Independent control of mean and noise by convolution of gene expression distributions

Karl P. Gerhardt[1], Satyajit D. Rao[1], Evan J. Olson[1], Oleg A. Igoshin[1,2,3,4], Jeffrey J. Tabor[1,2,4]*.

[1]Department of Bioengineering, [2]Systems, Synthetic, and Physical Biology (SSPB) Ph.D. Program, [3]Center for Theoretical Biophysics, [4]Department of Biosciences, Rice University, 6100 Main Street, Houston, Texas, 77005, USA

*Contact: jeff.tabor@rice.edu

Abstract

Gene expression noise can reduce cellular fitness or facilitate processes such as alternative metabolism, antibiotic resistance, and differentiation. Unfortunately, efforts to study the impacts of noise have been hampered by a scaling relationship between noise and expression level from a single promoter. Here, we use theory to demonstrate that mean and noise can be controlled independently by expressing two copies of a gene from separate inducible promoters in the same cell. We engineer high- and low-noise inducible promoters to validate this result in *Escherichia coli*, and develop a model that predicts the experimental distributions. Finally, we use our method to reveal that the response of a promoter to a repressor is less sensitive with higher repressor noise, and explain this result using a law from probability theory. Our approach can be applied to

investigate the effects of noise on diverse biological pathways or program cellular heterogeneity for synthetic biology applications.

Introduction

Protein copy numbers vary within populations of genetically-identical cells due to stochasticity in the molecular and cellular level events that impact gene expression[1–3]. This gene expression noise can be deleterious, causing metabolic or physiological challenges, or reduce the reliability with which a cell can carry out a task[4–7]. Indeed, evolution appears to have optimized genomic locus [8], promoter architecture and sequence[9], transcription and translation rate[10], and selected for negative feedback[11,12] to decrease noise in the expression of toxic, complex forming, highly-connected, and essential proteins[8,10,13–16]. On the other hand, evolution has also exploited noise as a means to regulate stress response[17–19], alternative metabolism[20], cell-fate determination[21], and pathways enabling cell populations to divide labor or hedge bets against unpredictable environments[22,23].

Gene expression noise can be considered to contain an intrinsic component, relating to stochasticity in the chemical reactions of gene expression processes, and an extrinsic component, relating to noise in global conditions or upstream factors[1,24,25]. At low numbers of expressed proteins, intrinsic noise dominates and inversely correlates with the mean. At higher protein copy numbers, extrinsic noise becomes dominant and invariant with mean.

Tools that specifically modulate noise in the expression of genes of interest are needed to study the impact of noise on biological processes. However, controlling noise independently of mean is a major challenge due to the coupling between these two parameters[9,13,14,26–28].

Several strategies for decoupling mean and noise have been reported. For example, researchers have altered promoter activation kinetics[9,29,30], operator site location and multiplicity [31,32], and introduced transcriptional feedback[33–35]. However, multiple strains must be engineered to achieve different noise levels for the same mean using these methods. Independent control of mean and noise in a single strain requires manipulation of two separate processes impacting protein copy number[36]. This result has been demonstrated by combining two small molecule responsive regulators in a cascade[37–40], altering both the frequency and bias of promoter state switching[41], tuning both transcription and mRNA degradation rates[42] or using a time-varying input to independently control promoter activation frequency and transcription rate[43].

One important limitation to all of these previous approaches is that they rely upon genetic parts, circuits, or pathways that are native to or have been optimized to function in a particular organism. As a result, substantial re-engineering may be required to achieve the same results in each new organism of interest. Additionally, there may be fundamental limitations on the levels of gene expression and noise that can be achieved using these approaches. For example, a two-step cascade primarily allows control of extrinsic noise as it relies on transmission of noise from the upstream regulator to the

output[39,40]. Likewise, modulation of promoter kinetics is expected to primarily modulate intrinsic noise[39], and is unlikely to be effective at high copy numbers.

Here, we show that combining the protein expression distributions from multiple promoters in a single cell is a simple and generalizable strategy to achieve robust and independent control of mean and noise over a wide area. To this end, we first use a simple theoretical model to reveal that the mean and noise of a population distribution can be independently controlled using two co-expressed and independently-regulated inducible promoters (IPs). We then implement this approach experimentally by constructing high- and low-noise generating IPs activated by the addition of two separate inducer molecules in *E. coli*. Next, we show that mean and noise of total gene expression can be manipulated using inducer combinations to control the activity of each IP. We characterize the steady-state behavior of cells harboring the IP pair and present a simple mathematical model to predict mean and noise from inducer concentrations. Next, we show that our experimental gene expression profiles can be predicted with high accuracy by simulating convolutions between the distributions contributed by each IP. Finally, we use our approach to independently tune mean and noise in the expression of a bacterial transcriptional repressor and analyze how each affects the activity of a target promoter independently.

Results

Model for independent control of mean and noise from two IPs

We considered two copies of a gene, encoding products G_1 and G_2, in a single cell. Here, the total amount of gene product, G, is $G_1 + G_2$. The mean (μ) value of G across a population of such cells is obtained from

$$\mu(G) = \mu(G_1) + \mu(G_2) \#(1)$$

and the variance (σ^2) from

$$\sigma(G)^2 = \sigma(G_1)^2 + \sigma(G_2)^2 + 2Cov(G_1, G_2) \#(2)$$

where Cov is the covariance. If G_1 and G_2 are regulated such that their expression is independent and covariation is negligible, the noise (η, defined as the standard deviation (σ) divided by the mean) of G is described by a weighted average of the noise from each source

$$\eta(G) = \left(\frac{\eta(G_1)^2\mu(G_1)^2 + \eta(G_2)^2\mu(G_2)^2}{(\mu(G_1) + \mu(G_2))^2}\right)^{\frac{1}{2}} \#(3)$$

and the distribution of G in the cell population is described by a convolution

$$p_G(g) = \int_{-\infty}^{\infty} p_2(g - g_1)p_1(g_1)dg_1 \#(4)$$

where p_1, p_2, and p_G are the probability density functions of G_1, G_2, and G, respectively. In the presence of significant correlation between G_1 and G_2, noise can be solved by accounting for covariance in Eq. (2).

Based on Eqs. (1-4), we reasoned that mean and noise of G can be independently controlled by regulating G_1 and G_2 expression from low- and high-noise IPs (Fig. 1a). In this approach, $\eta(G)$ can be varied between $\eta(G_1)$ and $\eta(G_2)$ while maintaining constant $\mu(G)$ by tuning the relative expression of G_1 and G_2 with ratios of IP inputs (Fig. 1b). $\eta(G)$ can be tuned this way at different values of $\mu(G)$ by controlling the absolute expression of G_1 and G_2 with the amount of IP inputs. We also reasoned that the distribution of G can be predicted from a convolution of the distributions of G_1 and G_2 if IP expression is independent (Fig 1c). IP expression may be independent if their dominant sources of noise are intrinsic and/or pathway specific[25]. One attractive feature of this approach is the direct relationship between tunability and the difference in noise produced by G_1 and G_2. IPs that produce large differences in noise over the same range of means are therefore desirable when implementing this method.

Engineering a high-noise promoter induced by AHL

To engineer a system capable of tuning noise over a wide range, we designed two IPs that produce similar mean expression levels with very different noise values. First, we engineered a high-noise IP that incorporates positive autoregulation through the 3-oxo-C_6-acylhomoserine lactone (AHL)-dependent transcriptional activator LuxR and its target promoter P_{lux}. Specifically, we expressed a bicistronic mRNA encoding the

reporter gene superfolder green fluorescent protein (*sfgfp*) and *luxR* as the output of P_{lux}. To achieve high noise levels and strong inducibility, we generated a small library of variants of this IP with *luxR* ribosome binding site (RBSs) of different strengths (Fig. 2a, Supplementary Fig. 1). We initially expected that stronger *luxR* RBSs would increase positive feedback strength by increasing the translational burst size of *luxR*. This effect, we believed, would result in more *luxR* expression per AHL molecule, a steeper AHL-mean transfer function, and higher noise at intermediate IP outputs[44–46]. To examine the performance of these high noise IP designs, we separately transformed each construct into bacteria, treated the resulting strains with different AHL concentrations and measured the resulting sfGFP fluorescence distributions via flow cytometry. Overall, the behavior of this high-noise IP family is closely consistent with our expectations (Fig. 2b-d, Supplementary Fig. 1, Supplementary Table 1). However, to our surprise weaker (not stronger) *luxR* RBSs result in increased steepness in the AHL-mean transfer function and higher noise in the AHL-noise and mean-noise transfer functions, respectively (Fig. 2b, Supplementary Fig. 1).

