ado9336>
43. Rodriguez A, Diehl JD, Wright GS, Bonar CD, Lundgren TJ, Moss MJ et al. Synonymous codon substitutions modulate transcription and translation of a divergent upstream gene by modulating antisense RNA production. *Proc Natl Acad Sci.* 2024;121. <https://doi.org/10.1073/pnas.2405510121>
44. Birkholz N, Kamata K, Feussner M, Wilkinson ME, Cuba Samaniego C, Migur A et al. Phage anti-CRISPR control by an RNA- and DNA-binding helix–turn–helix protein. *Nature.* 2024;631:670–7. <https://www.nature.com/articles/s41586-024-07644-1>
45. Kim GY, Yang J, Han YH, Seo SW. Synthetic redesign of *Escherichia coli* W for faster metabolism of sugarcane molasses. *Microb Cell Fact.* 2024;23:242. <https://microbialcellfactories.biomedcentral.com/articles/10.1186/s12934-024-02520-z>
46. Liu LF, Wang JC. Supercoiling of the DNA template during transcription. *Proc Natl Acad Sci.* 1987;84:7024–7. <https://doi.org/10.1073/pnas.84.20.7024>
47. Tsao Y-P, Wu H-Y, Liu LF. Transcription-driven supercoiling of DNA: Direct biochemical evidence from in vitro studies. *Cell.* 1989;56:111–8. <https://linkinghub.elsevier.com/retrieve/pii/0092867489909896>
48. Neuman KC, Nagy A. Single-molecule force spectroscopy: optical tweezers, magnetic tweezers and atomic force microscopy. *Nat Methods.* 2008;5:491–505. <https://www.nature.com/articles/nmeth.1218>
49. Takizawa T, Meaburn KJ, Misteli T. The meaning of gene positioning. *Cell.* 2008;135:9–13.
50. Bryant JA, Sellars LE, Busby SJW, Lee DJ. Chromosome position effects on gene expression in *Escherichia coli* K-12. *Nucleic Acids Res.* 2014;42:11383–92. <http://academic.oup.com/nar/article/42/18/11383/2435330/Chromosome-position-effects-on-gene-expression-in>
51. Chen X, Zhang J. The genomic landscape of position effects on protein expression level and noise in yeast. *Cell Syst.* 2016;2:347–54. <https://linkinghub.elsevier.com/retrieve/pii/S2405471216301077>
52. Balleza E, Kim JM, Cluzel P. Systematic characterization of maturation time of fluorescent proteins in living cells. *Nat Methods.* 2017;15:47–51. <https://doi.org/10.1038/nmeth.4509>
53. Hensel Z. A plasmid-based *Escherichia coli* gene expression system with cell-to-cell variation below the extrinsic noise limit. *PLoS ONE.* 2017;12:1–15.
54. Taniguchi Y, Choi PJ, Li G-W, Chen H, Babu M, Hearn J et al. Quantifying *E. coli* Proteome and transcriptome with single-molecule sensitivity in single cells. *Science.* 2010;329:533–8. <https://www.science.org/doi/10.1126/science.1188308>
55. Silva JPN, Lopes SV, Grilo DJ, Hensel Z. Plasmids for independently tunable, low-noise expression of two genes. *Ellermeier CD, editor. mSphere.* 2019;4:1–9. <https://doi.org/10.1128/mSphere.00340-19>
56. Raser JM, O’Shea EK. Noise in gene expression: origins, consequences, and control. *Science.* 2005;309:2010–3. <https://doi.org/10.1126/science.1105891>
57. Ozbudak EM, Thattai M, Kurtser I, Grossman AD, van Oudenaarden A. Regulation of noise in the expression of a single gene. *Nat Genet.* 2002;31:69–73. <http://www.nature.com/articles/ng869z>
58. Swain PS, Elowitz MB, Siggia ED. Intrinsic and extrinsic contributions to stochasticity in gene expression. *Proc Natl Acad Sci.* 2002;99:12795–800. <http://doi.org/10.1073/pnas.162041399>
59. Paulsson J. Summing up the noise in gene networks. *Nature.* 2004;427:415–8. <https://www.nature.com/articles/nature02257>
60. Deloupy A, Sauveplane V, Robert J, Aymerich S, Jules M, Robert L. Extrinsic noise prevents the independent tuning of gene expression noise and protein mean abundance in bacteria. *Sci Adv.* 2020;6. <https://doi.org/10.1126/sciadv.abc3478>

Publisher’s note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.