OPEN ACCESS   SOURCE DATA   TRANSPARENT PROCESS

molecular systems biology

REPORT

Synthetic mammalian transgene negative autoregulation

Vinay Shimoga[1,2,5], Jacob T White[1,2,5], Yi Li[1,2], Eduardo Sontag[3] and Leonidas Bleris[1,2,4,*]

[1] Bioengineering Department, The University of Texas at Dallas, Richardson, TX, USA, [2] Center for Systems Biology, The University of Texas at Dallas, Richardson, TX, USA, [3] Department of Mathematics, Rutgers University, New Brunswick, NJ, USA and [4] Electrical Engineering Department, The University of Texas at Dallas, Richardson, TX, USA
[5]These authors contributed equally to this work
* Corresponding author. Electrical Engineering, Bioengineering Department, The University of Texas at Dallas, 800 West Campbell Road, NSERL 4.708, Richardson, TX 75080, USA. Tel.: + 1 972 883 5785; Fax: + 1 972 883 5785; E-mail: bleris@utdallas.edu

Biological networks contain overrepresented small-scale topologies, typically called motifs. A frequently appearing motif is the transcriptional negative-feedback loop, where a gene product represses its own transcription. Here, using synthetic circuits stably integrated in human kidney cells, we study the effect of negative-feedback regulation on cell-wide (extrinsic) and gene-specific (intrinsic) sources of uncertainty. We develop a theoretical approach to extract the two noise components from experiments and show that negative feedback results in significant total noise reduction by reducing extrinsic noise while marginally increasing intrinsic noise. We compare the results to simple negative regulation, where a constitutively transcribed transcription factor represses a reporter protein. We observe that the control architecture also reduces the extrinsic noise but results in substantially higher intrinsic fluctuations. We conclude that negative feedback is the most efficient way to mitigate the effects of extrinsic fluctuations by a sole regulatory wiring.
Molecular Systems Biology **9**: 670; published online 4 June 2013; doi:10.1038/msb.2013.27
Subject Categories: metabolic and regulatory networks; synthetic biology
Keywords: cellular noise; human cells; negative feedback; transgenes

Introduction

Information in cells propagates through intricate, diverse biochemical pathways. Within these complex networks certain small-scale interaction patterns appear more frequently than others (Milo *et al*, 2002; Alon, 2006, 2007). The convergence of pathways to particular motifs may be attributed to their inherent topological and functional properties, with a range of theoretical and experimental results supporting this hypothesis.

A network motif of particular interest is negative feedback, which appears in high frequency in bacterial (Alon, 2006), yeast (Lee *et al*, 2007), and mammalian cells (Odom *et al*, 2006). The negative-feedback loop consists of a single node that represses its own synthesis, and has been shown to accelerate transcriptional response time (Rosenfeld *et al*, 2002) and reduce gene expression noise in bacteria and yeast (Thattai and van Oudenaarden, 2001; Becskei and Serrano, 2002; Dublanche *et al*, 2006; Nevozhay *et al*, 2009). Remarkably, theoretical and experimental results show that negative feedback might either amplify or reduce noise in gene expression (Thattai and van Oudenaarden, 2001; Simpson *et al*, 2003; Austin *et al*, 2006; Cox *et al*, 2008; Nacher and

Ochiai, 2008; Singh and Hespanha, 2009; Marquez-Lago and Stelling, 2010), highlighting the need for additional experimental investigation, particularly in human cells.

Even though endogenous motifs are composed of relatively few elements, they are typically embedded as 'modules' in larger networks that exhibit complex behavior. Therefore, synthetic gene circuits, orthogonal to endogenous cellular processes, are a suitable experimental platform for elucidating their topological and functional properties.

Events controlling synthesis and degradation are independent for different proteins in a cell, and are often called 'intrinsic' or 'local' noise (Elowitz *et al*, 2002; Blake *et al*, 2003; Paulsson, 2004; Raser and O'Shea, 2005). A strongly expressing constitutive promoter is expected to have little intrinsic noise, while a weak promoter will have high intrinsic noise (Bar-Even *et al*, 2006; Newman *et al*, 2006). These variations propagate along pathways, with the consequence that protein distributions along a pathway appear correlated (Swain *et al*, 2002; Volfson *et al*, 2006). However, even proteins from different regulation pathways may show correlation, owing to stochastic variations in quantities that affect the regulation of all genes (Swain *et al*, 2002; Volfson *et al*, 2006), such as polymerase copies. As a consequence, two identical, independently

regulated promoters are expected to have the same extrinsic noise, which originates through global effects (Elowitz *et al*, 2002; Swain *et al*, 2002; Raser and O'Shea, 2005).

A two-reporter experimental platform (Elowitz *et al*, 2002) has been instrumental for studying extrinsic and intrinsic noise as well as pathway-specific effects (Colman-Lerner *et al*, 2005; Pedraza and van Oudenaarden, 2005). Results in *Escherichia coli* and yeast cells show that extrinsic noise dominates the total noise (Colman-Lerner *et al*, 2005), especially at high protein abundances (Taniguchi *et al*, 2010), whereas fluctuations at low-protein copy numbers are owing to intrinsic noise. A limitation of this experimental platform is the requirement for two identically regulated reporters with equal variances. To overcome this constraint, we develop an approach that permits intrinsic and extrinsic noise breakdown for two non-identical reporters, and use this methodology to study noise in synthetic mammalian transgene negative autoregulation.

Results

Integration of the circuits and initial characterization

We integrated the negative autoregulation and the control architectures (Figure 1) in Tet-On immortalized human kidney cells (Materials and methods, Generation of stable cell lines). We originally engineered (Bleris *et al*, 2011) these architectures using a bidirectional promoter that transcribes two genes fused in multiple cloning sites (MCSI and MCSII) upstream from minimal CMV promoters. The bidirectional promoter is activated in the presence of Doxycycline (Dox) by the transcription factor rtTA that, for all reported experiments, is produced stably from the cell line.

For the implementation of the control architecture (Figure 1A), we chose the transcriptional repressor LacI to be cloned in the MCSI and we fused sequences containing a tandem repeat of the wild-type LacO between the Pcmv region and the start codon of the dsRed monomer reporter gene (fused in the MCSII). We fused the zsGreen1 protein upstream from LacI, using an internal ribosome entry site (IRES), a nucleotide sequence that allows for translation initiation in the middle of a messenger RNA (mRNA) sequence. We constructed the transcriptional negative autoregulatory motif by inverting the promoter region of the control motif copying the wild-type LacO sequence (Figure 1B). As a result, the LacI protein inhibits the transcription of its own mRNA (and the co-expressed zsGreen1), and the constitutive output is now measured by the dsRed fluorescent. The negative-feedback strength can be tuned by IPTG induction.

Subsequently, we integrated the circuits stably in cells (Materials and methods, Generation of stable cell lines). For fixed Dox in a monoclonal cell population, the architecture depicted in Figure 1A is a simple negative regulation and serves as the control. We note that by changing the Dox levels, the output of the control architecture (Figure 1A) will depend on both X (rtTA) and Y (LacI) thereby emulating a Type I incoherent feedforward architecture (Alon, 2006). Previously (Bleris *et al*, 2011), we transiently transfected plasmids carrying these circuits and studied the behavior of the reporter proteins. Our experiments showed that the output node of an incoherent feedforward motif is largely invariant to the changes in the DNA fragment (i.e., primarily the copy number).

We first induced the stable clones with a wide range of IPTG and Dox concentrations, and the output was quantified after 24 h using flow cytometry (Materials and methods, Data processing). A gate based on the forward and side scatter is
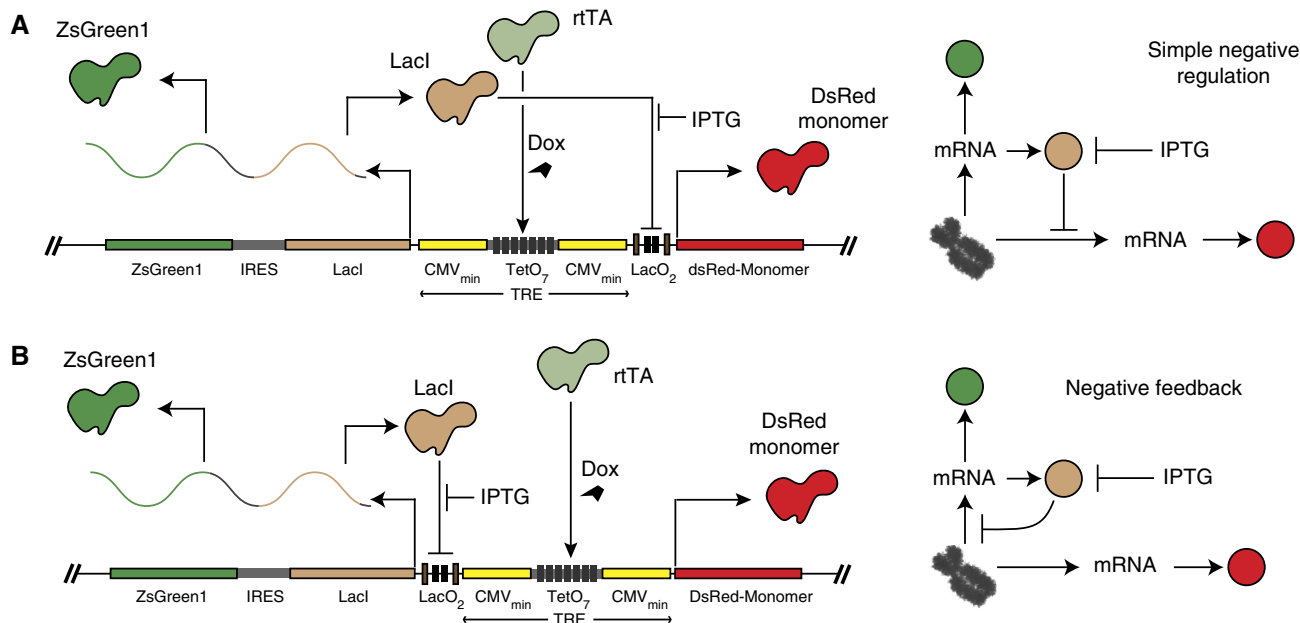
Figure 1 The synthetic architectures integrated in human kidney cells. (**A**) The control architecture: the bidirectional promoter under the control of rtTA transcribes the ZsGreen1–IRES–LacI and dsRed monomer transcripts. The dsRed monomer is inhibited by LacI. (**B**) The negative feedback: The bidirectional promoter under the control of rtTA transcribes the ZsGreen1–IRES–LacI and dsRed monomer transcripts. The ZsGreen1–IRES–LacI transcript is inhibited by LacI.

first used to select single cell events followed by a gate that retains the constitutive protein-positive events (i.e., dsRed for the negative feedback and zsGreen1 for the control architecture) at the threshold of negative cells (Supplementary Figure 1).
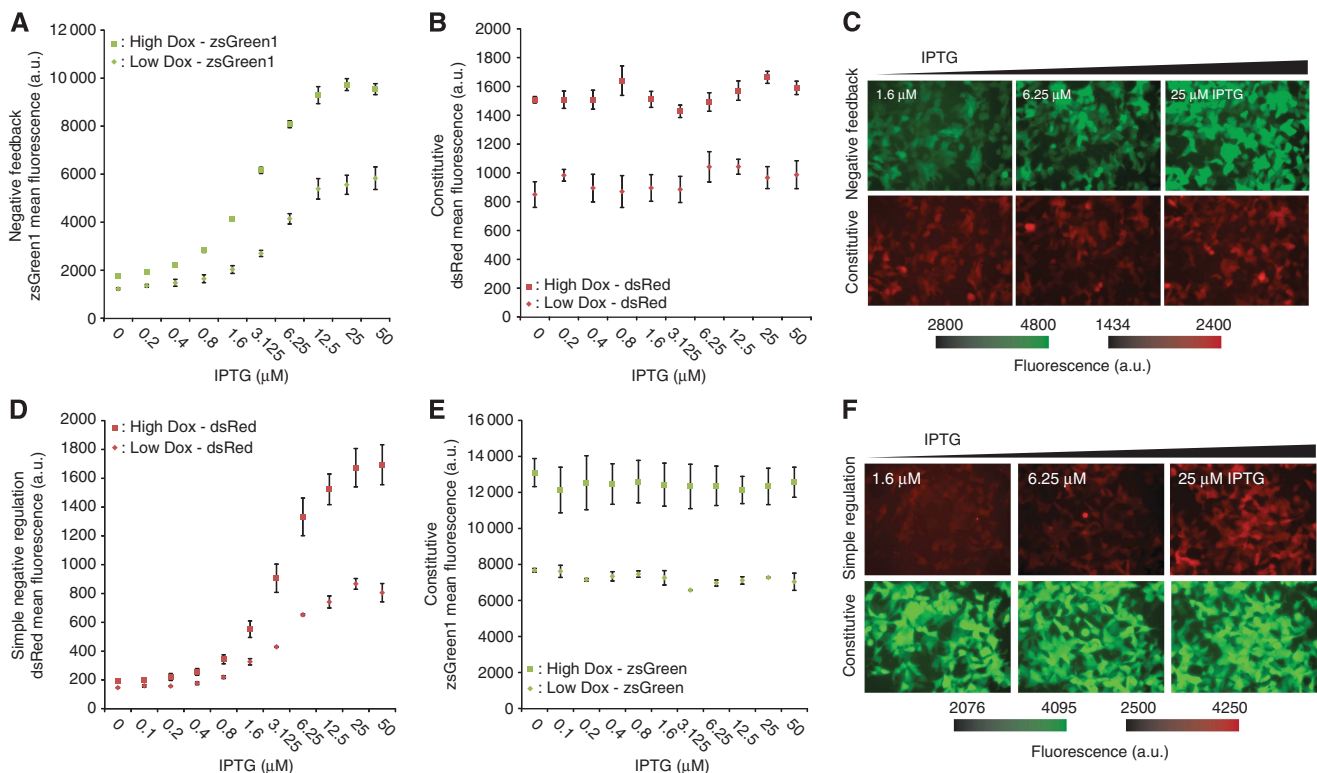
A Dox titration of the negative-feedback architecture shows that the mean protein concentration increases for both the constitutive side, dsRed, and the side controlled by negative feedback. In the case of dsRed, we have a three-fold increase, while for zsGreen1 we have a two-fold increase in the mean protein concentration (Supplementary Figure 2a). For the control architecture, the mean protein concentration increased by five-fold the constitutive side (zsGreen1) and 2.5-fold the regulated side, dsRed (Supplementary Figure 2b). These results are in agreement with our transient transfection experiments (Bleris *et al*, 2011), and show that the incoherent feedforward architecture is superior to negative feedback in controlling the output mean levels due to changes in the input.