To understand these effects, we developed a deterministic molecular model of the high noise IP family (Supplementary Methods, Supplementary Table 2). The model details the binding interaction between LuxR and AHL and positive autoregulation by transcriptionally active LuxR:AHL. We find that this family of high noise IP designs is only sensitive to feedback when the LuxR:AHL concentration is sensitive to LuxR fluctuations, i.e. when LuxR is limiting relative to AHL. At low *luxR* translation rates and intermediate AHL concentrations, the sfGFP output is bistable and exhibits hysteresis.

In this regime, cell populations undergo abrupt jumps from low to high states resulting in steep mean transfer functions (Supplementary Fig. 2).

Within this bistable window, we expected noise-driven transitions between states to result in high gene expression variability. To verify this prediction, we constructed a stochastic model of the positive feedback circuit and computed AHL-mean and AHL-noise transfer functions using Gillespie SSA simulations (Supplementary Methods, Supplementary Table 3). Mean and noise values quantitatively match experimental values for all four RBSs (Supplementary Fig. 3). For weak RBSs, noise reaches a maximum at intermediate AHL concentrations while stronger RBSs show monotonic decreases in noise, further demonstrating ultrasensitive circuit transitions due to bistability.

Taken together, these modeling results capture the performance of our high noise IP family and reveal an unexpected inverse relationship between *luxR* RBS strength and the magnitude of feedback in the circuit. Among the high-noise IP variants we tested, the variant containing the B0031 RBS generated the highest overall noise while maintaining unimodality. With this variant (hereafter named IP_h), we have constructed a high-noise IP that satisfies our design criteria.

Engineering a low-noise promoter induced by aTc

To produce low-noise gene expression distributions over a wide range of mean values, we designed an IP wherein *sfgfp* is expressed under control of the Tetracycline

Repressor (TetR)-regulated $P_{Ltet\text{-}O1}$ promoter with *tetR* expressed from a constitutive promoter on the same plasmid (Fig. 2e). We cloned the DNA encoding this IP into separate plasmid backbones with ColE1 (50-70 copies/cell), p15a (20-30 copies/cell), and SC101* (3-4 copies/cell)[47] origins of replication (Fig. 2e). Initially, we hoped to find differences in output noise between origin of replication variants by virtue of the scaling between copy number and intrinsic noise or by differences in plasmid copy number variability[25,47,48]. Mean output range (the difference between high and low states), detection threshold (inducer concentration a half output range), and steepness all increased with plasmid copy number (Fig. 2f, Supplementary Table 4). A deterministic molecular model of the low-noise IP (Supplementary Table 5, Supplementary Methods) captures these experimental behaviors (Supplementary Fig. 4).

For all origin of replication variants, we also observed that the sfGFP noise decreases monotonically as a function of both inducer and sfGFP mean until reaching a noise floor of about 0.25 (Fig 2g, h). At low induction, higher copy number variants produce lower noise but also correspondingly higher sfGFP mean, such that all variants collapse onto a similar initial trajectory. These behaviors suggest that noise in sfGFP from this IP is dominated by intrinsic and global extrinsic sources rather than transmitted noise from TetR or differences in copy number stringency between origins of replication. At intermediate induction, mean-noise transfer functions diverge slightly, with lower copy number variants decreasing more rapidly than higher copy number variants (Fig. 2h). A stochastic model of this low-noise IP (Supplementary Methods) recapitulates these experimental results and predicts monotonic decreases in noise as a function of aTc for

all origin of replication variants, and lower noise at intermediate induction for lower copy number variants (Supplementary Fig. 3, Supplementary Table 6).

These models recapitulate the behavior of our low-noise IP and reveal that the observed performance differences that arise on different plasmid backbones are attributable to differences in repressor and promoter copy number. The SC101* variant generates the lowest overall noise and similar sfGFP mean output levels as IP_h. Thus, we renamed this variant IP_l, and carried it forward for further studies.

Independent control of mean and noise with low- and high-noise IPs

To achieve independent control of mean and noise by summing gene expression from low- and high-noise IPs, we co-transformed bacteria with plasmids encoding IP_h and IP_l (Fig. 3a). We exposed populations of the co-transformed bacteria to a 25 x 25 (625 total) panel of AHL and aTc concentrations and measured the sfGFP distributions by flow cytometry (Fig. 3b-d). We found that AHL-mean transfer functions are sigmoidal and shift higher with the level of aTc (Fig. 3b). This behavior is consistent with summation of sfGFP from IP_h and IP_l. Both the AHL-noise and mean-noise transfer functions decrease non-monotonically, peak at intermediate AHL concentrations, and shift lower with the level of aTc (Fig. 3c,d). These properties are consistent with noise being determined by the relative contribution of IP_l and IP_h to total sfGFP (Eq. 3). As intended, exposure to different inducer combinations produces an area in mean-noise space over which our system can be tuned (Fig. 3d). Thus, we can independently control mean and noise by summing gene expression from high and low noise IPs.

As predicted by our model, a wide range of sfGFP noise values can be achieved at virtually the same mean (Fig. 3d-g). At low total bacterial fluorescence levels, differences between distributions with similar mean and disparate noise is masked by *E. coli* autofluorescence (Fig. 3f). However, as sfGFP levels increase and the contribution of autofluorescence to total cellular fluorescence becomes negligible, differences between distributions with similar mean and different noise become dramatic (Fig. 3e,f). Thus, while we can tune mean and noise at low as well as high expression levels, detecting those changes at low mean requires analysis after autofluorescence subtraction.

We next determined whether the experimental mean-noise data of this system could be quantitatively described by summation of mean and weighted average of noise from each IP. Building upon a previous model of noise for regulated gene expression [49], we assigned each IP two equations describing their inducer-mean and inducer-noise transfer functions, related these functions by Eqs. (1) and (3), and fit their parameters to the experimental mean-noise data (Supplementary Fig. 5, Supplementary Methods). Following this procedure, we observe close agreement between experiment and model predictions, enabling accurate prediction of mean and noise from AHL and aTc concentrations (Fig. 3b-d, Supplementary Fig. 6, Supplementary Table 7). The accuracy of this model further supports our hypothesis of additive gene expression from our two IPs.

IP_h/IP_l outperforms previous mean-noise control systems

To our knowledge, no metric has been proposed to describe the ability of a genetically-encoded system to independently tune gene expression mean and noise. The dynamic range, or ratio of output gene expression levels in the fully active versus fully inactive states, is a 1-dimensional metric frequently used to quantify IP performance. However, mean and noise are tunable over a 2-dimensional area[37–39,41–43].

To quantify the performance of mean-noise control systems, we developed metrics that we term dynamic area (F_A) and dynamic noise (F_η) (Methods). F_A measures the fold-change in mean-noise area over which a system can be tuned, while F_η measures the largest fold-change in noise a system can achieve at a constant mean. Practically, F_A describes the capacity of a system to produce different combinations of both mean and noise, while F_η captures the ability to modulate noise, specifically, at a constant mean. We computed F_A and F_η using the experimental mean-noise dataset measured for our system, and found values of 11.39 and 6.9, respectively (Fig. 3g). By this same analysis, we find that our system performs better than any previously described noise-mean tuning system (which range from 2.22 to 10.17 in F_A and 1.61 to 4.66 in F_η) we are aware of, in any organism (Supplementary Fig. 7, Supplementary Table 8).

