CONSTRUCTION AND CHARACTERIZATION OF A NOISE CONTROL SYSTEM IN HUMAN EMBRYONIC KIDNEY CELLS (HEK293)

A Senior Thesis Presented to

The Faculty of the Department of Molecular Biology

The Colorado College

In Partial Fulfillment of the Requirements for the Degree

Bachelor of Arts

By

Qiu Chang Wu

26th day of March 2018

_____

Professor Phoebe Lostroh, PhD

Primary Thesis Reader

_____

Professor Jennifer Garcia, PhD

Secondary Thesis Reader

ABSTRACT

Non-genetic cellular heterogeneity is often overlooked in the study of molecular biology. Population averaged measurements of clonal cell populations are made under the assumption that genetic homogeneity implies cellular homogeneity. On the contrary, such assumption discounts the vast variability that exists within a clonal cell population. When gene expression of individual genes is observed across a clonal population of mammalian cells, variation in expression of the same gene differ between cells. This phenomenon is defined as gene expression noise and has been shown to have a functional role in processes like cell fate decisions and viral latency reactivation. While many efforts have been made to measure gene expression noise, less is known about how to control noise. Here, we present our work towards controlling noise using a synthetic genetic circuit we call a noise rheostat. The circuit we built places two small-molecule inducible transcription systems linked in series, driving expression of a green fluorescent protein reporter gene. The inducible transcription system includes an abscisic acid inducible synthetic transcription factor with its cognate promoter and a doxycycline inducible transcription factor with its cognate promoter. By using two inducible transcription systems in series, we create lags in transcription of the terminal output and thus produce different noise levels. We performed transient transfections and characterized the system through dosage experiments using flow cytometry. The datasets analyzed demonstrate that gene expression noise is dialable while maintaining gene expression mean. These results offer a promising prototype for the first mammalian noise rheostat. We propose that this tool will be useful in the study of noise biology as it provides the ability to separate the control of gene expression mean from gene expression noise.

INTRODUCTION

A short history of synthetic biology

The French biologist Stephane Leduc wrote the first known reference of the term "synthetic biology". In a book titled "La Biologie Synthetique", he proposed that the study of biology should not merely be observational but reproducible through synthetic methods. The modern field of synthetic biology follows the philosophy proposed by Leduc. That is that much could be learned about nature through synthetic reproductions in a controlled laboratory setting.

In 2006, the first major source of funding to support the emerging field of synthetic biology came from the National Science Foundation. The formation of the Synthetic Biology Engineering Research Center (SynBERC) was a milestone for the field as it was public acknowledgement of the scientific potential of synthetic biology. In 2007, the United States Department of Energy funded national laboratories across the country (Oak Ridge National Laboratory and Lawrence Berkeley National Laboratory) to focus on the advancement of synthetic biotechnologies for bioenergy development. As of now, major sources of public funding come from the National Science Foundation, National Institute of Heath, Department of Defense, and Department of Energy.

In the contemporary study of synthetic biology, the goal is to understand basic biology through an engineering and physics lens. Systems biology, a field that seeks to holistic study biological systems through mathematical and computational modeling, is inextricably linked to synthetic biology. By viewing cells as machines, scientists hope to apply engineering principles to explore how cellular and molecular processes are wired and thus give rise to phenotypic manifestations (Cameron et al. 2014). The first applications of synthetic biology were to understand cellular regulation wiring through *de novo* engineering of regulation behavior from

known circuit architectures using synthetic genetic circuits. In this introduction of synthetic biology, I hope to cover a few landmark papers that have given rise to major topics of study within synthetic biology and thereby contextualize my project within the scope of this field.