We then performed IPTG titrations (0.0625–25 µM) at two Dox concentrations, 625 ng/ml (defined as 'low') and 5000 ng/ml (defined as 'high'). The Dox values were selected in order to study the effect of the expression level on the total noise, and to provide sufficient separation of the mean output levels (Supplementary Figure 3). Importantly, the ability to control the transcription levels of the transgene allows us to emulate the effect of the variable transcriptional activity

expected at different genomic locations. Consistently with our experiments (Supplementary Figure 3), previous experimental (Bar-Even *et al*, 2006) and theoretical (Paulsson, 2004) studies show that noise scales inversely with the protein abundance.

For the negative feedback, the titrations performed at low and high concentrations of Dox show a corresponding five-fold increase for the high Dox case and a three-fold increase in the low Dox case in mean fluorescence of zsGreen1 (Figure 2A). As expected, the dsRed protein levels remains constant over the entire range of IPTG concentrations (Figure 2B). We provide selected microscopy snapshots of the induced negative-feedback populations in Figure 2C. The IPTG titrations of the simple negative regulation clone result to a four-fold and eight-fold increase in dsRed protein levels at low and high Dox concentrations, respectively, whereas the constitutively synthesized protein levels were unchanged (Figure 2D and E, respectively). We provide selected microscopy snapshots of the induced simple regulation populations in Figure 2F. We also include all the flow cytometry data and the corresponding histograms for the IPTG titrations for both transgenes and two Dox conditions (Supplementary Figures 4–7).

To further probe the behavior of the circuits, we quantified the number of copies of integrations for our circuits using real-time quantitative PCR (Materials and methods, and Supplementary Material, Integration Copies). We found that

Figure 2 IPTG titrations for the negative feedback and simple regulation transgenes. (**A–C**) The negative feedback: zsGreen protein in green (under negative autoregulation) and dsRed in red (constitutively synthesized). Saturated Dox concentration with squares and low DOX concentration with diamonds. Error bars show the s.d. of triplicate experiments. (**A**) The mean zsGreen fluorescence. (**B**) The mean dsRed fluorescence. (**C**) Microscopy images of IPTG titrations at high concentration of Dox. (**D–F**) Simple regulation architecture: the dsRed in red (under regulation) and the zsGreen protein in green (constitutively synthesized); saturated Dox concentration with squares and low DOX concentration with diamonds. Error bars show the s.d. of triplicate experiments. (**D**) The mean dsRed fluorescence. (**E**) The mean zsGreen fluorescence. (**F**) Microscopy images of IPTG titrations at high concentration of Dox. Source data for this figure is available on the online supplementary information page.

the negative-feedback clone is a single integration while the simple negative regulation is two copies (Supplementary Material, Table I). We created a new transgene of the simple negative regulation with a single integration and showed that their behavior is consistent (Supplementary Figure 8).

Analysis of noise

After the initial characterization of the transgenes, our next objective was to extract the two noise components from the experimental data. The total noise observed in a fluorescent reporter distribution arises through the combination of global (extrinsic) fluctuations together with the fluctuations in that protein's local regulation machinery (intrinsic). In the standard two-reporter formulation (Elowitz *et al*, 2002), the extrinsic noise becomes the normalized covariance of the two reporters that are independently regulated and identically distributed. We extend this analysis to cases where it is not feasible to construct two identically regulated reporters or it is impossible to obtain identical reporter statistics.

We use a multiplicative noise model (Supplementary Material, Theory), where the total fluctuations of one reporter are the product of an extrinsic random variable and intrinsic random variable, while the second reporter fluctuations are the product of the same extrinsic random variable but its own intrinsic random variable. The three-model components are assumed independent. As described in detail in the Supplementary Material, we convert this multiplicative model to a linear model and show that the extrinsic noise is the normalized covariance of two constitutive reporters and the intrinsic noise is the difference between the observed CV square and the extrinsic CV square. Importantly, to decouple the extrinsic noise of the regulated and constitutive reporters, we add the sensitivity coefficient α to the extrinsic random variable for the regulated reporter.

We define as Y the constitutive reporter, X the regulated reporter (controlled by an inducer, in our case IPTG), and α is the coefficient that is 1 for two constitutive promoters with identical reporter statistics but varies depending on the regulation of X. Considering the case of negative feedback, when LacI is fully inactivated by IPTG at saturation, both reporters are equally sensitive to extrinsic fluctuations; thereby the extrinsic noise will be the normalized covariance as defined previously (Elowitz *et al*, 2002). We can safely postulate that the unregulated reporter should have equal sensitivity to extrinsic fluctuations for all IPTG conditions. Consequently, we calculate the regulated reporter sensitivity coefficient α for each sample in the IPTG titration such that the extrinsic noise of the unregulated reporter is the same as in the full IPTG well (Supplementary Material, Table II). In summary, we use the following noise breakdown:

$$n_{\text{extX}}^2 = \alpha_i \text{Cov}(X_i, Y_i) \quad (1)$$

$$n_{\text{intX}}^2 = n_{\text{totX}}^2 - n_{\text{extX}}^2 \quad (2)$$

$$n_{\text{extY}}^2 = \text{Cov}(X_i, Y_i)/\alpha_i \quad (3)$$

$$n_{\text{intY}}^2 = n_{\text{totY}}^2 - n_{\text{extY}}^2 \quad (4)$$

$$\alpha_N = \frac{n_{\text{totX}}}{n_{\text{totY}}} \quad (5)$$

$$\alpha_i = \frac{\text{Cov}(X_i, Y_i)\alpha_N}{\text{Cov}(X_N, Y_N)}, i \neq N \quad (6)$$

where n_{totX}^2 and n_{totY}^2 are experimentally determined CV squares of reporters X and Y, $\text{Cov}(X_i, Y_i)$ is the covariance of the logarithms of X and Y, and the inducer concentrations are indexed $i = 1, 2, \dots, N$, where well N is fully induced.

The effect of negative-feedback regulation on noise

We next obtain the intrinsic, extrinsic, and total noise of the two architectures (Figure 1) for the IPTG titrations of Figure 2. The negative-feedback architecture experiments show that stronger feedback (i.e., low IPTG concentration) reduces extrinsic noise (and the total noise) but mildly increases intrinsic noise (Figure 3A and D for high and low Dox, respectively). As expected, the noise levels remain flat for the control protein output (Figure 3C and F), with the extrinsic noise significantly higher than the intrinsic noise. In addition, for the negative-feedback architecture, when comparing the high and low Dox cases (Figure 3A and D), we observe that the strong induction leads to lower total noise. We also plot the noise versus the mean levels of the output protein, to probe directly the impact of negative feedback. Indeed, as illustrated in Figures 3B and E the negative feedback reduces noise. In contrast, the simple negative regulation remains in the same range for decreasing mean (Figure 3H).

When we examine the noise breakdown of the simple negative regulation architecture, we observe that the noise scales with protein abundance (Figure 3G for high Dox and Supplementary Figure 9 for low Dox). The effect is not as pronounced as with the negative feedback but we partially attribute this to a post-processing of the data, performed in order to discard the portion of the events that merge with the background signal (Supplementary Figures 6–7). The post-processing is particularly necessary for the low Dox case and although qualitatively consistent these results are not taken into consideration (Supplementary Figures 7 and 9). We emphasize that for the negative autoregulation there is no post-processing beyond the constitutive protein gating (Supplementary Figures 4–5).

For the simple negative regulation architecture, the results show that stronger regulation (i.e., low IPTG concentration) reduces extrinsic noise but significantly, as compared with the negative feedback, increases intrinsic noise (Figure 3G), resulting to approximately flat total noise. As expected, the control constitutive protein noise breakdown remains flat for high Dox (Figure 3I).

In order to validate our noise decomposition, we performed alternative analyses of the raw experimental data and simulations. First, we used a method to filter out extrinsic noise, simply by processing the flow cytometry data using a smaller forward versus side scatter gate (Newman *et al*, 2006). This approach has been reported to filter out extrinsic noise in
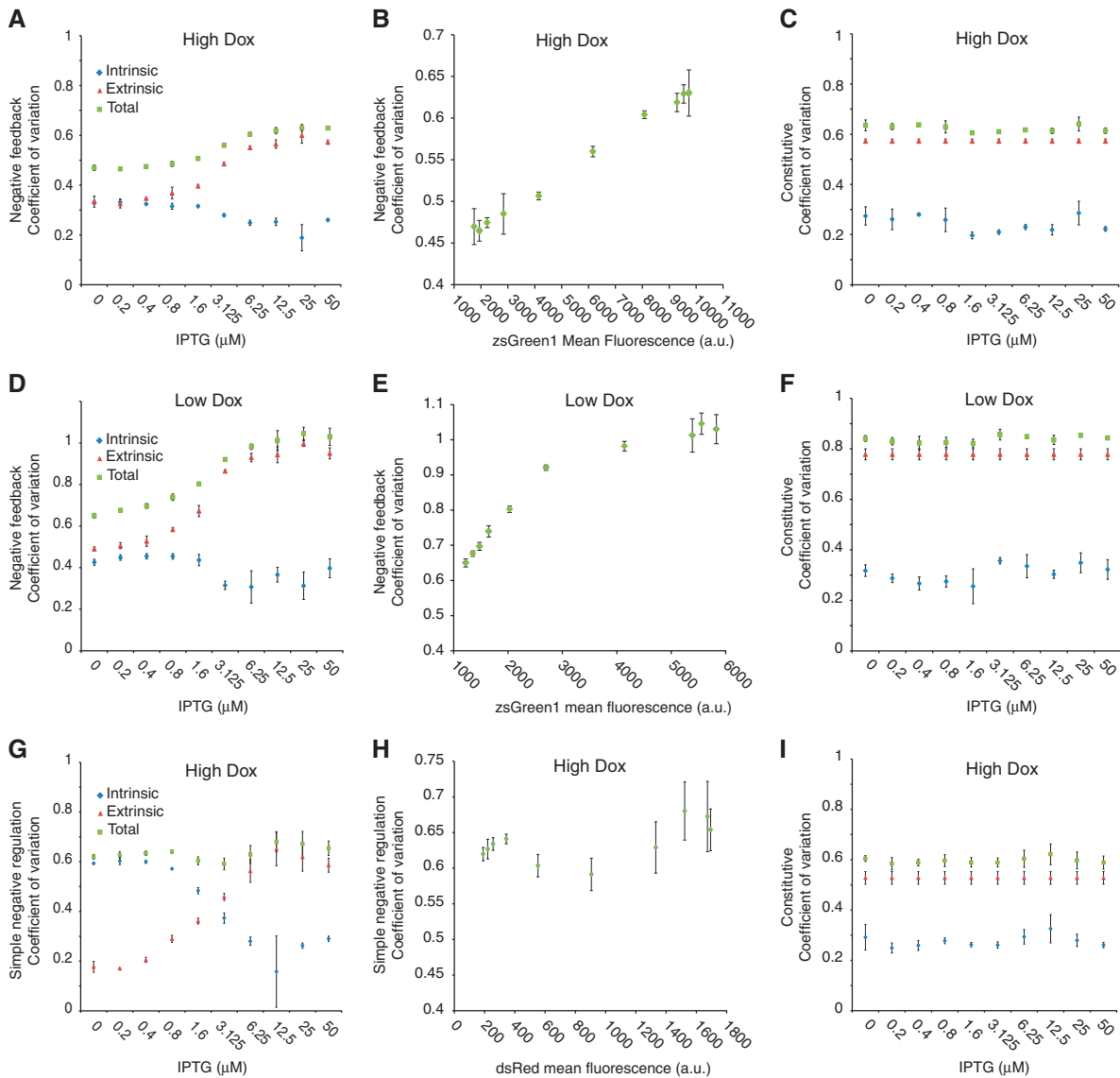
Figure 3 Noise for the negative-feedback loop and the simple regulation transgenes. Intrinsic noise with blue diamonds, extrinsic noise with red triangles, and total noise with green squares. Error bars are the s.d. of triplicates. (**A–F**) Negative-feedback architecture: (**A**) coefficient of variation for high Dox. (**B**) Coefficient of variation versus mean zsGreen1 protein levels for high Dox. (**C**) Coefficient of variation for the constitutive dsRed protein for high Dox. (**D**) Coefficient of variation for low Dox. (**E**) Coefficient of variation versus mean zsGreen1 protein levels for low Dox. (**F**) Coefficient of variation for the constitutive dsRed protein for low Dox. (**G–I**) Simple regulation architecture: (**G**) coefficient of variation for the dsRed protein for high Dox. (**H**) Coefficient of variation versus mean dsRed protein levels for high Dox. (**I**) Coefficient of variation for the constitutive zsGreen1 protein for high Dox. Source data for this figure is available on the online supplementary information page.

mammalian cells (Singh *et al*, 2010). Indeed, we find (Supplementary Figure 10) that reducing the gate decreased the total noise due to a decrease in the extrinsic noise, while the intrinsic noise remains the same. Furthermore, the overall trend of the drop in noise with decreasing concentrations of IPTG did not change. We also used simulations to gain additional insight into the ways our method is able to decompose noise. Specifically, as discussed in the Supplementary Material Theory section, we first vary the strength of transcription of a single bidirectional promoter coding for two fluorescent proteins, leading to perfectly correlated fluorescence quantities, which our decomposition shows to have only extrinsic noise and no intrinsic noise. Next, we vary the transcriptional activity of two fluorescent genes

independently, which leads to uncorrelated fluorescence quantities; our method returns only intrinsic noise and no extrinsic noise. Subsequently, we show that the decomposition is correct for mixtures of intrinsic and extrinsic noise.