Convolution model predicts IP_h/IP_l distributions

We hypothesized that the sfGFP distributions produced by our combined IP_h/IP_l system could be predicted by simulating a convolution between distributions generated by IP_l and IP_h individually (Fig. 4a). To examine this hypothesis, we simulated each of the 625

experimental populations resulting from exposure to the AHL and aTc panel in Figure 3 by summing randomly sampled fluorescence events between populations induced with only AHL and populations induced with only aTc (hereafter termed marginal distributions) (Fig. 4b, Methods). These simulated distributions show remarkable similarity to their experimental counterparts and frequently capture subtle, higher-order behaviors observed in experimental distributions such as skew and bimodality (Fig. 4b). While the distribution predictions are highly accurate overall, sfGFP levels are systematically overestimated in distributions with very low mean. This occurs due to autofluorescence and leaky sfGFP fluorescence (i.e. sfGFP fluorescence in the absence of inducer) being counted twice during the summation of two cell fluorescence events. Our distribution predictions are also less accurate when the AHL-induced population is near the inflection point of the AHL-mean transfer function (Fig. 3b).

We quantified similarity between each pair of experimental and predicted distributions using the Bhattacharyya coefficient[50] (c_B), a metric ranging from 0 to 1 measuring overlap between two probability distributions (Fig. 4b,c, Methods). The average c_B for all populations is remarkably high, at 0.92 with a standard deviation of 0.094. However, due to the previously-described effect of overestimating autofluorescence and leaky sfGFP expression, c_B shows strong correlation with expression mean below about 1,000 MEFL (Supplementary Fig. 8). Mean and standard deviation of simulated distributions show strong concordance with their experimental counterparts (ρ_c of 0.969 and 0.990 respectively) and this concordance is further improved (ρ_c of 0.995 and 0.993 respectively) after compensation for autofluorescence and leaky expression (Fig. 4 c,d,

Supplementary Tables 9 and 10, Methods). Overall, this approach of simulating convolutions between two experimental marginal distributions enables simple and accurate prediction of the entire distributions generated by our system.

Repressor noise decreases steepness of promoter response

The transcription factor-promoter transfer function is the quantitative relationship between transcription factor expression level and target promoter activity. It has previously been shown that the shape, including the sensitivity, of a transcription factor-promoter transfer function can strongly depend on the level and context of transcription factor expression[51–55]. Based on previous experiments measuring the impact of noise on biological processes[29,37,43], we hypothesized that increasing noise in transcription factor expression would produce shallower transcription factor-promoter transfer functions. We used IP_h/IP_l to characterize the effect of repressor noise on mean expression from a target promoter. To that end, we first fused $PhlF^{AM}$ [51], a TetR family repressor from *Pseudomonas fluorescens*, to the C-terminus of sfGFP on both IP_h and IP_l. We then co-transformed *E. coli* with plasmids expressing the modified IP_h and IP_l along with an output plasmid containing mCherry expressed under the $PhlF^{AM}$ repressed P_{PhlF} promoter (Fig. 5a). We grew these bacteria under different combinations of AHL and aTc, allowed sfGFP and mCherry levels to reach steady-state, and quantified both fluorescent proteins by flow cytometry (Fig. 5b-e). As before, mean-noise transfer functions decrease non-monotonically with respect to AHL induction, peak at intermediate AHL concentrations, and shift lower upon addition of aTc (Fig. 5b). Conversely, induction with only aTc results in comparatively low noise which

monotonically decreases with higher mean (Fig. 5b). As a result, the system achieves different sfGFP-PhlF^{AM} noise while maintaining the same mean by applying different amounts and ratios of AHL and aTc. Compared to sfGFP-PhlF^{AM}, mCherry noise is higher overall and shows only modest differences at the same mean between different induction conditions (Fig. 5c). This behavior suggests most noise transmitted from PhlF^{AM} to mCherry is filtered out or represents only a small portion of the total mCherry noise. Mean mCherry fluorescence decreases in response to increasing mean sfGFP-PhlF^{AM} in a manner that strongly depends on noise (Fig. 5d). When induced with AHL, mean mCherry begins to decrease at lower mean sfGFP-PhlF^{AM}, but requires dramatically higher mean sfGFP-PhlF^{AM} to become fully repressed compared to when induced with aTc (Fig. 5d, Supplementary Fig. 9). Moreover, induction with combinations of AHL and aTc generates intermediate levels of this effect. Noise, therefore, makes the PhlF-promoter transfer function less steep, increases apparent PhlF^{AM} activity at low mean, and decreases apparent PhlF^{AM} activity at high mean. This striking result demonstrates how a single transcription factor can generate vastly different population level behaviors depending on the details of its expression.

To model this population behavior, we applied the Law of the Unconscious Statistician (LOTUS) to our system (Methods). LOTUS can be used to compute the mean of a function of a random variable and was previously used to model the effects of noise in an antibiotic resistance gene on cellular fitness[37]. Here, population mean mCherry is found through the probability distribution of sfGFP-PhlF^{AM} and an experimentally determined single-cell I/O relationship between mCherry and sfGFP-PhlF^{AM}. After fitting

the single-cell I/O relationship to fluorescence data, mean mCherry values predicted by LOTUS show a qualitative match with our data (Fig. 5e, Supplementary Fig. 10, Supplementary Table 11). While the qualitative behavior of the system is captured by this model, strong cell autofluorescence signal at low expression levels likely undermine the single-cell fit and a more quantitative agreement.

Discussion

We have shown that regulating the expression of two copies of the same gene that produce different levels of noise is a simple method for controlling and studying gene expression noise. In nature, gene duplication has been proposed as a mechanism to resolve the tradeoff between expression noise and environmental responsiveness[56]. This hypothesis was recently validated for a pair of duplicated transcription factors in *S. cerevisiae*, where one transcription factor is constitutively expressed with low noise, while the other increases from undetectable to highly expressed with high noise under environmental stress[57]. Noise in this transcription factor pair was associated with a negative fitness advantage, so this strategy likely evolved to minimize noise under normal conditions, while maintaining responsiveness under stressful conditions. For some genes, noise can have a positive or negative fitness advantage depending on environmental conditions. Natural examples of cells regulating gene duplicates which produce different levels of noise, as we have demonstrated here, may have evolved as a strategy to dynamically regulate noise in such genes.

Unlike previous approaches to modulating mean and noise in gene expression, our method does not require the use of a particular genetic part or circuit, a specific mechanism of noise reduction or amplification, or a particular host organism. Rather, it only requires the availability of orthogonal high- and low-noise generating IPs in an organism of interest, and the ability to co-transform them into a single cell.

Though several mean-noise tuning systems have been reported, there had been no method available to benchmark their performance against one another. Here, we propose dynamic area and dynamic noise to quantify the performance of mean-noise tuning systems. These metrics capture the magnitude of the mean-noise area and the noise accessible at a constant mean, respectively. Like dynamic range, these metrics are independent of the absolute values of mean and noise. By comparing these two variables, one could distinguish systems that are tunable in both dimensions from systems which are primarily tunable in just a single dimension.

Our model based on summation of gene expression from two IPs accurately predicts population-level mean and noise of fluorescence distributions from inducer concentrations. The success of this approach supports the hypothesis for additive gene expression from our two engineered sources, and may be adapted to describe future implementations of this method.

We can predict full gene expression distributions by simulating convolutions between experimental IP_l and IP_h distributions. While the predictions are systematically

overestimated when autofluorescence and leaky reporter gene fluorescence dominate (at low mean), they are otherwise surprisingly accurate and able to capture detailed population features that would not be predicted following a parametric approach. Going forward, our predictions would benefit from a method to reduce or deconvolve out autofluorescence from flow cytometry measurements. While convolution predictions require advance measurement of marginal distributions, the number of convolutions that can be predicted combinatorially increases with number of marginal distributions measured. We imagine this predictive ability could be utilized to forward engineer desired convolution distributions, including their higher moments and noise types, from existing promoter libraries[58–64], constitutive or otherwise. Likewise, our method could be used to combinatorially increase the number of gene expression distributions achievable with a constitutive promoter library by expressing the same gene from combinations of two or more library members.

Our results indicate that regulated promoters respond to their cognate transcription factors at lower concentrations, but with less sensitivity, when noise in the level of the transcription factor is high. As a result, a noisy transcription factor has higher apparent activity than a less noisy transcription factor at low expression levels, and the opposite is true at high levels. These findings could be used to anticipate the effect of changing the distribution of a transcriptional regulator, or create design principles for predicting and programming the shape of transcriptional dose response curves.