Jacob and Monod's landmark papers on the *trp* and *lac* operon in *Escherichia coli* initially sparked the interest in regulatory circuitry. Scientists were interested in how regulation at the transcriptional and translational level creates the necessary circuit architecture to produce different phenotypes. As the science of electronic circuitry was already understood, the first synthetic genetic circuits constructed and characterized in *E. coli* attempted to implement known electronic circuit behavior in a biological setting. These circuit behaviors allowed for biologist to determine the kinetic properties of biological components and frame biology through the lens of math and physics. Electronic circuits like toggle switches (Gardner et al. 2000, Appendix IV), oscillators (Elowitz et al. 2000, Appendix IV), and logic gates (Guet et al. 2002, Weiss et al. 2002, Appendix IV) were constructed as genetic circuitry. As oppose to a voltage output as in an electronic circuit, what was often measured in genetic circuits was mRNA or protein levels. The construction of these circuits was not without studying the kinetics in transcription and translation as well as the feedback and feed forward architectures that give rise to the signal behaviors observed. McAdams et al. (1995, 2000) have done the first set of mathematical modeling and simulations to determine output signals for numerous sets of circuit architectures. These landmark papers demonstrate the ability for biologist to manipulate native proteins and wiring architectures to produce modeled outputs within a living cell or organism. This simple idea gave rise to the field of synthetic biology.

In the past few decades, technological advancements have aided the growth of synthetic biology. With the rise of computational power through Moore's Law (Appendix IV),

mathematical modeling and simulations capabilities have increased in both speed and ease. With the advancement of biotechnologies, specifically in recombinant DNA assembly technologies, experimental implementation of biological circuitry has also increased in both speed and volume of circuitry tested. Cohen et al. in 1973 first introduced restriction cloning and revolutionized molecular biology. Consequently, other DNA recombination techniques have been developed using similar principles employed in restriction cloning. Today, assembly methods like BioBrick® Assembly (Knight, T. et al. 2003), Gibson Assembly® (Gibson et al. 2009), and Golden Gate Assembly (Engler et al. 2008, Sarrion-Perdigones et al. 2011) have increased the efficiency of cloning and synthetic circuitry construction.

In parallel to the development of new DNA assembly methods, the cost of *de novo* DNA synthesis and DNA sequencing have become exponentially cheaper since the human genome project in the 1990s (Smanski et al. 2016). Thus, with the decrease cost compounded with the increase efficiency of DNA synthesis, the growth of synthetic biology has been largely supported through these avenues.

Circuitry construction is only one facet of synthetic biology; the goal in circuitry construction is to provide tools for biological engineering. Metabolic and genomic engineering are two fields that have particularly benefitted from the tools developed. Applications include industrial manufacturing, drug biosynthesis, medicine, and programmable organism construction. Below, I will highlight examples from some of these applications.

Metabolic engineering (Appendix IV) has been around for many years, but synthetic biology has enhanced the capacity to engineer metabolic pathways to produce products of interest. Difficult-to-synthesize biomolecules and organic compounds have become easier to manufacture through improved metabolic engineering techniques. Artemisinin, an anti-malarial

drug manufactured by Novartis and Sanofi-Aventis, was difficult and rather expensive to organically synthesize (Schmid and Hofeinz 1983). At its highest rate, the cost of a kilogram of artemisinin was $1,200 USD between 2005 and 2008 (Anon 2008). The Keasling laboratory at University of California, Berkeley developed a way to synthesize artemisinic acid, a precursor to artemisinin, in *Saccharomyces cerevisiae* at high yield with minimal purification (Ro et al. 2006, Paddon et al. 2013). They done so by introducing non-native genes into yeast to construct the entire biosynthetic pathway of artemisinic acid. The company Amyris, which made this yeast strain available royalty-free to Sanofi-Aventis for manufacturing and distribution in the developing world, commercialized the process developed in the Keasling laboratory.

Artemisinin is but one molecule whose synthesis has been aided by synthetic biology. Synthesis of short chain alcohol (Yim et al. 2011) and isoprenoids (Ajikumar et al. 2010) for biofuel production have also been developed. In these cases, synthesis processes have been optimized *in vivo* to increase efficiency and yield. The synthetic metabolic pathways constructed often co-localized protein complexes to increase proximity of chemical reactions within a cell. This strategy is akin to creating an industrial factory line assembly *in vivo* (Smanski et al. 2016).