Finally, in order to further validate the consistency of our results, we examined the relationship between the chromosomal position of our transgene and the experimentally observed phenotype. We created new integrations for the negative feedback, selected three random colonies, and we performed titrations of IPTG for high Dox levels. A direct comparison between the main and new clones (Supplementary Figure 11) shows that the overall behavior is conserved with the absolute noise levels being marginally different. This difference is expected considering that distant

chromosomal sites often have significant differences in their transcriptional activity (Dar *et al*, 2012).

Discussion

Investigating the relationship between regulatory systems and cellular noise has inherently wide biological significance. Our results shed new light on one of the most abundant biological motifs, the negative-feedback loop. In particular, we used synthetic circuits stably integrated in human kidney cells and we studied the effect of negative feedback on cell-wide and gene-specific sources of uncertainty.

We developed an approach to extract the extrinsic and intrinsic noise contributions from experimental measurements of two reporter proteins that are controlled by non-identical promoters. Our experiments reveal that negative feedback reduces extrinsic noise while slightly increasing intrinsic noise. Importantly, negative feedback reduces the total noise. By comparing these results to simple negative regulation, we argue that negative feedback is the most efficient way to reduce extrinsic fluctuations by introducing a sole additional regulatory wiring.

It has been shown theoretically (Simpson *et al*, 2003) and confirmed experimentally (Austin *et al*, 2006) that negative autoregulation can filter out lower frequency noise. In addition, it has been shown experimentally (Rosenfeld *et al*, 2005) that extrinsic fluctuations have lower frequency components than intrinsic noise. It follows that negative autoregulation would remove extrinsic noise. Our results indeed show that negative autoregulation removes extrinsic noise; however, we also observe that simple negative regulation also removes extrinsic noise, de-correlating the negatively regulated reporter from the constitutive reporter. Time lapse experiments can be used to shed additional light to the properties of these architectures.

To conclude, our analysis shows that negative feedback reduces noise, but only from extrinsic sources, outside of the genetic components of the feedback loop. Furthermore, we show that the negative feedback raises intrinsic noise but the cost of this regulation is small when compared with simple negative regulation.

Materials and methods

Generation of stable cell lines

Tet-ON cells (Clontech), which stably express the transcription factor rtTA, were used for all the experiments. Cells were grown in 12-well plates (Greiner Bio-One) at 80% confluency and transfected (LTX Transfection reagent, Invitrogen) with the plasmid carrying the circuit and a Hygromycin linear selection marker (Clontech) in a 1:20 ratio. Cells were transferred to Petri dishes and incubated in 25 μg/ml hygromycin for a week. Hygromycin was then reduced to 15 μg/ml and cells were incubated until colonies developed. Colonies that were positive for florescent signals were selected by microscopy and picked using cloning rings and further expanded.

Cell culture

The cells were grown at 37°C and 5% CO_2. The cells were grown in Dulbecco's modified Eagle's medium (DMEM, Invitrogen, Cat # 11965-11810) supplemented with 0.1 mM MEM non-essential amino acids

(Invitrogen, Cat # 11140-050), 0.045 units/ml of penicillin and 0.045 μg/ml streptomycin (Penicillin–Streptomycin liquid, Invitrogen), and 10% fetal bovine serum (FBS, Invitrogen). The adherent culture was maintained in this medium by trypsinizing with Trypsin–EDTA (0.25% Trypsin with EDTAx4Na, Invitrogen) and diluting in a fresh medium upon reaching 50–90% confluence.

Flow cytometry

The cells were prepared for flow cytometry by trypsinizing each well with 0.5 ml 0.25% trypsin–EDTA, collecting the cell suspension, and centrifuging at 4000 r.p.m. for 2 min. Trypsin was removed and the pellet resuspended by short vortexing in 0.5 ml PBS buffer (Invitrogen). Cells were run on a LSR Fortessa (BD Biosciences) Flow cytometer equipped with the FACSDiva software program. One hundred thousand cells were counted in each run. DsRed was measured with a 561 nm laser and a 586 nm emission filter with a 582/15 band pass filter, and ZsGreen1 with a 488 nm laser and a 509 nm emission filter with a 515/20 band pass filter. Data analysis was performed using FlowJo and Matlab.

Data processing

First, we use the constitutive fluorescence protein-positive population of cells, based on a control experiment in which cells are uninduced (Supplementary Figure 1). Second, to remove outliers (explained in the supplement), we include cells that have a fluorescent intensity equal to 2.5 times the s.d. of the population.

Microscopy

All microscope images were taken from live cells grown in multi-well plates (Greiner Bio-One) in the DMEM supplemented with non-essential amino acids, penicillin/streptomycin, and 10% FBS. Cells were imaged using the Olympus IX81 microscope and a Precision Control environmental chamber. The images were captured using a Hamamatsu ORCA–03 Cooled monochrome digital camera. The filter sets (Chroma) are as follows: ET470/50x (excitation) and ET525/50m (emission) for ZsGreen1, ET560/40x (excitation) and ET630/75m (emission) for DsRed. For the negative feedback, the exposure times were ZsGreen1:500 ms and dsRed:1000 ms, while for the simple negative regulation, the exposure times were ZsGreen1:400 ms and dsRed:2000 ms. Data collection and processing was performed in the software package Slidebook 5.0 and Adobe Illustrator.

Copy number using real-time PCR

We performed real-time quantitative PCR to determine the absolute copies of integration for our circuits. The average copy numbers of dsRed of all stable clones were estimated by the delta delta Ct method as follows: $2^{-\Delta\Delta Ct} = ((1 + E_{DsRED})^{-\Delta Ct, DsRED})/((1 + E_{BRCA1})^{-\Delta Ct, BRCA1})$, where E_{DsRED} is the PCR amplification efficiency for dsRed and E_{BRCA1} for BRCA1 (endogenous reference gene)(Zheng *et al*, 2011). A control stable HEK293 cell line was generated by Flp-In system (Invitrogen) and contains a single copy of dsRed transgene (Li *et al*, 2012). To determine the PCR amplification efficiency, genomic DNAs from the control cell line were used to generate the dilution curve of $\log_2$(DNA amount, ng) versus Ct. E_{DsRED} was calculated as 1.07, and E_{BRCA1} as 0.98. For each stable clone, triplicates (50 ng of genomic DNA) were performed and the average copy numbers were calculated as the mean ± s.d.

Supplementary information

Supplementary information is available at the *Molecular Systems Biology* website (www.nature.com/msb).

Acknowledgements

Conflict of interest

The authors declare that they have no conflict of interest.

References

Alon U (2007) *An Introduction to Systems Biology: Design Principles of Biological Circuits*. Boca Raton, FL, USA: Chapman & Hall

Alon U (2007) Network motifs: theory and experimental approaches. *Nat Rev Genet* **8:** 450–461

Austin DW, Allen MS, McCollum JM, Dar RD, Wilgus JR, Sayler GS, Samatova NF, Cox CD, Simpson ML (2006) Gene network shaping of inherent noise spectra. *Nature* **439:** 608–611

Bar-Even A, Paulsson J, Maheshri N, Carmi M, O'Shea E, Pilpel Y, Barkai N (2006) Noise in protein expression scales with natural protein abundance. *Nat Genet* **38:** 636–643

Becskei A, Serrano L (2002) Engineering stability in gene networks by autoregulation. *Nature* **405:** 590–593

Blake WJ, Kaern M, Cantor CR, Collins JJ (2003) Noise in eukaryotic gene expression. *Nature* **422:** 633–637

Bleris L, Xie Z, Glass D, Adadey A, Sontag E, Benenson Y (2011) Synthetic incoherent feedforward circuits show adaptation to the amount of their genetic template. *Mol Syst Biol* **7:** 519

Colman-Lerner A, Gordon A, Serra E, Chin T, Resnekov O, Endy D, Pesce CG, Brent R (2005) Regulated cell-to-cell variation in a cell-fate decision system. *Nature* **437:** 699–706

Cox CD, McCollum JM, Allen MS, Dar RD, Simpson ML (2008) Using noise to probe and characterize gene circuits. *Proc Natl Acad Sci USA* **105:** 10809–10814

Dar RD, Razooky BS, Singh A, Trimeloni TV, McCollum JM, Cox CD, Simpson ML, Weinberger LS (2012) Transcriptional burst frequency and burst size are equally modulated across the human genome. *Proc Natl Acad Sci USA* **109:** 17454–17459

Dublanche Y, Michalodimitrakis K, Kummerer N, Foglierini M, Serrano L (2006) Noise in transcription negative feedback loops: simulation and experimental analysis. *Mol Syst Biol* **2:** 41

Elowitz M, Levine A, Siggia E, Swain P (2002) Stochastic gene expression in a single cell. *Science* **297:** 1183–1186

Lee T, Rinaldi NJ, Robert F, Odom DT, Bar-Joseph Z, Gerber GK, Hannett NM, Harbison CT, Thompson CM, Simon I, Zeitlinger J, Jennings EG, Murray HL, Gordon DB, Ren B, Wyrick JJ, Tagne JB, Volkert TL, Fraenkel E, Gifford DK *et al* (2007) Transcriptional regulatory networks in *Saccharomyces cerevisiae*. *Science* **298:** 799–804

Li Y, Moore R, Guinn M, Bleris L (2012) Transcription activator-like effector hybrids for conditional control and rewiring of chromosomal transgene expression. *Sci Rep* **2:** 897

Marquez-Lago TT, Stelling J (2010) Counter-intuitive stochastic behavior of simple gene circuits with negative feedback. *Biophys J* **98:** 1742–1750

Milo R, Shen-Orr S, Itzkovitz S, Kashtan N, Chklovskii D, Alon U (2002) Network motifs: simple building blocks of complex networks. *Science* **298:** 824–827

Nacher JC, Ochiai T (2008) Transcription and noise in negative feedback loops. *Biosystems* **91:** 76–82

Nevozhay D, Adams RM, Murphy KF, Josić K, Balázsi G (2009) Negative autoregulation linearizes the dose–response and suppresses the heterogeneity of gene expression. *Proc Natl Acad Sci USA* **106:** 5123–5128

Newman JRS, Ghaemmaghami S, Ihmels J, Breslow DK, Noble M, DeRisi JL, Weissman JS (2006) Single-cell proteomic analysis of S. cerevisiae reveals the architecture of biological noise. *Nature* **441:** 840–846

Odom DT, Dowell RD, Jacobsen ES, Nekludova L, Rolfe PA, Danford TW, Gifford DK, Fraenkel E, Bell GI, Young RA (2006) Core transcriptional regulatory circuitry in human hepatocytes. *Mol Syst Biol* **2:** 0017

Paulsson J (2004) Summing up the noise in gene networks. *Nature* **427:** 415–418

Pedraza J, van Oudenaarden A (2005) Noise propagation in gene networks. *Science* **307:** 1965–1969

Raser JM, O'Shea EK (2005) Noise in gene expression: origins, consequences, and control. *Science* **309:** 2010–2013

Rosenfeld N, Elowitz MB, Alon U (2002) Negative autoregulation speeds the response times of transcription networks. *J Mol Biol* **323:** 785–793

Rosenfeld N, Young J, Alon U, Swain P, Elowitz M (2005) Gene regulation at the single-cell level. *Science* **307:** 1962–1965

Simpson ML, Cox CD, Sayler GS (2003) Frequency domain analysis of noise in autoregulated gene circuits. *Proc Natl Acad Sci USA* **100:** 4551–4556

Singh A, Hespanha JP (2009) Optimal feedback strength for noise suppression in autoregulatory gene networks. *Biophys J* **96:** 4013–4023

Singh A, Razooky B, Cox CD, Simpson ML, Weinberger LS (2010) Transcriptional bursting from the HIV-1 promoter is a significant source of stochastic noise in HIV-1 gene expression. *Biophys J* **98:** L32–L34

Swain PS, Elowitz MB, Siggia ED (2002) Intrinsic and extrinsic contributions to stochasticity in gene expression. *Proc Natl Acad Sci USA* **99:** 12795–12800

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

Thattai M, van Oudenaarden A (2001) Intrinsic noise in gene regulatory networks. *Proc Natl Acad Sci USA* **98:** 8614–8619

Volfson D, Marciniak J, Blake WJ, Ostroff N, Tsimring LS, Hasty J (2006) Origins of extrinsic variability in eukaryotic gene expression. *Nature* **439:** 861–864

Zheng S, Houseman EA, Morrison Z, Wrensch MR, Patoka JS, Ramos C, Haas-Kogan DA, McBride S, Marsit CJ, Christensen BC, Nelson HH, Stokoe D, Wiemels JL, Chang SM, Prados MD, Tihan T, Vandenberg SR, Kelsey KT, Berger MS, Wiencke JK (2011) DNA hypermethylation profiles associated with glioma subtypes and EZH2 and IGFBP2 mRNA expression. *Neuro-Oncology* **13:** 280–289