Our method could be used to study other noise-dependent biological phenomena such as transient stochastic resistance[19,65], persister cell formation[17,18], stochastic differentiation[21], noise-induced cooperative behaviors[22,23], or gene circuit stability [66] in cell populations. Such studies could provide a greater experimental basis for the fitness advantages conferred by noise or be used to create design principles for engineering desirable cell population behaviors using noise-driven processes.

Gene expression convolution could also be used to engineer noisy phenotypes into populations of living cells. Example behaviors include genetically-identical populations that automatically differentiate into specified ratios of cell sub-types[67], form Turing-type patterns using activator and inhibitor morphogens with similar diffusion rates[68], or stochastically lyse in order to release enzymes to enable the population to metabolize complex agricultural feedstocks[69]. Taken together, gene expression convolution is a simple strategy for studying and controlling gene expression noise in a wide range of organisms and biological pathways.

Methods

Molecular biology

The plasmids used in this study are listed in Supplementary Table 12**.** Plasmid maps are shown in Supplementary Fig. 11. DNA assembly was performed by the Golden Gate method[70] and cloning was performed in strain NEB 10β (New England Biolabs). A PhlF mutant ($PhlF^{AM}$) with improved repression activity was amplified from the genome of strain sAJM.1506[51] and used in this study.

Cell growth and chemical induction

IP characterization and convolution experiments

Experiments were performed in strain MG1655 in M9 media + 100 mM HEPES (pH 6.6), at 37°C and 250 RPM of shaking. Media was supplemented with ampicillin (50 µg/mL), spectinomycin (100 µg/mL) and chloramphenicol (35 µg/mL) as appropriate to maintain plasmids.

Frozen preculture aliquots of each experimental strain were generated by growing transformants to exponential phase ($OD_{600} \approx 0.1$), adding glycerol to 18% (v/v), recording OD_{600}, and freezing 100 µL aliquots in PCR tubes at -80°C.

Panels of chemical inducer concentrations were prepared by diluting varying amounts of AHL (Sigma-Aldrich, K3007) in media, and aTc (Takara, 63130) in 100% ethanol, in wells of 96-well plates. Inducer concentrations in each panel well were prepared to 200X the desired final concentration. Panel plates were sealed with adhesive foil, stored at -30°C, and warmed to room temperature before use in experiments.

Experimental cultures were prepared by diluting a volume of preculture in media to achieve a cell density of $OD_{600} = 2 \times 10^{-5}$. Culture media was distributed among wells of 24-well plates (1 mL/well) and supplemented with the desired chemical inducer concentration by multichannel pipetting solution from chemical inducer panels. Culture

plates were then sealed with adhesive foil and grown for 6 h to $OD_{600} \leq 0.3$, after which time they were iced for ≥ 15 min and measured by flow cytometry.

Transcription factor noise experiments

Experiments were performed as described above but in LB media. This was found to be necessary due to high metabolic burden from mCherry expression.

Flow cytometry

Flow Cytometry was performed with a BD FACScan flow cytometer outfitted with blue (488 nm, 30 mW) and yellow (561 nm, 50 mW) solid-state lasers (Cytek). sfGFP fluorescence was measured in the FL1 channel with a 510/20 nm emission filter and mCherry fluorescence was measured in the FL3 channel with 650 nm long pass filter. Event rates of 1,000–3,500 events/s were used, and all events were captured until 20,000 events occurred within an SSC-FSC area characteristic of the strain. Calibration beads (Spherotech, RCP30-5A) were measured at the end of each cytometry session. Flow cytometry files were processed using FlowCal[71]. A gate fraction of 0.3 was used to gate events in the SSC and FSC channels and FL1/FL3 arbitrary fluorescence units were calibrated to MEFL/MECY using calibration bead data collected during each respective cytometry session (Supplementary Fig. 12).

Population mean and variance calculation

Flow cytometry data for each sample was analyzed by a custom Python script (Supplementary Data 1) which calculates arithmetic mean and noise from fluorescence

distributions. The script first trims a small number of outlier observations which can heavily influence sample noise. Trimming is performed by first calculating a smooth estimation of the probability density function corresponding to the log-fluorescence distribution of the sample via kernel density estimation. The range of fluorescence values to keep is then determined by identifying the points nearest to the median at which the density estimate falls below a 0.5% threshold. Plots of the sample histogram which display the effect of the trimming are generated and assessed to ensure the trimming functions as expected. Mean and variance of a likewise analyzed, untransformed, MG1655 sample are then subtracted from the mean and variance of experimental samples.

System performance analysis

The concave hull of $\log_{10}$ transformed mean-noise pointsets was used to calculate both F_A and F_η. R packages “alphahull” and “sp” were used to find concave hulls and convert hulls to polygon objects, respectively. Alpha parameter was chosen manually. We report $\log_{10}(F_A)$ as the area of the polygon defined by the hull, and $\log_{10}(F_\eta)$ as the length of the longest vertical chord spanning the hull. The mean-noise pointset of the system reported here and R script used to obtain F_A and F_η are included in Supplementary Data 2.

Convolution distribution simulation and analysis

Fluorescence events were randomly sampled (n=50,000) from populations which received no AHL and summed with likewise sampled events from populations which

received no aTc. To simulate the population which received neither AHL nor aTc, events were sampled from that population twice and then summed.

Experimental and simulated fluorescence distributions were converted to density estimates, p_e and p_s, respectively, to calculate Bhattaccharyya coefficients:

$$c_B = \int_0^\infty \sqrt{p_e(x_i)p_s(x_i)}dx \#(5)$$

where x is FL1 fluorescence (MEFL).

Compensation of simulated distribution mean and standard deviation was performed by subtracting the fluorescence mean and variance of the fully uninduced population (signal from cell autofluorescence and leaky sfGFP expression) from the uncompensated mean and variance. Weights of $1/y^2$ were used for linear regression. Agreement between $\log_{10}$-scale experimental and simulated mean and standard deviation was measured by Lin's concordance correlation coefficient (ρ_c)[72].

LOTUS model for transcriptional output of a noisy regulator

Mean transcriptional output from the P_{PhlF} promoter was modeled using LOTUS:

$$\mu(c) = \int_0^\infty f(x)p_e(x)dx \#(6)$$

where mean mCherry expression, $\mu(c)$, of a population is found through a single cell relationship, $f(x)$, between FL1 and FL3 fluorescence (Supplementary Methods) and the experimental FL1 probability density estimate $p_e(x)$. Model predictions in FL1 and FL3 were converted to sfGFP and mCherry by background subtraction.

Statistical analysis

Points and density estimates throughout the text are single replicates collected over one to three separate experiments as indicated in figure legends. Standard errors on model fits are shown in Supplementary Information.

Code availability

Code used to calculate arithmetic mean and noise from fluorescence distributions is included in Supplementary Data 1. Code used to calculate systems performance metrics is included in Supplementary Data 2.

Data availability

The datasets generated during and/or analyzed during the current study are available from the corresponding author upon request.

References

1. Swain, P. S., Elowitz, M. B. & Siggia, E. D. Intrinsic and extrinsic contributions to stochasticity in gene expression. *Proc. Natl. Acad. Sci. U.S.A.* **99**, 12795–12800 (2002).

2. Cai, L., Friedman, N. & Xie, X. S. Stochastic protein expression in individual cells at the single molecule level. *Nature* **440**, 358–362 (2006).
3. McAdams, H. H. & Shapiro, L. A bacterial cell-cycle regulatory network operating in time and space. *Science (New York, N.Y.)* **301**, 1874–7 (2003).
4. Schmiedel, J. M., Carey, L. B. & Lehner, B. Empirical mean-noise fitness landscapes reveal the fitness impact of gene expression noise. *Nature Communications* **10**, 3180 (2019).
5. Wang, Z. & Zhang, J. Impact of gene expression noise on organismal fitness and the efficacy of natural selection. *PNAS* **108**, E67–E76 (2011).
6. Duveau, F. *et al.* Fitness effects of altering gene expression noise in Saccharomyces cerevisiae. *eLife* **7**, e37272 (2018).
7. Keren, L. *et al.* Massively Parallel Interrogation of the Effects of Gene Expression Levels on Fitness. *Cell* **166**, 1282-1294.e18 (2016).
8. Batada, N. N. & Hurst, L. D. Evolution of chromosome organization driven by selection for reduced gene expression noise. *Nature Genetics* **39**, 945–949 (2007).
9. Raser, J. M. & O'Shea, E. K. Control of stochasticity in eukaryotic gene expression. *Science* **304**, 1811–1814 (2004).
10. Fraser, H. B., Hirsh, A. E., Giaever, G., Kumm, J. & Eisen, M. B. Noise Minimization in Eukaryotic Gene Expression. *PLoS Biol* **2**, (2004).
11. Ramsey, S. A. *et al.* Dual feedback loops in the GAL regulon suppress cellular heterogeneity in yeast. *Nat. Genet.* **38**, 1082–1087 (2006).
12. Ji, N. *et al.* Feedback control of gene expression variability in the Caenorhabditis elegans Wnt pathway. *Cell* **155**, 869–880 (2013).