Using synthetic biology, construction of programmable organisms is a developing area of research. Its application includes areas within environmental remediation and medicine. Some synthetic biologists have been interested in developing programmable organisms as biosensors to sense toxic chemicals at low concentrations within the environment (Khalil and Collins 2010). Others have been interested in the human therapeutic applications with programmable organisms. This includes programming bacteria to kill tumors (Anderson et al. 2006), constructing artificial microbiome communities for treatment of gastrointestinal disorders

(Biteen et al. 2015), and modifying immune cells for targeted attacks against pathogens and cancerous cells (Esensten et al. 2017).

Although synthetic biology holds incredible potential in the ability to provide control within a biological system, variability in the form of noise limits how well a biological system may be controlled (Knuuttila and Loettgers 2014). By using genetic circuitry with predictable outputs, both the origin and implication of variability within a biological system may be better understood.

Introduction to Gene Expression Noise

In biology, prior to the understanding of epigenetics, we assumed that genetic homogeneity implied phenotypic homogeneity. Thus, it became standard practice to take population averaged measurements as surrogate measures for average behavior of a clonal population. However, this practice discounts the role of variability, or noise, that exist between genetically identical cells. In this section, I will be talking about noise and its functional role within a biological system. Noise in gene expression is defined as the variability in the abundance of gene products (i.e. mRNAs, non-coding RNAs, and proteins) between cells within a clonal population (Raser and O'Shea 2005). This noise is suggested to explain for phenotypic differences between cells despite genetic homogeneity.

The origin of noise in gene expression has been attributed to the inherent stochastic nature of gene expression. The stochasticity, or randomness, comes from the binding events of proteins to DNA, RNA, and/or protein. Binding requires multiple components to first find each other, interact, and initiate a programmed protocol. Because binding events are often not directed, interactions happen as the molecules move randomly throughout the cell (i.e. stochastic

kinetics). At the epigenetic, transcriptional, and translational levels, the kinetics and stoichiometric fluctuations of mRNAs and proteins between cells propagate as noise throughout gene expression (Raj and Van Oudenaarden 2008). This may be due to the small number of molecules, in DNA, mRNA and protein, which transcription and translation begins with. Thus, both processes are highly affected by stochastic kinetics.

There are two different types of gene expression noise: intrinsic and extrinsic noise. They are defined by where the noise originates, but both types of noise contribute to the total gene expression noise. Intrinsic noise refers to the variation caused by the regulation of the gene and the biochemical processes of gene expression. This property is inherent to the gene's DNA sequence and depends on regulatory elements in the DNA like enhancer, repressor, and/or activating sequences. Meanwhile, extrinsic noise refers to the global differences between cells, e.g. number of polymerases, regulatory proteins, location of these molecules, and etc. (Elowitz et al. 2002). These differences may stem from micro-fluctuations within the environment, asymmetric distribution of molecules during cell division, and cell cycle asynchrony (Swain et al. 2002, Sigal et al. 2006).

Total gene expression noise can be easily measured as mRNA or protein level using single cell analyses like single molecule fluorescence *in situ* hybridization (smFISH) for mRNA (Lyubimova et al. 2013); immunofluorescence, microscopy, and flow cytometry for proteins (Wu and Singh 2012). On the other hand, measuring intrinsic and extrinsic noise was not possible experimentally prior to Elowitz et al. (2002). In this landmark paper, Elowitz et al. built two fluorescent reporter constructs driven by the same promoter and integrated into different loci (*galK* and *intC*) in *E. coli* (Appendix V, figure 1 a). Elowitz hypothesize that because the same promoter controls the gene expression of the two different fluorescent proteins, the two reporters

should theoretically co-variate. However, that co-variation may be affected by local biochemical environments around each reporter and thus the ratio of each fluorescent produced may be different between cells. This is defined experimentally as the intrinsic noise and manifest as cells fluorescence different colors due to mixing of the two reporter signals, but the intensity of each cell will be the same (Appendix V, figure 1 c). If the intrinsic noise co-variation is the same, then all other noise should come from the global differences between the cells. This is defined experimentally as the extrinsic noise. It means that the reporter constructs produce the same ratio of proteins but differing amounts between each cell such that the intensity would differ, but the color of the fluorescence would be the same (Appendix V, figure 1 d).