Supplementary Material
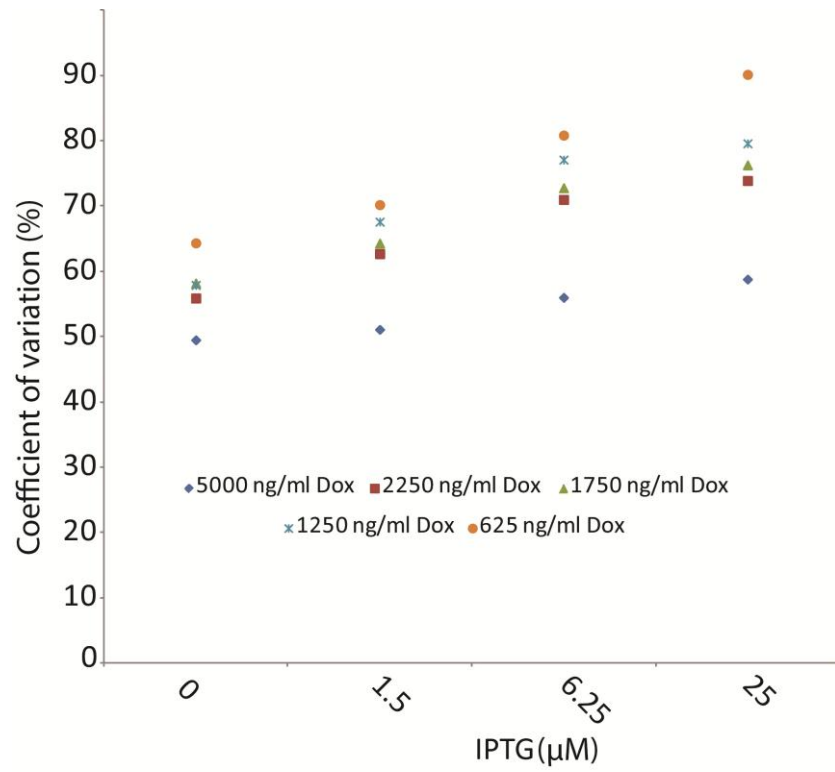
Supplementary Figures
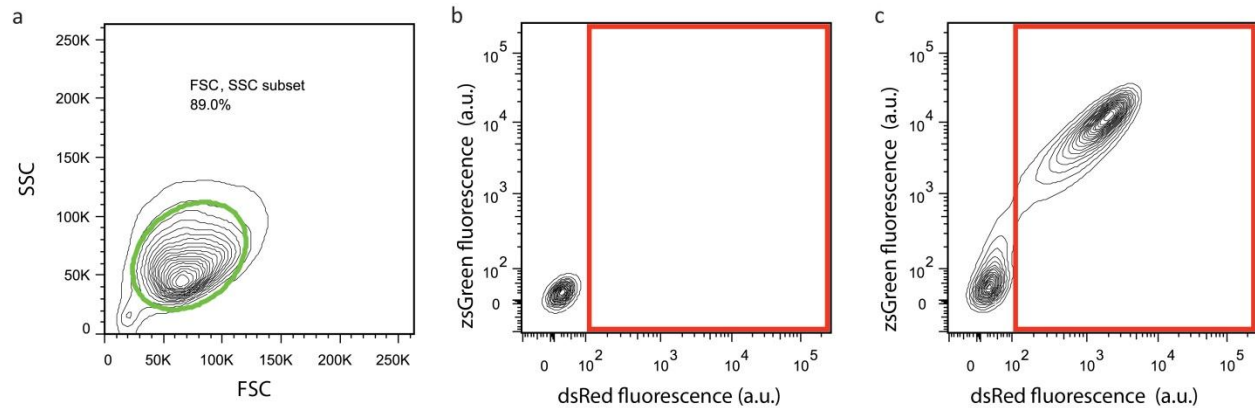
Supplementary Figure1. Flow cytometry data gating and processing. (a) A gate based on the forward and side scatter is first used to select single cell events. **(b)** The gated population is projected for dsRed and zsGreen. The particular case is a population of H293 cells without any fluorescence. Subsequently, we gated the constitutive protein positive events (e.g. dsRed for the negative feedback) at the threshold of negative cells**. (c)** The particular case is a population is the negative feedback. ■
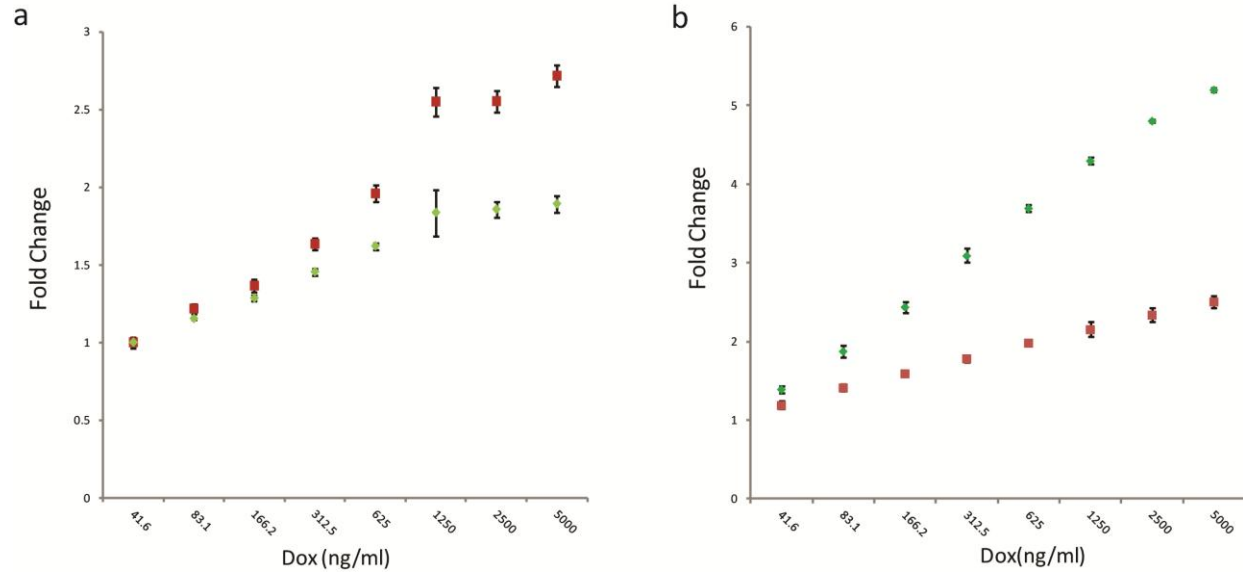
Supplementary Figure 2. Doxycycline titrations for the negative feedback loop and the control architecture. The zsGreen protein in green diamonds and the dsRed in red squares. Error bars show the standard deviation of triplicate experiments. **(a)** Absolute change of mean fluorescent levels of dsRed and zsGreen for the NFL. **(b)** Absolute change of mean fluorescent levels of dsRed and zsGreen for the control architecture.■

Supplementary Figure 3. Coefficient of variation versus IPTG concentration for the negative feedback loop for various concentrations of doxycycline. ∎
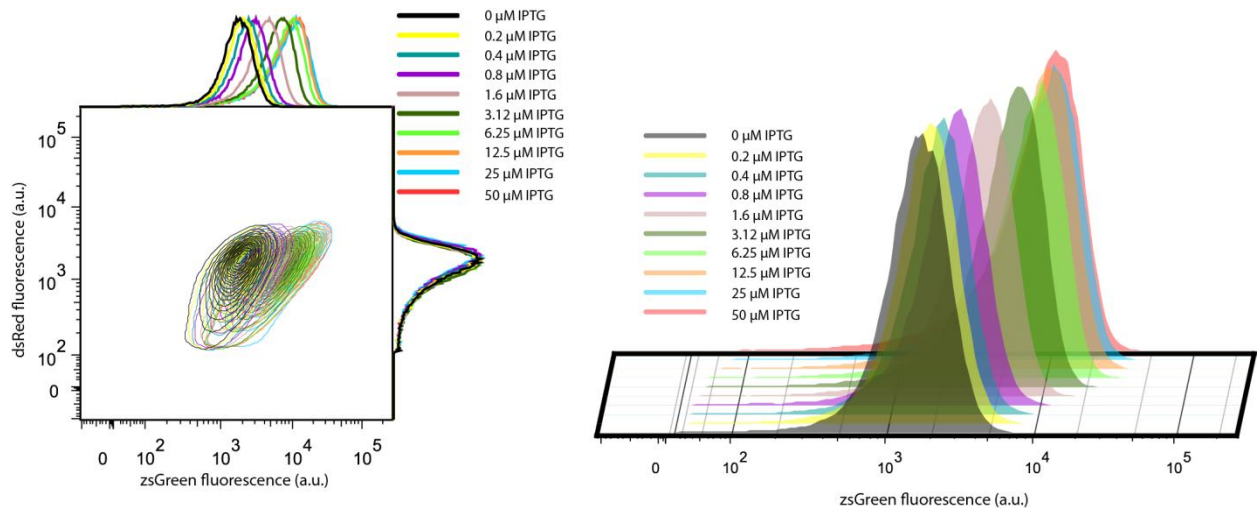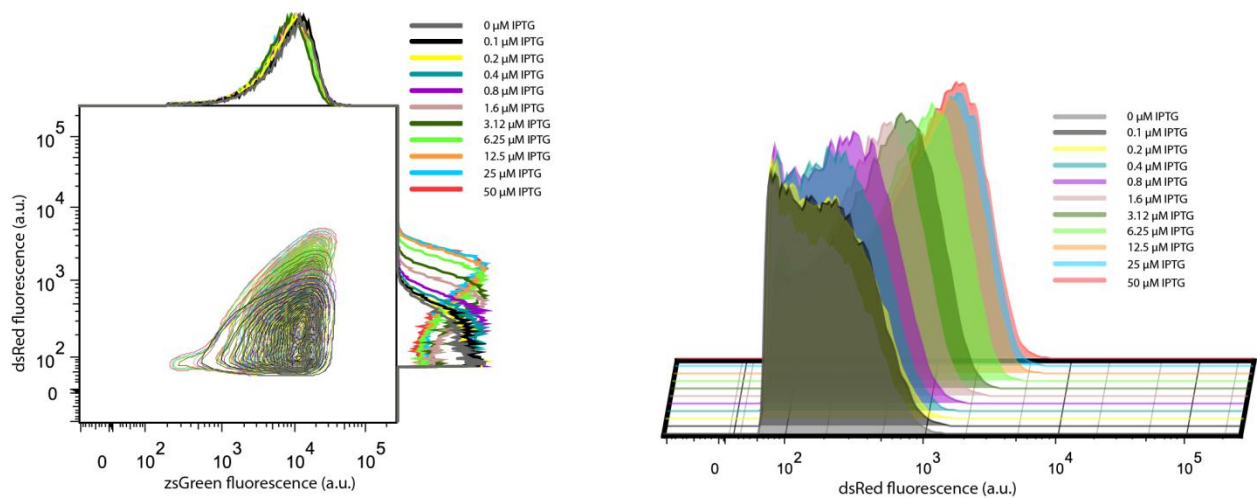
Supplementary Figure 4. IPTG Titrations for the negative feedback loop. The high Doxycycline case. DsRed positive cells are illustrated and the corresponding histograms for each output are also presented. ∎
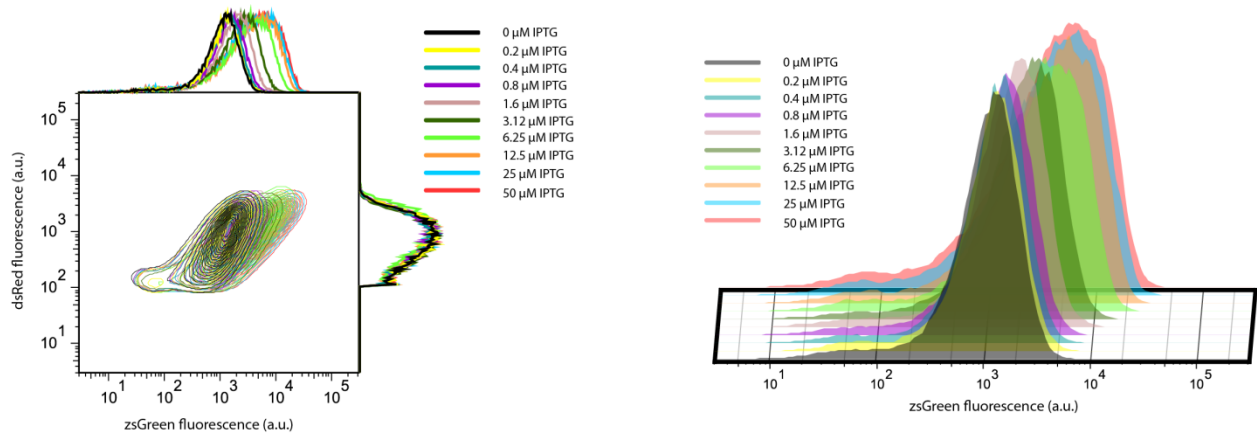