13. Taniguchi, Y. *et al.* Quantifying E. coli Proteome and Transcriptome with Single-Molecule Sensitivity in Single Cells. *Science* **329**, 533–538 (2010).
14. Newman, J. R. S. *et al.* Single-cell proteomic analysis of S. cerevisiae reveals the architecture of biological noise. *Nature* **441**, 840–846 (2006).
15. Lehner, B. Selection to minimise noise in living systems and its implications for the evolution of gene expression. *Mol. Syst. Biol.* **4**, 170 (2008).
16. Barroso, G. V., Puzovic, N. & Dutheil, J. Y. The Evolution of Gene-Specific Transcriptional Noise Is Driven by Selection at the Pathway Level. *Genetics* **208**, 173–189 (2018).
17. Balaban, N. Q., Merrin, J., Chait, R., Kowalik, L. & Leibler, S. Bacterial persistence as a phenotypic switch. *Science* **305**, 1622–1625 (2004).
18. Rotem, E. *et al.* Regulation of phenotypic variability by a threshold-based mechanism underlies bacterial persistence. *Proc. Natl. Acad. Sci. U.S.A.* **107**, 12541–12546 (2010).
19. El Meouche, I., Siu, Y. & Dunlop, M. J. Stochastic expression of a multiple antibiotic resistance activator confers transient resistance in single cells. *Scientific Reports* **6**, 19538 (2016).
20. Carey, J. N. *et al.* Regulated Stochasticity in a Bacterial Signaling Network Permits Tolerance to a Rapid Environmental Change. *Cell* **173**, 196-207.e14 (2018).
21. Balázsi, G., van Oudenaarden, A. & Collins, J. J. Cellular Decision-Making and Biological Noise: From Microbes to Mammals. *Cell* **144**, 910–925 (2011).
22. Chen, L., Wang, R., Zhou, T. & Aihara, K. Noise-induced cooperative behavior in a multicell system. *Bioinformatics* **21**, 2722–2729 (2005).

23. Ackermann, M. *et al.* Self-destructive cooperation mediated by phenotypic noise. *Nature* **454**, 987–990 (2008).

24. Elowitz, M. B., Levine, A. J., Siggia, E. D. & Swain, P. S. Stochastic gene expression in a single cell. *Science* **297**, 1183–1186 (2002).

25. Raser, J. M. & O'Shea, E. K. Noise in gene expression: origins, consequences, and control. *Science* **309**, 2010–2013 (2005).

26. Bar-Even, A. *et al.* Noise in protein expression scales with natural protein abundance. *Nature Genetics* **38**, 636–643 (2006).

27. Kim, K. H., Choi, K., Bartley, B. & Sauro, H. M. Controlling E. coli Gene Expression Noise. *IEEE Trans Biomed Circuits Syst* **9**, 497–504 (2015).

28. Hornung, G. *et al.* Noise–mean relationship in mutated promoters. *Genome Res.* **22**, 2409–2417 (2012).

29. Blake, W. J. *et al.* Phenotypic consequences of promoter-mediated transcriptional noise. *Mol. Cell* **24**, 853–865 (2006).

30. Murphy, K. F., Adams, R. M., Wang, X., Balázsi, G. & Collins, J. J. Tuning and controlling gene expression noise in synthetic gene networks. *Nucleic Acids Res* **38**, 2712–2726 (2010).

31. Becskei, A., Kaufmann, B. B. & van Oudenaarden, A. Contributions of low molecule number and chromosomal positioning to stochastic gene expression. *Nat. Genet.* **37**, 937–944 (2005).

32. Murphy, K. F., Balázsi, G. & Collins, J. J. Combinatorial promoter design for engineering noisy gene expression. *Proc Natl Acad Sci U S A* **104**, 12726–12731 (2007).

33. Becskei, A. & Serrano, L. Engineering stability in gene networks by autoregulation. *Nature* **405**, 590–593 (2000).
34. Nevozhay, D., Adams, R. M., Murphy, K. F., Josic, K. & Balázsi, G. Negative autoregulation linearizes the dose-response and suppresses the heterogeneity of gene expression. *Proc. Natl. Acad. Sci. U.S.A.* **106**, 5123–5128 (2009).
35. To, T.-L. & Maheshri, N. Noise can induce bimodality in positive transcriptional feedback loops without bistability. *Science* **327**, 1142–1145 (2010).
36. Briat, C. & Khammash, M. Computer control of gene expression: Robust setpoint tracking of protein mean and variance using integral feedback. in *2012 IEEE 51st IEEE Conference on Decision and Control (CDC)* 3582–3588 (2012). doi:10.1109/CDC.2012.6426720.
37. Zhuravel, D. *et al.* Phenotypic impact of regulatory noise in cellular stress-response pathways. *Syst Synth Biol* **4**, 105–116 (2010).
38. Blake, W. J., KAErn, M., Cantor, C. R. & Collins, J. J. Noise in eukaryotic gene expression. *Nature* **422**, 633–637 (2003).
39. Aranda-Díaz, A., Mace, K., Zuleta, I., Harrigan, P. & El-Samad, H. Robust Synthetic Circuits for Two-Dimensional Control of Gene Expression in Yeast. *ACS Synth Biol* **6**, 545–554 (2017).
40. Bonny, A. R., Fonseca, J. P., Park, J. E. & El-Samad, H. Orthogonal control of mean and variability of endogenous genes in a human cell line. *Nature Communications* **12**, 292 (2021).
41. Hung, M. *et al.* Modulating the frequency and bias of stochastic switching to control phenotypic variation. *Nature Communications* **5**, 4574 (2014).

42. Mundt, M., Anders, A., Murray, S. M. & Sourjik, V. A System for Gene Expression Noise Control in Yeast. *ACS Synth. Biol.* **7**, 2618–2626 (2018).
43. Benzinger, D. & Khammash, M. Pulsatile inputs achieve tunable attenuation of gene expression variability and graded multi-gene regulation. *Nat Commun* **9**, 3521 (2018).
44. Louis, M. & Becskei, A. Binary and graded responses in gene networks. *Sci. STKE* **2002**, pe33 (2002).
45. Hooshangi, S., Thiberge, S. & Weiss, R. Ultrasensitivity and noise propagation in a synthetic transcriptional cascade. *Proc. Natl. Acad. Sci. U.S.A.* **102**, 3581–3586 (2005).
46. Rossi, N. A. & Dunlop, M. J. Customized Regulation of Diverse Stress Response Genes by the Multiple Antibiotic Resistance Activator MarA. *PLoS Comput. Biol.* **13**, e1005310 (2017).
47. Lutz, R. & Bujard, H. Independent and tight regulation of transcriptional units in Escherichia coli via the LacR/O, the TetR/O and AraC/I1-I2 regulatory elements. *Nucleic Acids Res* **25**, 1203–1210 (1997).
48. Wong Ng, J., Chatenay, D., Robert, J. & Poirier, M. G. Plasmid copy number noise in monoclonal populations of bacteria. *Phys. Rev. E* **81**, 011909 (2010).
49. Pedraza, J. M. & Oudenaarden, A. van. Noise Propagation in Gene Networks. *Science* **307**, 1965–1969 (2005).
50. Bhattacharyya, A. On a Measure of Divergence between Two Multinomial Populations. *Sankhyā: The Indian Journal of Statistics (1933-1960)* **7**, 401–406 (1946).