The experimental results of Elowitz et al. show that each individual cell not only produce differing levels of green and red fluorescence despite identical promoters, but also both fluorescent protein levels differs between the cells within the population. This is particularly clear when the flow cytometry data was plotted for individual cells with green fluorescence against the red fluorescence (Appendix V, figure 1 b). The scatterplot shows a positive linear correlation between the green and red fluorescence. The vector that follows the best-fit line is mathematically defined as the extrinsic noise (Swain et al. 2003). This is understood as the green and red fluorescence intensity should co-vary due to the identical promoters. However, because of global differences between each individual cell, different cells may fall somewhere along that vector. Meanwhile, the orthogonal (perpendicular) vector is mathematically defined as the intrinsic noise (Swain et al. 2003). The variation between the two identical promoters is due to local biochemical processes and thus deviates perpendicularly from the extrinsic noise. The total noise is defined as a combination of intrinsic and extrinsic noise (Appendix V, figure 1 e). This total noise appears to be promoter dependent, as seen in (Appendix V, figure 1 b): the noise of

the D22 strain is higher than the M22 strain as they have different promoters controlling the output of the reporters.

Further work has been done to understand why total noise differs in a gene dependent manner. For example, environmental sensing and stress sensing genes have high gene expression noise profiles in both bacterial and eukaryotic systems. Eldar and Elowitz (2010) suggest that this is an evolutionarily driven phenomenon as a means of bet-hedging survival. This is manifested in persister phenomenon: a small population of bacterial cells can survive in nutrient limited or high antibiotic environments because of slight advantages individual cells may have (Grote et al. 2015). Housekeeping genes have a comparably lower gene expression noise level, suggesting that there exists a tight range of tolerable expression levels for essential genes (Richard and Yvert 2014). The low noise level is attributed to the tight regulation of housekeeping genes.

In a paper by McCullagh et al. (2010), they were able to show how transcriptional regulators affected the noise of gene expression. They studied two seemingly redundant mitogen activated protein (MAP) kinase response regulators Dig1 and Dig2. Both inhibited the transcription factor Ste12, which is required for mating. When *dig1* was knocked out, the noise of gene expression increased at Ste12 controlled genes. In addition, formation of discrete nuclear foci of Ste12, overexpression of Ste12 controlled genes, and decreased mating efficiency were observed. Meanwhile, *dig2* knockouts did not exhibit the same phenotype. This suggests Dig1 and Dig2 do not have redundant functions. Rather Dig1 is not only responsible for reducing noise at Ste12, but that noise is not tolerated in the mating pathway.

Gene regulation not only decreases noise, as in the case of housekeeping genes, but also propagates noise. Ornstein et al. (2010) uses that premise to propose a way to screen for novel

regulons (regulation networks) in *S. cerevisiae* by look for noise co-variation between genes. The hypothesis is that co-variation is a surrogate reading for noise propagation through regulatory components as seen in McCullagh et al. Ornstein et al. was not only able to show noise covariation of known regulons to validate this technique but was also able to find new regulon candidates for further experimental exploration.

In mammalian cell systems, noise has been proposed to contribute to cell-fate decisions (Kalmar et al. 2009, Trapnell et al. 2014). Consider a theoretical framework first proposed by Conrad Waddington called the epigenetic landscape (Waddington 1957). Some biophysicists suggest using the epigenetic landscape in applications to understanding cellular differentiation. This landscape consists of many local minima and maxima in which a cell state can exist. A stem cell state is on an unstable equilibrium. Only through differentiation does the cell state move to a stable equilibrium or another unstable equilibrium. In Waddington's theory, he suggests that cell states can change due to "attractor forces" which allow the cell state to "move" from equilibria to equilibria. In recent years, noise has been proposed to act as this "attractor force" (Chang et al. 2008, Pujadas and Feinberg 2012). Chang et al. (2008) demonstrated that high noise in Sca-1 is a mechanism for differentiation of hematopoietic stem cells (HSCs) into the myeloid and erythroid precursors. Using Waddington's model, the noisiness of Sca-1 provides a population with cells that have extremely high levels of Sca-1 and extremely low levels of Sca-1. These cells have increase proclivity to differentiate into erythroid (low Sca-1) or myeloid (high Sca-1). What was even more fascinating was that when Chang et al. sorted high and low Sca-1 populations, the original high noise distribution was re-populated despite initial sorting, suggesting that the noisy distribution is a meta-stable equilibrium.