Supplementary Figure 5. IPTG Titrations for the negative feedback loop. The low Doxycycline case. DsRed positive cells are illustrated and the corresponding histograms for each output are also presented.
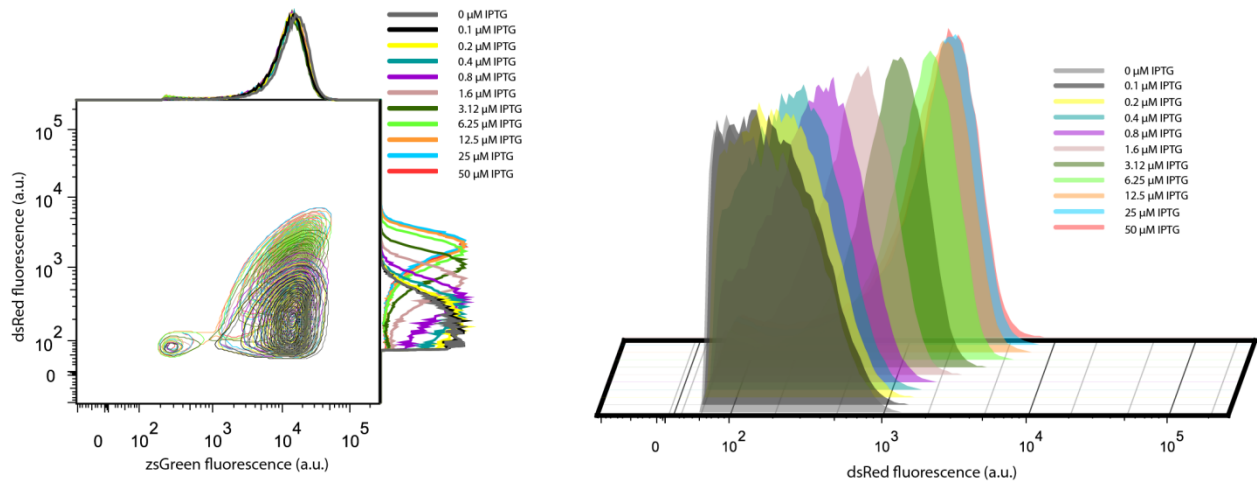
Supplementary Figure 6. IPTG Titrations for the cascade. The high Doxycycline case. DsRed positive cells are illustrated and the corresponding histograms for each output are also presented. ∎


Supplementary Figure 7. IPTG Titrations for the cascade. The low Doxycycline case. DsRed positive cells are illustrated and the corresponding histograms for each output are also presented. ∎
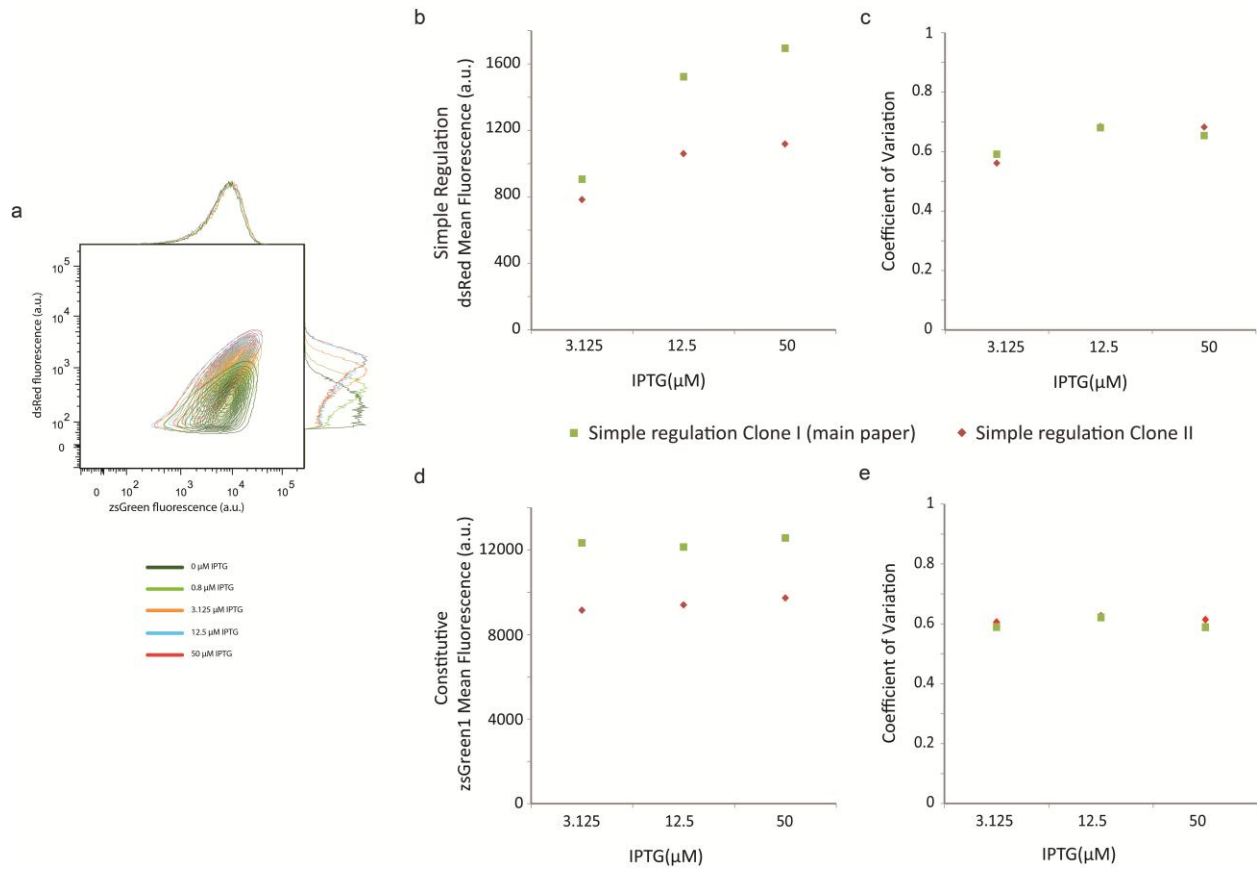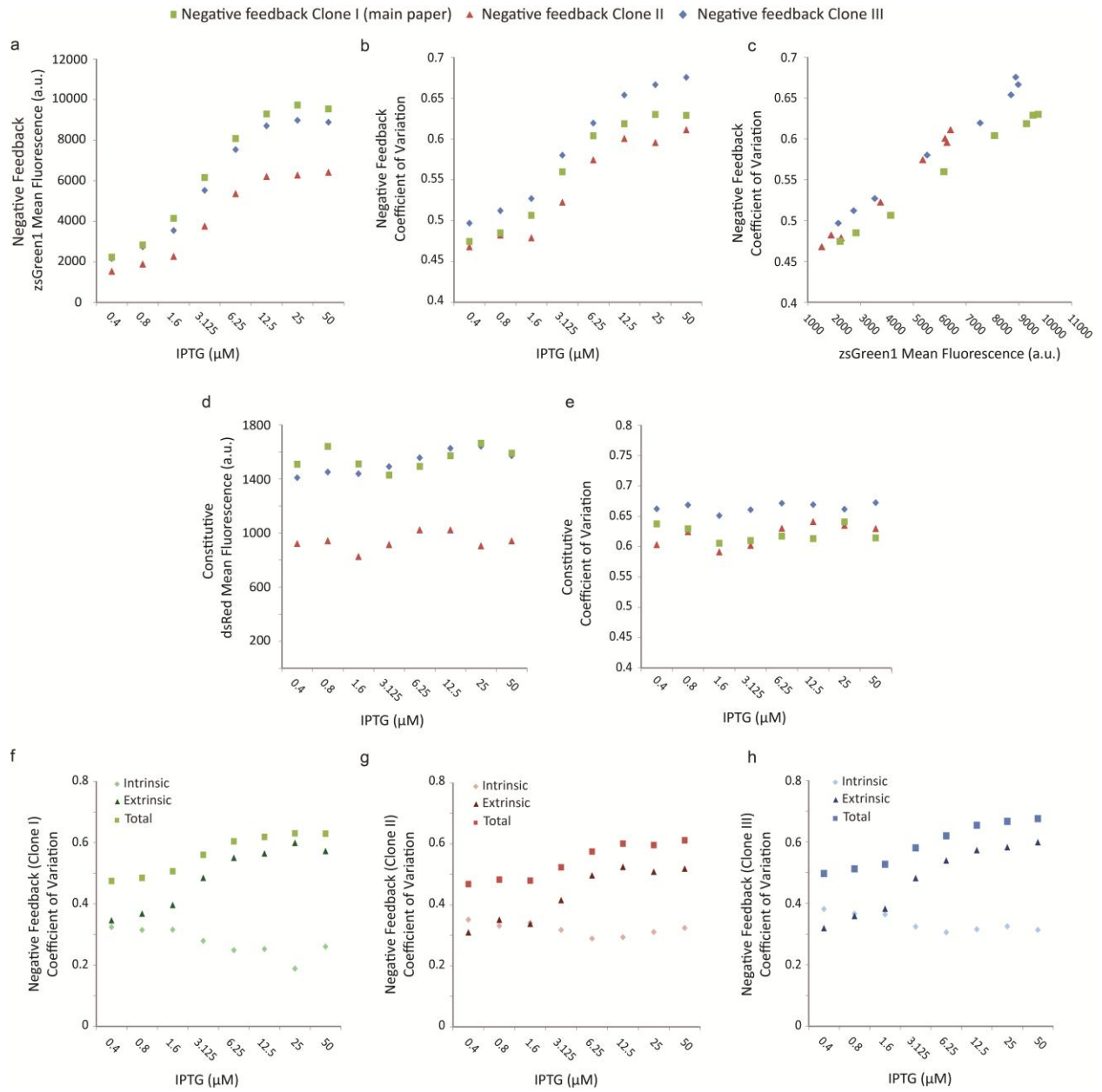
Supplementary Figure 8. Comparison of mean fluorescence and coefficient of variation between the main paper simple regulation clone and another simple regulation transgene.■
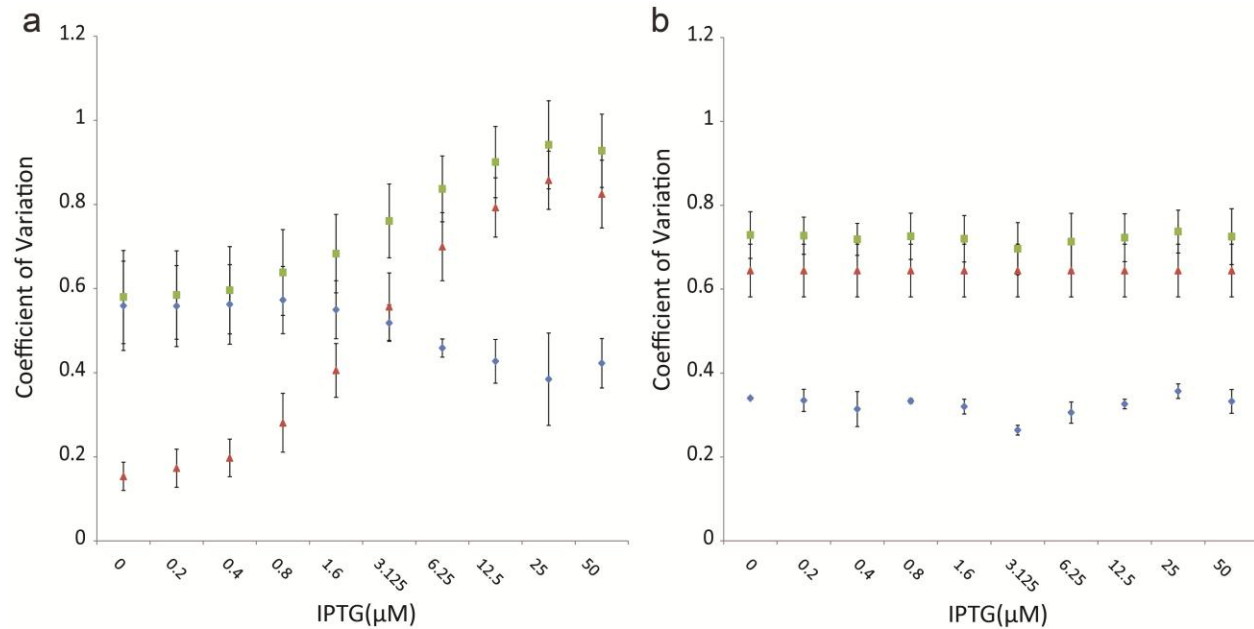
Supplementary Figure 9. Coefficient of variation for the control architecture. Local, global and total noise in gene expression of the control architecture. **(a)** Coefficient of variation of dsRed protein for low DOX, **(b)** Coefficient of variation of zsGreen protein for low DOX.∎
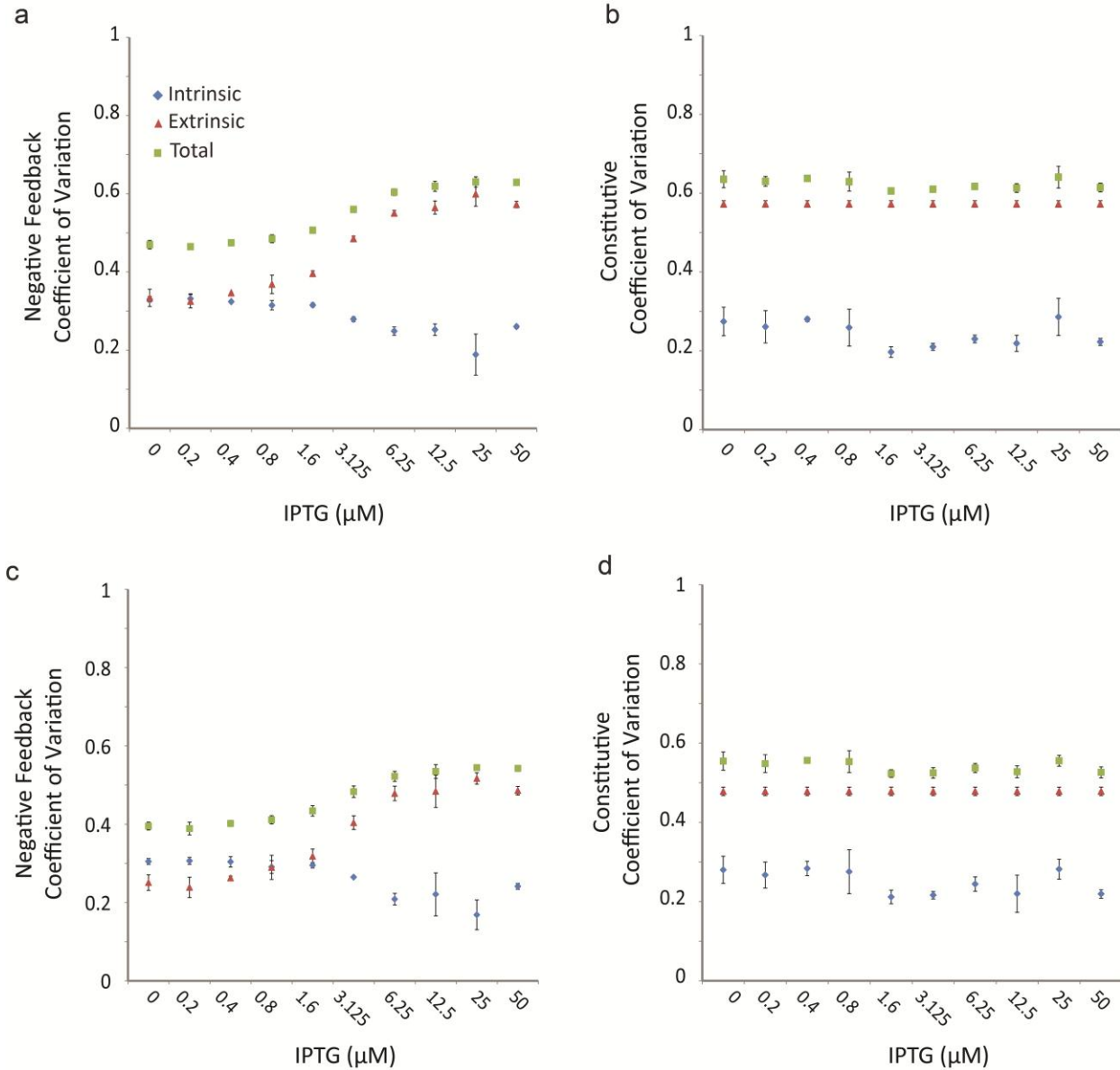
Supplementary Figure 10. Effect of forward scattering vs. side scatter gate on the extrinsic noise. The top panels (a and b) correspond to the original gate (SSC ~20k-120k and FSC ~20k-120k gate, supplementary figure 1) while panels c and d were prepared using smaller gate (SSC ~40k-50k and FSC ~70k-80k gate).■