51. Meyer, A. J., Segall-Shapiro, T. H., Glassey, E., Zhang, J. & Voigt, C. A. Escherichia coli ‘Marionette’ strains with 12 highly optimized small-molecule sensors. *Nat. Chem. Biol.* **15**, 196–204 (2019).
52. Schmidl, S. R., Sheth, R. U., Wu, A. & Tabor, J. J. Refactoring and optimization of light-switchable Escherichia coli two-component systems. *ACS synthetic biology* **3**, 820–31 (2014).
53. Anderson, J. C., Voigt, C. A. & Arkin, A. P. Environmental signal integration by a modular AND gate. *Mol Syst Biol* **3**, (2007).
54. Yokobayashi, Y., Weiss, R. & Arnold, F. H. Directed evolution of a genetic circuit. *Proc. Natl. Acad. Sci. U.S.A.* **99**, 16587–16591 (2002).
55. Egbert, R. G. & Klavins, E. Fine-tuning gene networks using simple sequence repeats. *PNAS* **109**, 16817–16822 (2012).
56. Lehner, B. Conflict between Noise and Plasticity in Yeast. *PLOS Genetics* **6**, e1001185 (2010).
57. Chapal, M., Mintzer, S., Brodsky, S., Carmi, M. & Barkai, N. Resolving noise–control conflict by gene duplication. *PLOS Biology* **17**, e3000289 (2019).
58. Mutalik, V. K. *et al.* Precise and reliable gene expression via standard transcription and translation initiation elements. *Nat. Methods* **10**, 354–360 (2013).
59. Engler, C. *et al.* A Golden Gate Modular Cloning Toolbox for Plants. *ACS Synth. Biol.* **3**, 839–843 (2014).
60. Fonseca, J. P. *et al.* A Toolkit for Rapid Modular Construction of Biological Circuits in Mammalian Cells. *ACS Synth. Biol.* **8**, 2593–2606 (2019).

61. Lee, M. E., DeLoache, W. C., Cervantes, B. & Dueber, J. E. A Highly Characterized Yeast Toolkit for Modular, Multipart Assembly. *ACS Synth. Biol.* **4**, 975–986 (2015).
62. Moore, S. J. *et al.* EcoFlex: A Multifunctional MoClo Kit for E. coli Synthetic Biology. *ACS Synth. Biol.* **5**, 1059–1069 (2016).
63. Tschirhart, T. *et al.* Synthetic Biology Tools for the Fast-Growing Marine Bacterium Vibrio natriegens. *ACS Synth. Biol.* **8**, 2069–2079 (2019).
64. Guiziou, S. *et al.* A part toolbox to tune genetic expression in Bacillus subtilis. *Nucleic Acids Res* **44**, 7495–7508 (2016).
65. Meouche, I. E. & Dunlop, M. J. Heterogeneity in efflux pump expression predisposes antibiotic-resistant cells to mutation. *Science* **362**, 686–690 (2018).
66. Turcotte, M., Garcia-Ojalvo, J. & Süel, G. M. A genetic timer through noise-induced stabilization of an unstable state. *PNAS* **105**, 15732–15737 (2008).
67. Aronson, M. S., Ricci-Tam, C., Zhu, X. & Sgro, A. E. Exploiting noise to engineer adaptability in synthetic multicellular systems. *Current Opinion in Biomedical Engineering* **16**, 52–60 (2020).
68. Karig, D. *et al.* Stochastic Turing patterns in a synthetic bacterial population. *PNAS* **115**, 6572–6577 (2018).
69. Egbert, R., Brettner, L., Zong, D. & Klavins, E. Self-destructive altruism in a synthetic developmental program enables complex feedstock utilization. *bioRxiv* 086900 (2017) doi:10.1101/086900.
70. Engler, C., Kandzia, R. & Marillonnet, S. A one pot, one step, precision cloning method with high throughput capability. *PloS one* **3**, e3647 (2008).

71. Castillo-Hair, S. M. *et al.* FlowCal: A user-friendly, open source software tool for automatically converting flow cytometry data from arbitrary to calibrated units. *ACS synthetic biology* (2016) doi:10.1021/acssynbio.5b00284.

72. Lin, L. I.-K. A Concordance Correlation Coefficient to Evaluate Reproducibility. *Biometrics* **45**, 255–268 (1989).

Acknowledgements

We thank Dr. Joel Moake for the use of his flow cytometer. This material is based upon work supported by the National Science Foundation under Grant No. CAREER 1553317, the Office of Naval Research under Grant No. N00014-18-1-2611, and the Welch Foundation under Grant No. C-1856.

Author Contributions

K.P.G. and J.J.T. conceived of the project. J.J.T. and O.A.I. supervised the project. K.P.G. designed and constructed plasmids, performed experiments, and analyzed data. E.J.O. developed and wrote the code for calculation of sample mean and noise. S.D.R. developed molecular models and performed stochastic simulations. K.P.G. and E.J.O. developed the model for mean and noise. K.P.G. developed and conducted performance analysis and convolution simulation technique. K.P.G. and J.J.T. wrote the manuscript.

Competing Financial Interests

The authors declare no competing financial interests.

Figures

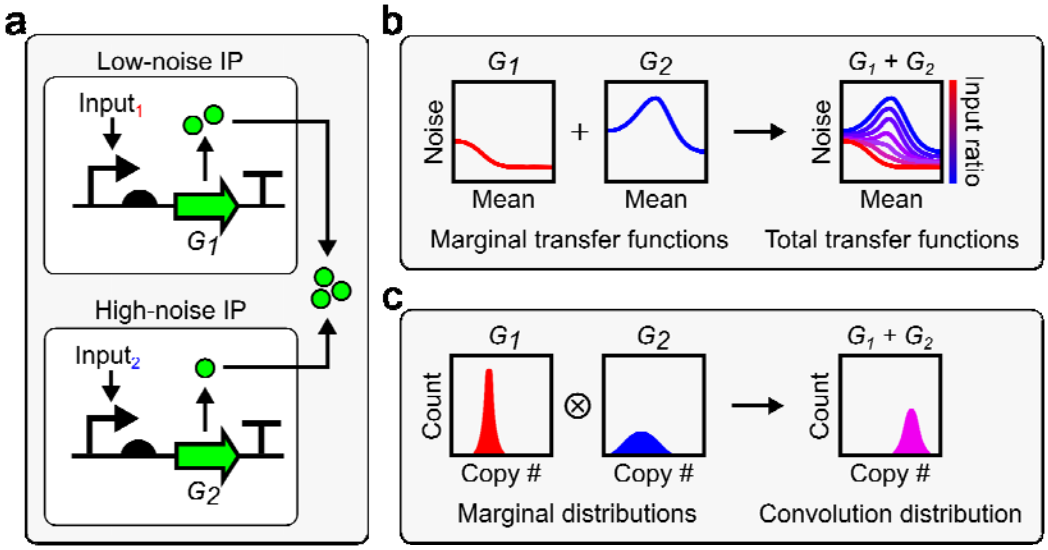

Fig. 1. Summation of gene expression from low and high noise IPs. (**a**) Two copies (and) of the same gene are expressed from two independently regulated IPs: one which produces low noise distributions and one which produces high noise distributions. and sum inside cells. (**b**) Mean and noise of and , separately, are functions of their respective IP's inputs (marginal transfer functions) and have single mean-noise trajectories. Mean and noise of the sum of and is a function of the amount and ratio of both IP's inputs (total transfer functions) and can be tuned within the area defined by their marginal transfer functions. (**c**) When and are summed, their marginal distributions form a convolution ( ).