Introduction to Noise Control and Previous Works

Although technologies are currently available to measure noise, less has been done to construct reliable methods for controlling noise. Thus, previous studies have largely been phenomenological. In this section, I hope to summarize work that has been done on the construction of control mechanisms to better study noise in gene expression.

The Collins lab at MIT was the first to attempt construction of synthetic genetic circuits to control noise in *S. cerevisiae* (Blake et al. 2003, Blake et al. 2006). The first noise control circuit was proposed in the 2003 paper, where the in-series architecture of inducible promoters was introduced. Subsequently, Raser and O'Shea (2004) determined a set of kinetic parameters and circuit architectures that may be used to change the noise of gene expression. In that paper, Raser and O'Shea proposed the usage of targeted mutant screens at the promoter of interest. Blake et al. (2006) applied that idea to improve the initial noise control circuit by combining promoter mutations with the circuit architecture first determined in Blake et al. (2003). Unfortunately, the papers mentioned above were unable to maintain the mean of gene expression, and thus were unable to separate the effects they saw from the change in mean and the change in noise. Another approach taken to change the noise of gene expression is through drug screens. Dar et al. (2014) used a massive drug screen to search for drugs that changes the noise of the Long Terminal Repeat (LTR) promoter of HIV in order to demonstrate how noise changed reactivation of viral latency during dormancy. Although drug screens for noise are valuable, the noise-modulating mechanisms of the drugs are unknown. Thus, it is not possible to conclude that noise is the sole producer of the observed variations in viral reactivation.

Andras-Diaz et al. (2017) mathematically modeled and constructed the first robust experimental noise control system in *S. cerevisiae*. In the paper, they coined the term 'noise

rheostat' as homage to the rheostat's function as a variable resistor within an analogue circuit (Appendix IV). The noise rheostat's circuit architecture consisted of a constitutively expressed inducible transcription factor with its cognate promoter controlling the transcription of a second inducible transcription factor. This, we will call the first node of control. By dosing the ligand of the first node, we can control the availability of the second inducible transcription factor. This second transcription factor controls the expression of a fluorescent reporter that is located downstream of the cognate promoter. This, we will call the second node, in which by dosing the second ligand, we can control the gene expression activity of the terminal output. The output of the system, the gene expression of the fluorescent reporter, is therefore being controlled by the doses of ligands that control the activity of each inducible transcription factor (Appendix V, figure 2). When the fluorescent reporter is measured at steady state, i.e. where the rate of decay is equal to the rate of production, the variation of fluorescent protein levels in the population differs depending on the ligand concentrations.

The variation of the reporter level between individual single cells within this uniformly treated population is called noise. The total noise (combination of intrinsic and extrinsic) is measured as the co-efficient of variation (CV) or CV^2 (Raser and O'Shea 2005). This measurement is a descriptor of how variable the gene expression is within the population. The noise rheostat can control the magnitude of the noise while maintaining the mean gene expression of the population (figure 1). This is rather important because unlike previously built noise control systems, this system was able to separate the effect of changing the mean of expression from the effect of changing the amount of noise.

Andras-Diaz et al. were able to also show that the noise rheostat can control not only the noise of the fluorescent reporter, but also noise of endogenous genes within the yeast genome in

a robust and replicable manner with predictable ligand dose combinations. In addition, they were able to demonstrate that noise control through this device produced a phenotypic change within the cell system. With such results, the device holds incredible promise in the study of noise biology.