Supplementary Figure 11. Comparison of mean fluorescence and coefficient of variation between the main paper negative feedback clone and two different negative feedback transgenes.∎
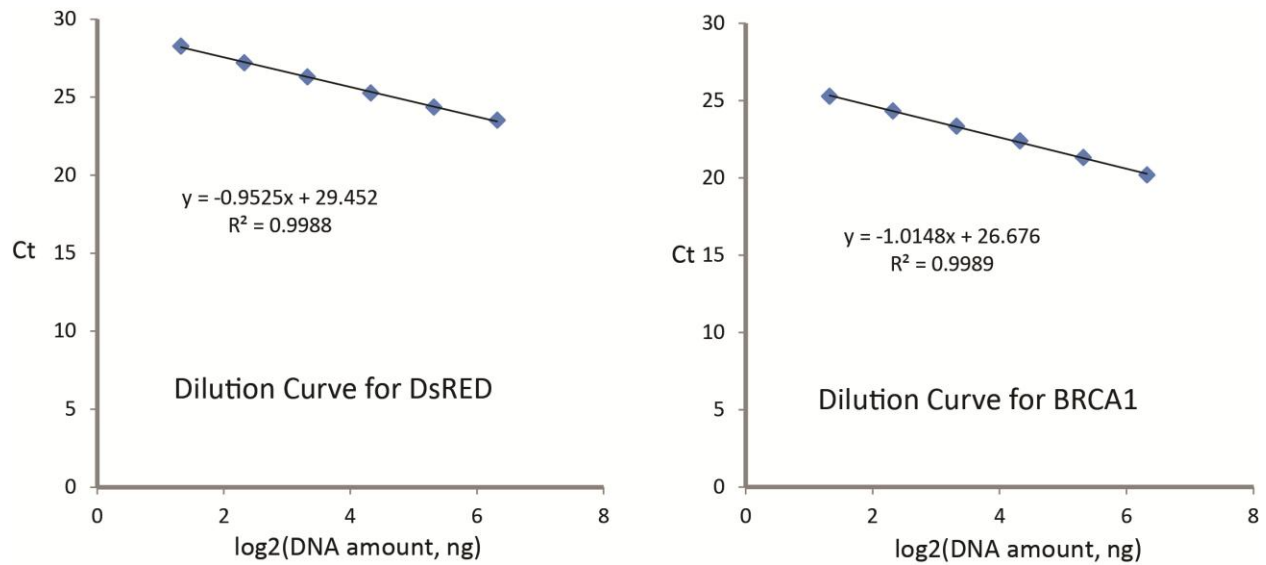
Supplementary Figure 12. Determination of PCR amplification efficiencies for DsRED and BRCA1 gene targets. The dilution curves (DsRED and BRCA1) were plotted as $\log_2$(DNA amount, ng) versus Ct. The PCR amplification efficiency E was calculated as: $2^{(-1/\text{slope of the dilution curve})}-1$. E_{DsRED} was determined as 1.07, and E_{BRCA1} as 0.98. ∎
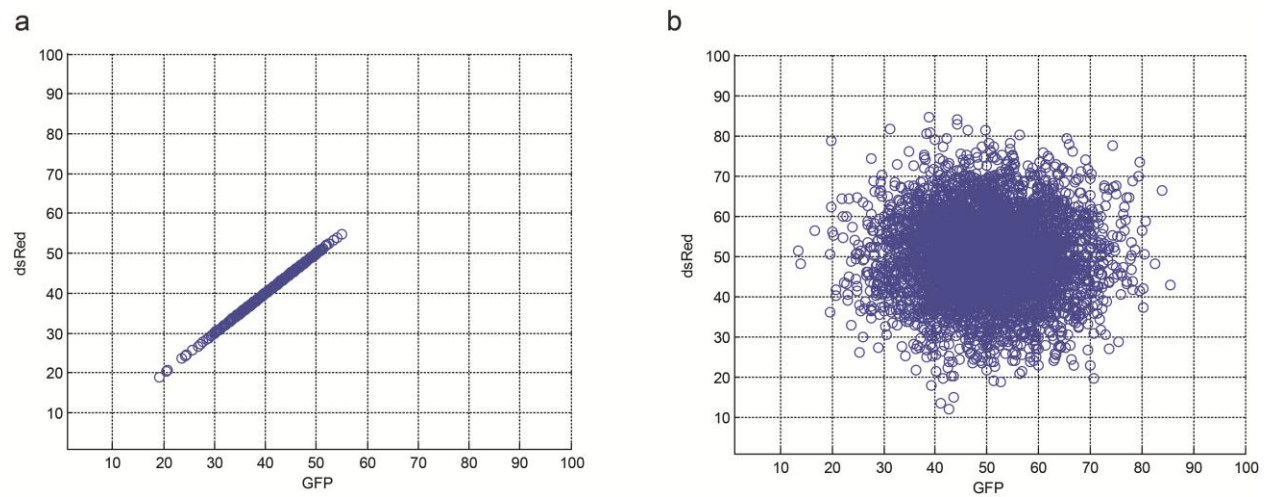
Supplementary Figure 13. Simulations where the intrinsic and extrinsic noise change separately. **(a)** We vary the strength of transcription of a single bidirectional promoter coding for two fluorescent proteins,

leading to perfectly correlated fluorescence quantities. **(b)** We vary the strength of transcription of two

fluorescent genes independently.■
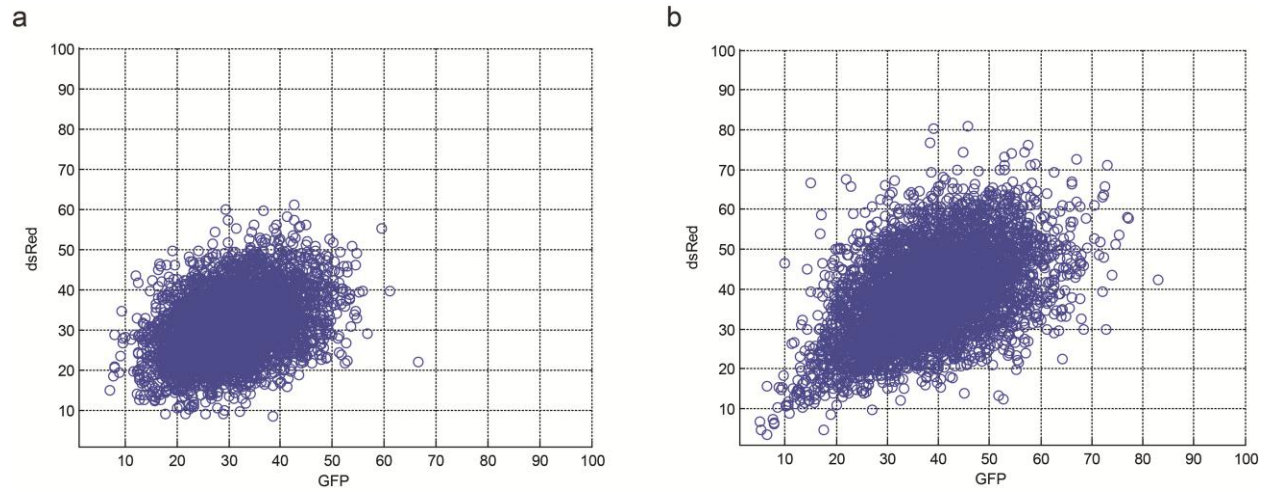
Supplementary Figure 14. Simulations where the intrinsic and extrinsic noise change simultaneously.■

Transgene copy number

Real-time quantitative PCR has been used as an alternative to Southern blot or fluorescence in situ hybridization for detection of gene copy numbers[1]. Various studies demonstrated that this method is accurate enough compared to Southern blot. For example, in Table 2 from "Determination of Cytochrome P450 2D6 (CYP2D6) gene copy number by real-time quantitative PCR"[2], the estimations of CYP2D6 gene copies from real-time quantitative PCR match with those from Southern blotting. The average copy numbers of DsRED of all stable clones were estimated by the delta delta Ct method as follows: $2^{-\Delta\Delta Ct} = ((1 + E_{DsRED})^{-\Delta Ct,DsRED}) /((1+E_{BRCA1})^{-\Delta Ct,BRCA1})$, where E_{DsRED} is the PCR amplification efficiency for DsRED and E_{BRCA1} for BRCA1 (endogenous reference gene)[3].

The PCR primers are: DsRED forward primer: 5'- ctccaccacggtgtagtcct-3'; DsRED reverse primer: 5'- agaccgtgtacaaggccaag-3'; BRCA1 forward primer: 5'- gagcgtcccctcacaaataa-3'; and BRCA1 reverse primer: 5'- tgctccgtttggttagttcc-3'. The control stable HEK293 cell line was generated by Flp-In system (Invitrogen) and contains one copy of DsRED transgene[4]. All genomic DNA samples were extracted using DNeasy Blood and Tissue kit (Qiagen). To determine the PCR amplification efficiency, genomic DNAs from the control cell line were used to generate the dilution curve of $\log_2$(DNA amount, ng) vs. Ct. E_{DsRED} was calculated as 1.07, and E_{BRCA1} as 0.98. The PCR conditions were as: 95 degree for 3 minutes, followed by 40 cycles of 95 degree for 15 seconds and 60 degree for 30 seconds. For each stable clone, triplicates (50 ng of genomic DNA) were performed and the average copy numbers were calculated as the mean ± SD. For statistical analysis, z scores were calculated against estimated integer copy numbers, and -1.96<z<1.96 was determined as no statistical difference (corresponding to 95% confidence interval).

Determination of PCR amplification efficiency: 80 ng, 40 ng, 20 ng, 10 ng, 5 ng and 2.5 ng of genomic DNAs were extracted from the control cell line using DNeasy Blood and Tissue kit (Qiagen). The PCR primers are: DsRED forward primer: 5'- ctccaccacggtgtagtcct-3'; DsRED reverse primer: 5'-

agaccgtgtacaaggccaag-3'; BRCA1 forward primer: 5'- gagcgtcccctcacaaataa-3'; and BRCA1 reverse

primer: 5'- tgctccgtttggttagttcc-3'. The PCR conditions were as: 95 degree for 3 minutes, followed by 40

cycles of 95 degree for 15 seconds and 60 degree for 30 seconds. The dilution curves were plotted

(**Supplement Fig. 12**) as $\log_2$(DNA amount, ng) vs. Ct. The PCR amplification efficiency E was

calculated as: $2^{(-1/\text{slope of the dilution curve})}-1$.

Supplementary Table I. Transgene clones and the resulting number of integrations.

Clone	Gene Copy Average	Standard Deviation
Negative Feedback (NF Clone I: L6.89.14)	1.026716891	0.205356713
Simple Negative Regulation (SNR Clone I: V2.42.1)	1.986616106	0.063272704
Single integration clone	1.005016457	0.127722277
Simple Negative Regulation (SNR Clone II: V2.52.2)	1.024331368	0.09450017

Supplementary Table II: The values for alpha obtained for the manuscript clones.