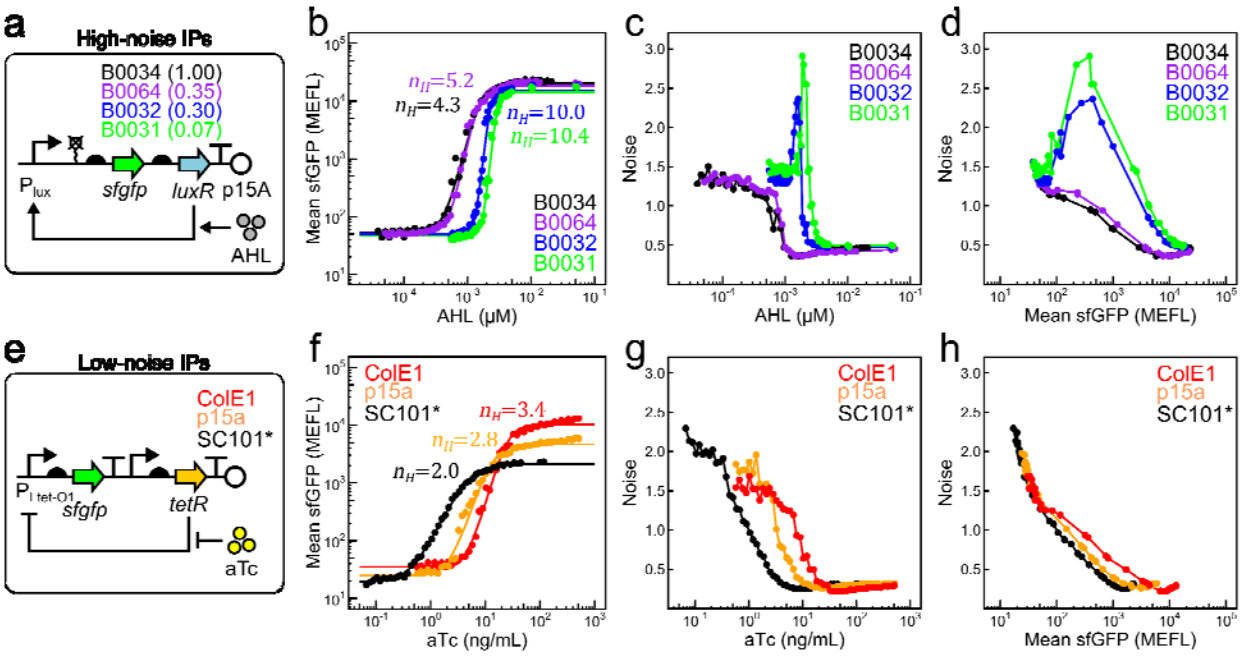

Fig. 2. Engineering high- and low-noise IPs controlled by AHL and aTc. (**a**) Library of high noise IPs based on LuxR:AHL-mediated positive transcriptional feedback. *LuxR* is encoded under RBSs of variable strength (indicated above the RBS) in different library members. Steady-state AHL-mean (**b**), AHL-noise (**c**), and mean-noise (**d**) transfer functions for high noise IP variants from (**a**). (**e**) Library of low noise IPs based on TetR:aTc-inducible gene expression without feedback. Different plasmid origins on which this IP is introduced are indicated. Steady-state aTc-mean (**f**), aTc-noise (**g**), and mean-noise (**h**) transfer functions for origin of replication variants from (**e**). Points within each group are single replicates collected from two separate experiments performed on two separate days. Smooth lines in (**b**) and (**f**) are fits to Hill functions with Hill coefficients (indicated.

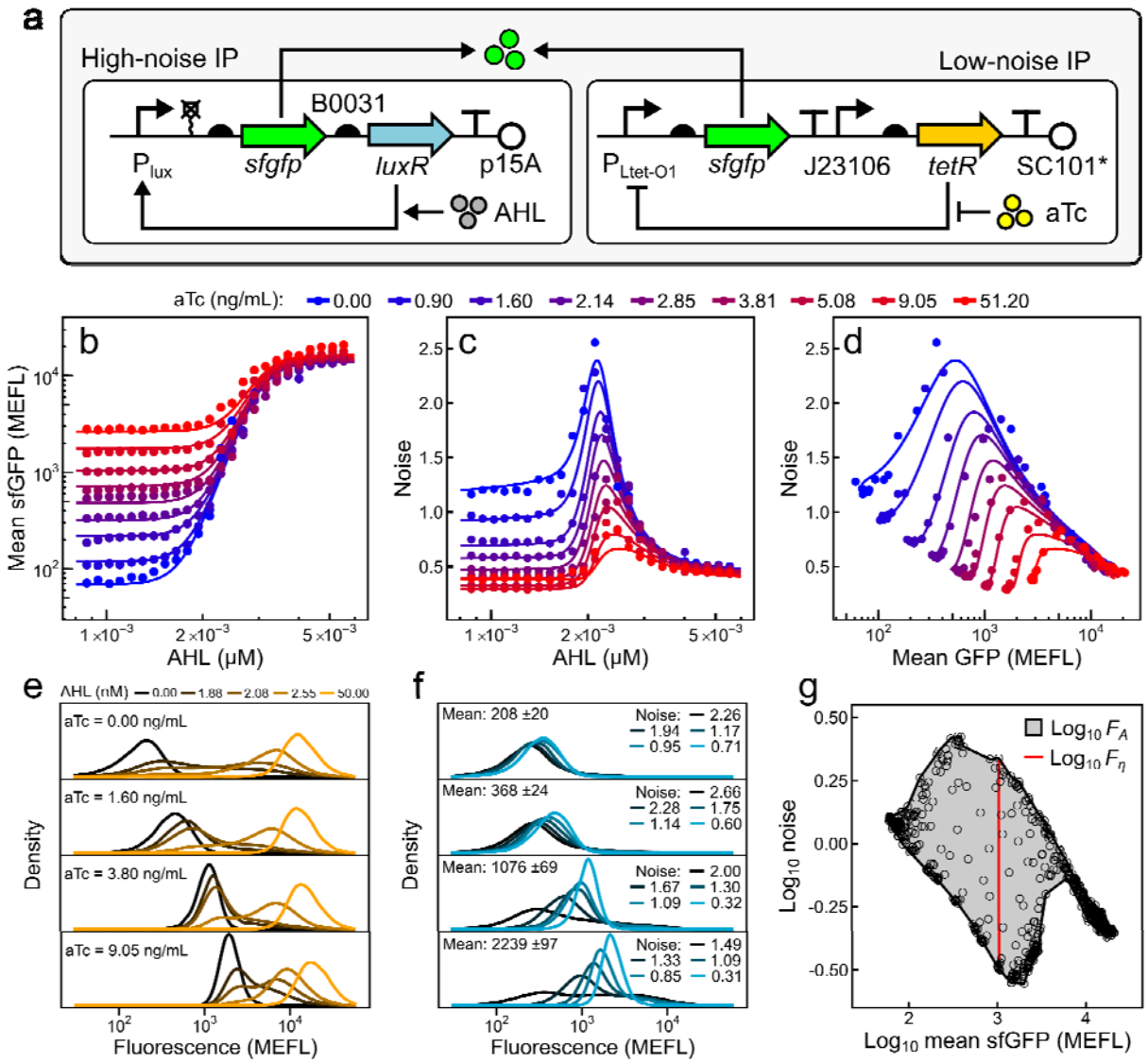

Fig. 3. Independent control of mean and noise by summation of gene expression from IP_h/IP_l. (**a**) Plasmids encoding IP_h and IP_l are co-transformed in MG1655 and induced with AHL and aTc, respectively. sfGFP from each IP sum inside cells. Steady-state AHL-mean (**b**), AHL-noise (**c**), and mean-noise (**d**) transfer functions of cells harboring plasmids in (**a**) with exposure to combinations of AHL and aTc. Mean and noise values (points) and model fits (lines) shown are a subset (9 out of 25 aTc concentrations) selected for evenly spaced visualization. Selected fluorescence density estimates of cell populations induced with AHL and aTc (**e**) over a wide range of inducer concentrations, or (**f**) to a wide range of noise at virtually the same mean. (**g**) Concave hull (black line) of the complete mean-noise pointset (circles) used to calculate dynamic area and dynamic noise. Points and density estimates are single replicates collected ofrom three separate experiments performed on three separate days.