Although the noise rheostat appears to function within a yeast cell system, one does not know whether the device may be used within a context of another cell type. Unfortunately, such issue is a prevailing problem in synthetic biology. The kinetic properties may differ between differing cell systems such that each synthetic circuit must be re-optimized through trial and error for new systems (Cameron et al. 2014).

As of late, there have been very interesting questions within mammalian cell biology regarding the necessity of noise for normal biological functions. However, there has yet to be a robust way to control noise within mammalian cell systems. Thus, the goal of this thesis is to construct a functional noise rheostat device to use within mammalian cell lines. Andras-Diaz et al. provides guidelines and engineering parameters for the construction of a noise rheostat in mammalian cell systems.

To build the in-series architecture of the rheostat in Andras-Diaz et al., I will first determine suitable parts that meet the engineering parameters highlighted. I first determined the fluorescent protein reporters being used as the output of the device. Then, I characterized different available transcription systems. The transcription systems for the individual nodes of the noise rheostat must have linearly inducible outputs or have a range of concentrations by which the output is indeed linearly inducible. Finding this range allows us to determine the general range of dosages for the two-node noise rheostat characterization to allow for maximization of data resolution. If the dosages chosen for that experiment are not within the log

linear range, noise control is unpredictable. The transcription systems also must not be affected by mammalian cellular processes and thus must be sourced from non-mammalian origins but optimized for mammalian cell system usage. Lastly, I will build the in-series circuit using the transcription systems the meet the engineering criteria above and characterize its dynamic properties to determine whether a noise rheostat could be built with the nodes chosen within mammalian cell lines.

MATERIALS AND METHODS

Bacterial strain and growth media

Liquid culture of *E. coli* DH5α were grown in Luria Broth (recipe in Appendix 1) at 37º C. The following antibiotics were added depending on the cloning vector: carbenicillin (100mg/L), chloramphenicol (25 mg/L), or kanamycin (50 mg/L).

Golden Gate Assembly and Transformations

A Golden Gate® Assembly was prepared with the following reagents: 0.3 μL of T4 ligase (New England Biolabs), 1 μL of T4 ligase buffer (New England Biolabs), 0.5 μL of restriction enzyme, 0.5 μL of each parts plasmid or GFP dropout entry vector (each parts plasmid concentration was normalized to 25 fmol and entry vectors normalized to 50 fmol), and brought to a final volume of 10 μL with PCR grade water. The restriction enzymes used were BsaI (10,000 U/mL from New England Biolabs) or BsmBI (10 U/μL from ThermoFisher Scientific).

Assemblies were incubated in a thermocycler with the following program: 25 cycles of digestion and ligation (37º C for 1 min, 16º C for 2 mins) followed by a final digestion step (42ºC

for 10 mins) before an enzymatic heat inactivation step (80° C for 10 mins). Assemblies were subsequently either stored at -20° C or used directly for bacterial transformations.

Two microliters of the resulting assembly reaction were added to 50 µL chemically competent *E. coli* DH5α cells (lab stock, full protocol in Appendix 1). Cells were left on ice for 30 mins before 1 min heat shock at 42° C. These cells were either directly plated on their respective antibiotic plates (carbenicillin or chloramphenicol) or recovered (for kanamycin resistance vectors) in 200 µL of SOC media (Thermofisher Scientific) for an hour at 37° C and plated onto kanamycin antibiotic plates.

Plasmid bank of plasmids used and built are in the Appendix 2.

DNA Construct Verifications

Individual white colonies from each plate were inoculated into Luria Broth with necessary antibiotics and grown for 16 hours at 37°C. Subsequently, they were pelleted for plasmid purification using a miniprep kit (Invitrogen). Plasmids concentrations were measured using NanoDrop™ 2000 (Thermofisher Scientific). 1µg of DNA was used for restriction digestion and incubated for 37°C for two hours. Digestion products were run on 1.5% agarose gel with 1x SYBR Safe DNA gel stain (Invitrogen) at 130 V for 30 mins. Gels were imaged on a BluView Transilluminator (Stellar Scientific) and digestion patterns were verified for correct lengths. Plasmids were subsequently retransformed and plasmid extractions were done using a midiprep kit (Invitrogen) to obtain purer plasmids for mammalian cell transfection.