IPTG (µM)	Negative Feedback High Dox	Negative Feedback Low Dox	Simple Negative Regulation High Dox	Simple Negative Regulation Low Dox
50	1	1.22	1.11	1.28
25	1.046293	1.280207	1.170337	1.332335
12.5	0.986329	1.211581	1.232468	1.230991
6.25	0.961434	1.195583	1.064725	1.084425
3.125	0.847654	1.111019	0.868676	0.861535
1.6	0.692237	0.864742	0.684977	0.626666
0.8	0.64337	0.749677	0.548974	0.431871
0.4	0.605457	0.677382	0.389802	0.303799
0.2	0.568188	0.649937	0.323275	0.265747
0	0.5828	0.630114	0.337724	0.23634

Theory

Stochastic events which govern the concentration of a single protein, such as the synthesis and degradation of that protein, are referred to as "intrinsic" or "local" noise. Such random fluctuations can propagate along regulation pathways, with the consequence that protein distributions along a pathway appear correlated[5, 6]. However, even proteins from different regulation pathways show correlation[5, 7]. This arises from stochastic variations in quantities which affect the regulation of all genes[5, 6], such as in polymerase copies or cell cycle phase. As a consequence, a strongly expressing constitutive promoter is expected to have little intrinsic noise, while a weak promoter will have high intrinsic noise[8, 9]. In addition, two identical, independently regulated promoters are expected to have the same extrinsic noise, which arises through global effects[5, 7].

The total noise observed in a fluorescent reporter distribution arises through the combination of these "global" or "extrinsic" fluctuations together with the fluctuations in that protein's local regulation machinery ("intrinsic" noise)[5]. The intrinsic noise and extrinsic noise squared, sum to the CV-squared of the fluorescent reporter[5]. Using this notation, let angle brackets indicate that an average is taken with extrinsic variables held fixed, and let an overbar indicate an average where intrinsic variables are fixed. Then the three noises, intrinsic, extrinsic, and total, can be written in terms of P, the observed distribution of reporter protein:

$$n_{tot}^2 = \frac{\overline{\langle P^2 \rangle} - \left(\overline{\langle P \rangle} \right)^2}{\left(\overline{\langle P \rangle} \right)^2}$$

$$n_{int}^2 = \frac{\overline{\langle P^2 \rangle - \langle P \rangle^2}}{\left(\overline{\langle P \rangle} \right)^2}$$

$$n_{ext}^2 = \frac{\overline{\langle P \rangle^2} - \left(\overline{\langle P \rangle} \right)^2}{\left(\overline{\langle P \rangle} \right)^2}$$

$$n_{tot}^2 = n_{int}^2 + n_{ext}^2$$

For intrinsic noise, the authors of[5] take the variance of the intrinsic variables, $\langle P^2 \rangle - \langle P \rangle^2$, then estimate the expected value of this variance, denoted by the overbar, and subsequently divide by the mean squared of P. For extrinsic noise, the authors take the expected value of P with respect to intrinsic variables, then the variance of $\langle P \rangle$, and finally divide by the mean squared of P. For the total noise (CV-squared), the variance of P is divided by the mean squared.

Note that with a single reporter, the noises can't be estimated unless both the intrinsic and extrinsic variables are observed. However, in the standard two-reporter experiment, the extrinsic noise becomes the normalized covariance of two reporters that are independently regulated and identically distributed. The reason[5] is that in a single cell, the extrinsic variable is fixed, so the quantity $\overline{\langle P \rangle^2}$ can be calculated as the average product of the two reporters $\overline{\langle P^{(1)} P^{(2)} \rangle}$, then since $\overline{\langle P^{(1)} \rangle} = \overline{\langle P^{(2)} \rangle}$ the extrinsic noise becomes the normalized covariance of the two reporters:

$$n_{ext}^2 = \frac{\overline{\langle P^{(1)} P^{(2)} \rangle} - \left(\overline{\langle P^{(1)} \rangle}\right)\left(\overline{\langle P^{(2)} \rangle}\right)}{\left(\overline{\langle P^{(1)} \rangle}\right)\left(\overline{\langle P^{(2)} \rangle}\right)}$$

and the intrinsic noise becomes the normalized RMS difference from $P^{(1)} = P^{(2)}$, so that the sum of intrinsic and extrinsic is twice the CV of one reporter.

In this paper we examine more complicated regulatory mechanisms where it is not feasible to construct two identically-regulated reporters (or impossible to obtain identical reporter statistics). We define a new formulation and we will obtain the previous results as a special case, where the extrinsic noise is the normalized covariance and the components sum to the total noise.

Let X be the observed reporter protein, and A and B are the intrinsic and extrinsic variables; if we assume a multiplicative model and that the variables are independently distributed, we can derive the following intrinsic/extrinsic noise breakdown:

$$\frac{E\big((X - \mu_X)^2\big)}{\mu_X{}^2} = \frac{E\big((A - \mu_A)^2\big)}{\mu_A{}^2} + \frac{E\big((B - \mu_B)^2\big)}{\mu_B{}^2}$$

Such a multiplicative model can be motivated as follows. Suppose gene X is activated by two factors; one (A) is an intrinsic variable such as a transcription factor, and the other (B) is an extrinsic variable, such as RNA polymerase. Suppose both factors must be present for transcription, in the complex ABX. We have four reaction equations:

$$X + A \leftrightarrow AX$$

$$X + B \leftrightarrow BX$$

$$AX + B \leftrightarrow ABX$$

$$BX + A \leftrightarrow ABX$$
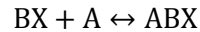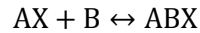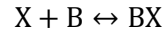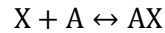
This results in the following algebraic equations at steady-state:

$$k_{r1}[AX] + k_{r2}[BX] = k_{f1}[A][X] + k_{f2}[B][X]$$

$$k_{f1}[A][X] + k_{r3}[ABX] = k_{r1}[AX] + k_{f3}[B][AX]$$

$$k_{f2}[B][X] + k_{r4}[ABX] = k_{r2}[BX] + k_{f4}[A][BX]$$

$$k_{f4}[A][BX] + k_{f3}[B][AX] = k_{r4}[ABX] + k_{r3}[ABX]$$

For the gene activity, we take the ratio of active complex ABX to total gene copies:

$$X_{active} = \frac{[ABX]}{[X] + [AX] + [BX] + [ABX]}$$

This simplifies to an expression in terms of A and B (we drop most of the constants):

$$X_{active} = \frac{[A][B] + [A]^2[B] + [A][B]^2}{k + [A] + [B] + [A]^2 + [B]^2 + [A]^2[B] + [A][B]^2}$$

Which, for small, unsaturated concentrations of A and B, looks like:

$$X_{active} \approx \frac{[A][B]}{k}$$

Intuitively, in this multiplicative approximation, a polymerase fluctuation of 10% is expected to change gene activity by 10% (with an unsaturated promoter). Compare this to an additive noise model: now, the same polymerase fluctuation of 1000 molecules is expected to change gene output by 1000 molecules, regardless of whether the output is currently regulated at 10000 molecules or at 100 molecules. Thus the multiplicative model makes physical sense for positive variables, where a reporter with 100 molecules cannot have an uncertainty of 1000 molecules.

We generalize this multiplicative model and assume the observed random variable is a function of its independent component sources (A and B, the intrinsic and extrinsic variables) of the following general form:

$$X = A^a B^b$$

i.e. where a and b are not necessarily both equal to 1. These sensitivity coefficients must appear as powers because multiplied coefficients fall out as a single constant in the next step. It is convenient to convert this to a linear model (for ease of calculation), by taking the logarithm:

$$\log(X) = a \log(A) + b \log(B)$$

For ease of notation, we drop the log functions and just use the original variable names.

$$X = a\,A + b\,B$$

Here we need to calculate the contributions of A (intrinsic) and B (extrinsic) to the total observed noise of X. In general, summing two independent random variables A and B with variance $Var(A)$ and $Var(B)$ results in the following variance:

$$\text{Var}(aA + bB) = a^2\text{Var}(A) + b^2\text{Var}(B) + 2ab\text{Cov}(A, B) = a^2\text{Var}(A) + b^2\text{Var}(B)$$

The last equality holds because the intrinsic and extrinsic components have been defined to have no covariance term: any fluctuation which affects two identical reporters is an extrinsic variable, and all remaining noise observed is intrinsic. The variances of logarithms returned by this method are approximately the normalized variances (CV-squares) of the original quantities; we discuss this point later.

Elowitz et al.[7] argues that two identically regulated reporters with the same mean and variance should have the same extrinsic noise, and uses this fact to calculate the noise components. We introduce the following modification to extend the intrinsic/extrinsic breakdown to cases where one reporter is not constitutive, and hence may not obey the Elowitz et al. assumptions. For a regulated reporter, noise may propagate along the regulatory pathway, changing the reporter's susceptibility to global fluctuations (supposing that this reporter is in the same cell as a constitutive reporter). We capture this asymmetric effect by adding a sensitivity coefficient to the regulated reporter for a two-reporter noise breakdown.

Let Y be the constitutive reporter and X a regulated reporter (controlled by an inducer, in our case IPTG). A is a function of all extrinsic variables, B and C are the intrinsic variables for each promoter, and α is a coefficient which is 1 for two constitutive promoters with identical reporter statistics (as in Elowitz et al.) but varies depending on the regulation of X. α represents an aggregated susceptibility to fluctuating variables which affect both reporters, similar to the quantity H used in Paulsson[10] to denote the logarithmic gain of an interaction:

$$X = A^\alpha B$$

$$Y = AC$$

With α placed as a power of A, we have defined $(\ln(X|B) - \langle\ln(X|B)\rangle)/(\ln(A) - \langle\ln(A)\rangle) = \alpha$, the logarithmic gain of fluctuations in A (extrinsic noise sources) transmitted to X, which we assume is

independent of the value of B and of the size of the fluctuation. Selecting $\alpha = 1$ results in the same noise breakdown as Elowitz et al.[7], where the extrinsic noise is the normalized covariance of both reporters, and the intrinsic noise is the total CV-squared minus the extrinsic noise.

Intuitively, the inverse tangent of α is the slope of the data on a log-log plot that lies along the 45-degree diagonal in the special case $\alpha = 1$ but does not if the two reporters experience different fluctuation magnitudes from extrinsic sources due to the presence of noise-changing regulatory components.

We take the logarithm to convert to a linear model to find the components.

$$\log(X) = \alpha \log(A) + \log(B)$$

$$\log(X) = \log(A) + \log(C)$$

Once more dropping the log notation for simplicity,

$$X = \alpha A + B$$

$$Y = A + C$$

Taking the covariance of the logarithms of the reporters (this can be done directly with cytometry data – gate as in figure S1, take the log of the raw reporter values, and calculate the covariance):

$$Cov(X, Y) = Cov(\alpha A + B, A + C)$$

$$= Cov(\alpha A, A) + Cov(B, A) + Cov(\alpha A, C) + Cov(B, C)$$

Because A, B, and C are defined as uncorrelated,

$$Cov(X, Y) = Cov(\alpha A, A) = \alpha \, Var(A)$$

We take the variances of the logarithms of X and Y, which are the experimentally determined total noises,

$$Var(X) = \alpha^2 Var(A) + Var(B)$$

$$Var(Y) = Var(A) + Var(C)$$

Next, we replace the variance of A with the experimental covariance term,

$$Var(X) = \alpha \, Cov(X, Y) + Var(B)$$

$$Var(Y) = Cov(X, Y)/\alpha + Var(C) \tag{I}$$

Our goal is to compute α_i (where $i = 1, 2, \ldots, N$ ranges over possible inducer concentrations), $Var(A_i)$, $Var(B_i)$, and $Var(C_i)$, from the knowledge of $Cov(X_i, Y_i)$, $Var(X_i)$, and $Var(Y_i)$ Here, the covariance terms correspond to the effects of the extrinsic noise, and the B and C terms are intrinsic noises. So far we have 2N equations (the expressions for $Var(X_i)$ and $Var(Y_i)$) in 3N unknowns (α_i and the variances of B and C for each well), but we also know that the inducer of the regulated protein (in our case IPTG) has no effect on the global fluctuations that contribute to noise in the constitutive reporter. This means the extrinsic noise term in (I), $Cov(X_i, Y_i)/\alpha_i$, (and directly $Var(Y_i)$ and $Var(C_i)$ also) should be the same for all inducer conditions, that is, we have $N - 1$ additional equations

$$\frac{Cov(X_i, Y_i)}{\alpha_i} = \frac{Cov(X_j, Y_j)}{\alpha_j} \tag{II}$$

for all i, j. Finally, $\alpha_N = 1$ for the unregulated control experiment (in our case fully induced with IPTG), so both constitutively produced reporters have the same extrinsic noise, as in Elowitz et al [7].

The choice of $\alpha_N = 1$ sets the noise components for the fully induced well, but we need to calculate α_i for the other wells. For wells without full IPTG induction, it follows from (II) that we can set α_i equal to $\frac{Cov(X_i, Y_i)\, \alpha_N}{Cov(X_N, Y_N)}$, i.e. the ratio of the covariance of the current well to the extrinsic noise of the unregulated control experiment. This forces the extrinsic noise of the constitutive reporter ($Cov(X_i, Y_i)/\alpha_i$) to the same value $Cov(X_N, Y_N)/\alpha_N$ for all wells. The computed α_i are then used to calculate the

extrinsic noise of the regulated protein $\alpha_i \text{Cov}(X_i, Y_i)$ for each well, and by subtraction we then obtain the intrinsic components.