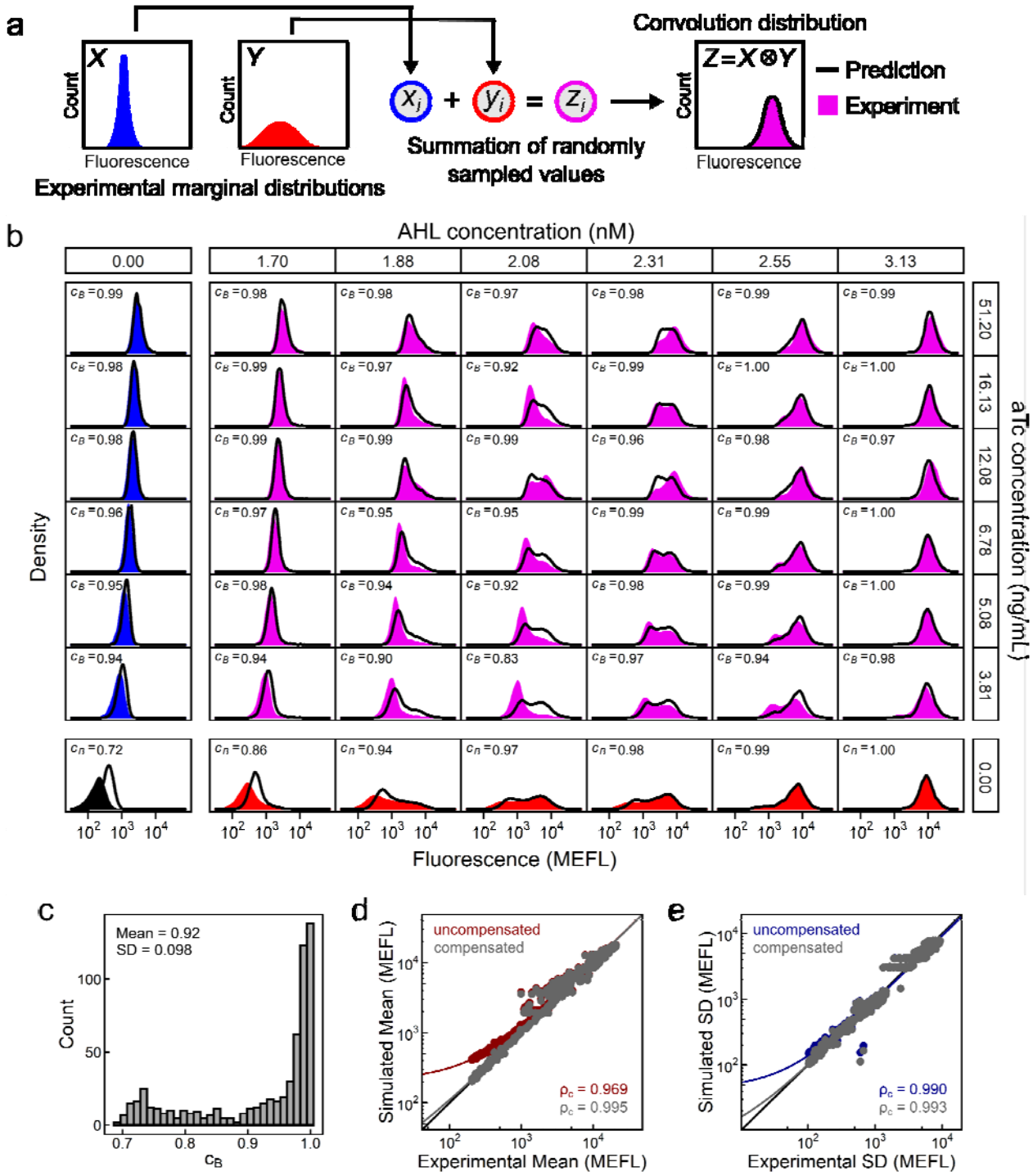

Fig. 4. Convolution predicts gene expression distributions from IP_h/IP_l. (**a**) Workflow for simulating convolution distributions. Fluorescence events from two experimental marginal distributions are randomly sampled and summed together. Summed events form a predicted convolution distribution which can be compared with an experimental counterpart. (**b**) Experimental (filled histograms) and predicted (black lines) fluorescence distributions of cell populations induced with combinations of AHL (vertically aligned, 7 of 25 shown) and aTc (horizontally aligned, 7 of 25 shown)

concentrations. Predicted distributions were simulated by summing randomly sampled events between populations with no AHL induction (first column) and no aTc induction (bottom row). Bhattacharyya coeffecicients (c_B) for each experimental-simulated distribution pair is listed in each subpanel. (**c**) Distribution of c_B for all 625 pairs of experimental and simulated distributions. (**d**,**e**) Comparison of mean and standard deviation of experimental and simulated distributions. Equivalence line (black line), linear regression lines (colored lines), and Lin's concordance coefficients ρ_c are shown (**Methods**). Compensated values were generated by subtracting the mean and variance of the fully uninduced population (black histogram in (**b**)) from the uncompensated mean and variance. Experimental density estimates are single replicates collected from three separate experiments performed on three separate days.

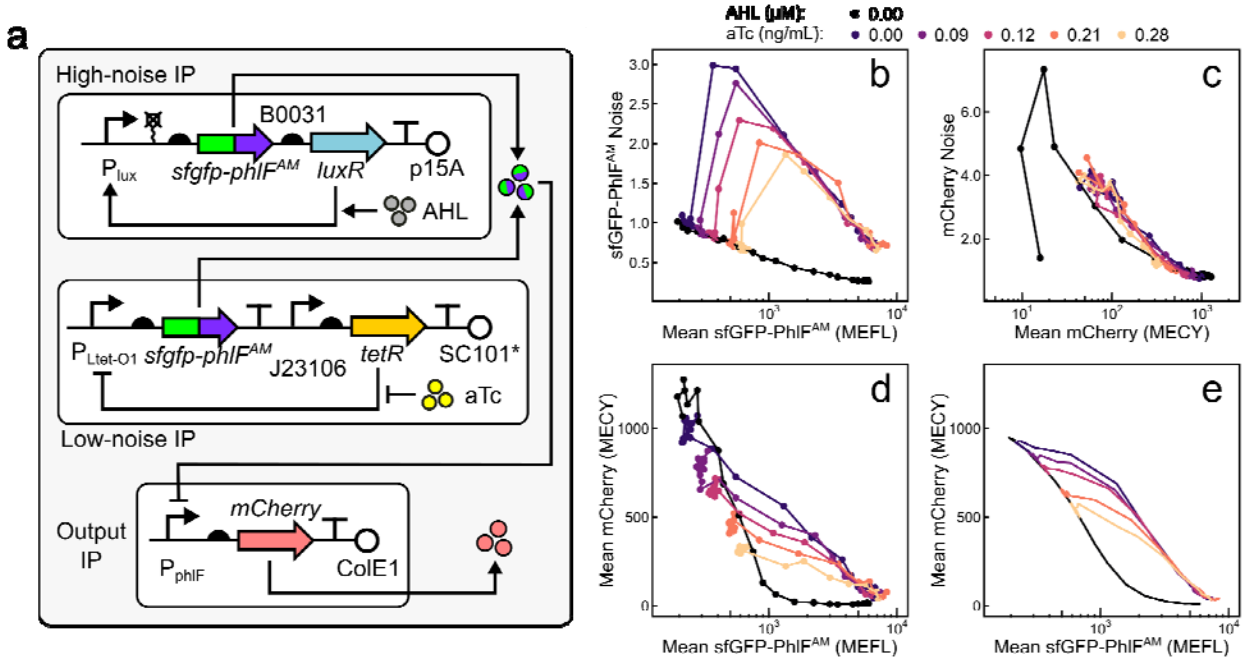

Fig. 5. Noise modulates repressor activity on a target promoter. (**a**) Schematic representation of plasmids used to control noise of a transcriptional regulator and monitor its output. PhlF^{AM} is fused to the C-terminus of sfGFP and expressed from IP_l and IP_h. PhlF^{AM} activity is monitored by measuring expression of mCherry from the P_{PhlF} promoter on the output IP. Steady-state sfGFP-PhlF^{AM} (**b**) and mCherry (**c**) mean-noise transfer functions of cells harboring plasmids in (**a**) with exposure to combinations of AHL and aTc. (**d**) Mean mCherry expression as a function of mean sfGFP-PhlF^{AM} expression. (**e**) LOTUS model prediction of mean mCherry as a function of sfGFP-PhlF^{AM}. Points represent a single replicates collected on a single day.