All of these terms $\text{Cov}(X_i, Y_i)$ are unbiased estimates of the sample covariance computed as follows, for cytometry data where each individual well i has m cells recorded with reporter measurements x_i^j and y_i^j indexed $j = 1, 2, \ldots, m$:

$$\text{Cov}(\log(X_i), \log(Y_i)) = \frac{1}{m-1} \sum_{j=1}^{m} \left[\left(\log(x_i^j) - \frac{1}{m} \sum_{k=1}^{m} \log(x_i^k) \right) \left(\log(y_i^j) - \frac{1}{m} \sum_{k=1}^{m} \log(y_i^k) \right) \right]$$

We note that in some experiments the two reporters in the fully induced well do not have the identical statistics required by the dual reporter theory of Elowitz et al. We attribute this to the difficulty in finding statistically identical genes and promoters, and also to measurement related issues. For these experiments, instead of assuming the both reporters have the same intrinsic and extrinsic noise, we may assume that they are proportionally the same. For example, if one reporter's CV-square is 1.2 times the other's we can assume its intrinsic and extrinsic CV-squares are also 1.2 times as large. Hence for the computation, instead of assuming $\alpha_N = 1$ for the fully induced well, we would assume α_N is the ratio of the CVs of the reporters. Otherwise the calculation proceeds the same way.

In summary, this is how we define the noise components, using as data n_{totX}^2 and n_{totY}^2, the experimentally determined CV-squares of reporters X and Y (which, as described below, we approximate with the variances of the logs of X and Y) and the covariance of the logarithms of X and Y, also determined experimentally (inducer concentrations are indexed $i = 1, 2, \ldots, N$, where well N is fully induced):

$$\alpha_N = \frac{n_{totX}}{n_{totY}}$$

$$\alpha_i = \frac{\text{Cov}(X_i, Y_i) \, \alpha_N}{\text{Cov}(X_N, Y_N)}, i \neq N$$

$$n_{extX}^2 = \alpha_i\, Cov(X_i, Y_i)$$

$$n_{intX}^2 = n_{totX}^2 - n_{extX}^2$$

$$n_{extY}^2 = Cov(X_i, Y_i)/\alpha_i$$

$$n_{intY}^2 = n_{totY}^2 - n_{extY}^2$$

As mentioned earlier, we have calculated the variances of the logarithms of the reporters instead of the CV-squared. This argument relies on a linearized approximation: the standard deviations of the log-scale variables are approximately the relative standard deviations of the original variables. Indeed, for small values of $\log(X) - \mu_{\log(X)}$:

$$E\left(\left(\log(X) - \mu_{\log(X)}\right)^2\right) \cong E\left(\left(\frac{X - \mu_X}{\mu_X}\right)^2\right) = \frac{E((X - \mu_X)^2)}{\mu_X^2}$$

Thus, our computed quantities are approximations of the squared coefficient of variation, which is a standard measure of noise[7]. Note that this strategy replaces the data normalization performed in Elowitz et al. but performs a similar function. This approximation is very close for tight distributions, but gets worse for broad distributions (i.e., the approximation is worse for larger values of $\log(X) - \mu_{\log(X)}$).

We can verify the approximation by calculating the standard deviation of the logarithm of the data and comparing it to the RSD of the original data. We improve the approximation by trimming the largest values of $\log(X) - \mu_{\log(X)}$ by dropping all values more than 2.5 standard deviations from the mean of the log of the data (these points are not dropped from the direct RSD verification, and the cutoff was obtained by trying several values).

Verification and decomposition of simulated noise

In our noise decomposition, we expect random quantities which affect the expression of both genes to show up as extrinsic noise, while we expect random quantities which affect only a single gene to show up

as intrinsic. We address the case where one reporter may be less sensitive to extrinsic noise sources due to noise-reducing regulatory pathways. To see this, first take the simplest case of a two-color experiment: suppose we have a plasmid with a constitutive bidirectional promoter P coding for reporters X and Y, and let the only source of uncertainty be the plasmid copy number. Then we have production rates of each reporter:

$$\frac{dX}{dt} = k_1 P - k_2 X$$

$$\frac{dY}{dt} = k_3 P - k_4 Y$$

At steady-state, we have the relations

$$X = \frac{k_1}{k_2} P \rightarrow \log(X) = \log\left(\frac{k_1}{k_2}\right) + \log(P)$$

$$Y = \frac{k_3}{k_4} P \rightarrow \log(Y) = \log\left(\frac{k_3}{k_4}\right) + \log(P)$$

And we want to find the extrinsic noise, the normalized covariance, which we have defined approximately by taking the covariance of the logarithm of the data:

$$n_{ext}^2 = \text{Cov}(\log(X), \log(Y))$$

$$= \text{Cov}\left(\log\left(\frac{k_1}{k_2}\right) + \log(P), \log\left(\frac{k_3}{k_4}\right) + \log(P)\right)$$

$$= \text{Cov}\left(\log\left(\frac{k_1}{k_2}\right), \log\left(\frac{k_3}{k_4}\right)\right) + \text{Cov}\left(\log\left(\frac{k_1}{k_2}\right), \log(P)\right) + \text{Cov}\left(\log(P), \log\left(\frac{k_3}{k_4}\right)\right)$$

$$+ \text{Cov}(\log(P), \log(P))$$

$$n_{ext}^2 = \text{Cov}(\log(P), \log(P)) = \text{Var}(\log(P))$$

To calculate intrinsic noise we need the total noise of each reporter:

$$n_{totX}^2 = Var(log(X)) = Var\left(log\left(\frac{k_1}{k_2}\right) + log(P)\right) = Var(log(P))$$

$$n_{totY}^2 = Var(log(Y)) = Var\left(log\left(\frac{k_3}{k_4}\right) + log(P)\right) = Var(log(P))$$

Which shows that in this example there is no intrinsic noise; hence a common promoter for two reporters is an extrinsic source of noise:

$$n_{intX}^2 = n_{totX}^2 - n_{ext}^2 = 0$$

$$n_{intY}^2 = n_{totY}^2 - n_{ext}^2 = 0$$

Suppose instead that there were two different plasmids with promoters P1 and P2 coding for reporters X and Y, and let their copy number be independent random variables. Setting up the problem the same way,

$$\frac{dX}{dt} = k_1 P_1 - k_2 X$$

$$\frac{dY}{dt} = k_3 P_2 - k_4 Y$$

At steady-state,

$$X = \frac{k_1}{k_2} P_1 \rightarrow log(X) = log\left(\frac{k_1}{k_2}\right) + log(P_1)$$

$$Y = \frac{k_3}{k_4} P_2 \rightarrow log(Y) = log\left(\frac{k_3}{k_4}\right) + log(P_2)$$

Calculating the extrinsic noise,

$$n_{ext}^2 = Cov(log(X), log(Y))$$

$$= Cov\left(log\left(\frac{k_1}{k_2}\right) + log(P_1), log\left(\frac{k_3}{k_4}\right) + log(P_2)\right)$$

$$= \mathrm{Cov}\left(\log\left(\frac{k_1}{k_2}\right), \log\left(\frac{k_3}{k_4}\right)\right) + \mathrm{Cov}\left(\log\left(\frac{k_1}{k_2}\right), \log(P_2)\right) + \mathrm{Cov}\left(\log\ (P_1), \log\left(\frac{k_3}{k_4}\right)\right)$$

$$+ \mathrm{Cov}(\log\ (P_1), \log(P_2))$$

$$n_{ext}{}^2 = \mathrm{Cov}(\log\ (P_1), \log(P_2)) = 0$$

For the total noise,

$$n_{totX}{}^2 = \mathrm{Var}(\log(P_1))$$

$$n_{totY}{}^2 = \mathrm{Var}(\log(P_2))$$

$$n_{intX}{}^2 = n_{totX}{}^2 - n_{ext}{}^2 = \mathrm{Var}(\log(P_1))$$

$$n_{intY}{}^2 = n_{totY}{}^2 - n_{ext}{}^2 = \mathrm{Var}(\log(P_2))$$

Hence in this case, where the random variable independently affects the two reporters, the extrinsic noise is zero, making these intrinsic noise sources.

Notice that the strength of our approach is when the two reporters are not identically regulated with identical statistics. We extend the applicability by assigning different extrinsic noise quantities to each reporter, so that now instead of there being a single extrinsic noise, each reporter has its own set of intrinsic and extrinsic contributions. The following example shows what can happen to extrinsic noise in the case of negative feedback. Suppose we have the extrinsic promoter case, but reporter X has negative feedback (and $k_5X \gg 1$):

$$\frac{dX}{dt} = k_1P\frac{1}{1 + k_5X} - k_2X \approx \frac{k_1P}{k_5X} - k_2X$$

$$\frac{dY}{dt} = k_3P - k_4Y$$

At steady-state,

$$X^2 = \frac{k_1 P}{k_2 k_5} \rightarrow \log(X^2) = \log\left(\frac{k_1}{k_2 k_5}\right) + \log(P) \rightarrow \log(X) = k + \frac{1}{2}\log(P)$$

$$Y = \frac{k_3}{k_4}P \rightarrow \log(Y) = \log\left(\frac{k_3}{k_4}\right) + \log(P)$$

Calculating the extrinsic noise,

$$n_{ext}^2 = \text{Cov}(\log(X), \log(Y))$$

$$= \text{Cov}\left(k + \frac{1}{2}\log(P), \log\left(\frac{k_3}{k_4}\right) + \log(P)\right)$$

$$= \text{Cov}\left(k, \log\left(\frac{k_3}{k_4}\right)\right) + \text{Cov}(k, \log(P)) + \text{Cov}\left(\log(P), \log\left(\frac{k_3}{k_4}\right)\right) + \text{Cov}\left(\frac{1}{2}\log(P), \log(P)\right)$$

$$n_{ext}^2 = \text{Cov}\left(\frac{1}{2}\log(P), \log(P)\right) = \frac{1}{2}\text{Var}(\log(P))$$

For the total noise,

$$n_{totX}^2 = \text{Var}(\log(X)) = \text{Var}\left(k + \frac{1}{2}\log(P)\right) = \frac{1}{4}\text{Var}(\log(P))$$

$$n_{totY}^2 = \text{Var}(\log(Y)) = \text{Var}\left(\log\left(\frac{k_3}{k_4}\right) + \log(P)\right) = \text{Var}(\log(P))$$

If we calculate the intrinsic noise using the Elowitz et al. approach, we find that the extrinsic noise exceeds the total noise for reporter X. However, for this simplified example, we know the only noise source is an extrinsic variable, and thus the intrinsic noise should turn out to be zero. This allows us to infer that α, as described in the previous section, has a value of ½, representing the fact that reporter X experiences half as much noise from the variable plasmid copy number as Y does.

$$n_{extX}{}^2 = \frac{\alpha}{2}\text{Var}(\log(P)) = \frac{1}{4}\text{Var}(\log(P))$$

$$n_{extY}{}^2 = \frac{1}{2\alpha}\text{Var}(\log(P)) = \text{Var}(\log(P))$$

$$n_{intX}{}^2 = n_{intY}{}^2 = 0$$

Recall that in the experiments, we must first estimate α in a case where the Elowitz et. al. assumptions hold, i.e., the reporters are identically regulated.

To confirm our analysis we used simulations to test the decomposition on noise for two extreme cases, where we control the levels of intrinsic and extrinsic noise. As illustrated in **Supplementary Fig. 13a** we first we vary the strength of transcription of a single bidirectional promoter coding for two fluorescent proteins, leading to perfectly correlated fluorescence quantities, which our decomposition shows to have only extrinsic noise and no intrinsic noise. Next, in **Supplementary Fig. 13b** we vary the strength of transcription of two fluorescent genes independently, which leads to uncorrelated fluorescence quantities; our method returns only intrinsic noise and no extrinsic noise.

Furthermore, in **Supplementary Fig. 14,** we simulated mixtures of both noise types, generated by varying the strength of transcription of two different reporter genes as in the intrinsic noise case in the previous figure, but also by varying the amount of a transcription factor which regulates both genes. For panel a, the noise breakdown gives an intrinsic noise of 0.22 and an extrinsic noise of 0.16 for both proteins. For panel b, where we give the common transcription factor extra variability, it raises the extrinsic noise substantially (but not the intrinsic noise; the common transcription factor is an "extrinsic variable"). The intrinsic noise changes slightly to 0.21 and the extrinsic noise jumps to 0.23.

References

1. Hoebeeck, J., Speleman, F. & Vandesompele, J. in (eds Hilario, E. & Mackay, J.) 205-226 (Humana Press, 2007).

2. Bodin, L., Beaune, P. H. & Loriot, M. Determination of Cytochrome P450 2D6 (CYP2D6) Gene Copy Number by Real-Time Quantitative PCR. *Journal of Biomedicine and Biotechnology* **2005**, 248-253 (2005).

3. Zheng, S. *et al*. DNA hypermethylation profiles associated with glioma subtypes and EZH2 and IGFBP2 mRNA expression. *Neuro-oncology* **13**, 280-289 (2011).

4. Li, Y., Moore, R., Guinn, M. & Bleris, L. Transcription activator-like effector hybrids for conditional control and rewiring of chromosomal transgene expression. *Scientific Reports* **2:897** (2012).

5. Swain, P. S., Elowitz, M. B. & Siggia, E. D. Intrinsic and extrinsic contributions to stochasticity in gene expression. *Proceedings of the National Academy of Sciences* **99**, 12795-12800 (2002).

6. Volfson, D. *et al*. Origins of extrinsic variability in eukaryotic gene expression. *Nature* **439**, 861-864 (2006).

7. Elowitz, M., Levine, A., Siggia, E. & Swain, P. Stochastic Gene Expression in a Single Cell. *Science* **297**, 1183-1186 (2002).

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

9. Newman, J. R. S. *et al*. Single-cell proteomic analysis of S. cerevisiae reveals the architecture of biological noise. *Nature* **441**, 840-846 (2006).

10. Paulsson, J. Summing up the noise in gene networks. *Nature* **427**, 415-418 (2004).