Cell Culture and Transfection

HEK293T cells were grown at 37ºC with 100% humidity and 5% CO_2 in Dulbecco's Modified Eagle's Medium (DMEM, Gibco). DMEM was supplemented with Antibiotic-Antimyotic (From Gibco: 100 units/mL of penicillin, 100 µg/mL streptomycin, 250 ng/mL Amphotericin B) and 10% fetal bovine serum (FBS, HyClone™).

Transient Transfections

0.5 x 10^6 HEK293T cells were seeded on 6 well plates, cultured with 2 mL DMEM, and grown for 24 hours to 70-90% confluence. Cells were then washed with Phosphate Buffered Saline (PBS, Gibco) and 0.5 mL of Opti-mem™ Reduced Serum Medium was added. For each unique plasmid, 10 uL of Lipofectamine 2000 was incubated with 3µg of DNA in 0.5 mL of Opti-mem™ for 25 mins and subsequently added to the wells. The media was removed, cells were washed with PBS, and media was replaced with DMEM after 24 hours. Cells were subsequently harvested and analyzed using flow cytometry and microscopy.

Stable Cell Line Construction

For construction of stable cell lines, similar methods were used as in transient transfection. However, each unique construct was added with a PiggyBac transposase expression vector in the 0.5 mL Opti-mem™ 10 uL Lipofectamine mixture. This mixture was added into the six well plate containing seeded HEK293T cells in 0.5mL Opti-mem™ and incubated for 24 hours before removal of transfection medium and replacement with DMEM. Cells were incubated in 48 hours in DMEM and subsequently supplemented with 300 ug/mL with hygromycin or 200 ug/mL blasticidin. Clonal cell lines were construct using flow cytometry assisted cell sorting for single cell sort and clonal expansion.

Activation and Fluorescence Measurements by Flow Cytometry

The one node and two node experiments were activated with their respective small molecules 72 hours post transfection. A BD^{TM} LSR II flow cytometry was used to obtain all fluorescence readings. To determine steady readings, fluorescence readings were obtained 24 hours, 36 hours, 48 hours, and/or 72 hours post activation. 10,000 events were collected and gated for live and singlet cells using forward and side scatter. The following lasers were used to obtain readings for green fluorescent protein (GFP), red fluorescent protein (RFP), and tagged blue fluorescent protein (TagBFP): 488nm, 561nm, and 405nm respectively.

Confocal Microscopy for Landing Pad Cell Lines

Landing pad cell lines co-transfected with insertion vectors and recombinase expression vectors were imaged seven days after transfection. Cells were plated onto microscopy compatible 35 mm plates and allowed for adhere for 24 hours prior to imaging. Using a spectral confocal microscope (Nikon) with laser lines at 405nm, 488 nm, 561 nm, cells were imaged for blue, red, green fluorescence.

Data Analysis

All data analysis was done using MatLAB R2017a flow cytometry analysis package. All populations were gated prior to analysis. Gating parameters and statistical analysis equations are found in Appendix 3.

RESULTS

Verification of functional fluorescent markers constructs

To build the genetic circuits proposed in the introduction, the constructs required the usage of at least three functional fluorescent markers whose emission spectra did not overlap between output channels. In this study, we chose to use three different fluorescent markers: enhanced green fluorescent protein (GFP), red fluorescent protein (RFP), and blue fluorescent protein (BFP). The excitation and emission spectra of each fluorescent protein (Appendix V, figure 3) had the least overlap of any three combinations of commercially available fluorescent proteins. These markers enabled us to track single cell activity using flow cytometry and microscopy. Prior to the construction of full length genetic circuits, all fluorescent markers genes tagged with localization signals were driven by constitutive promoters and stably integrated in HEK293T cells. This was to ensure that the fluorescent proteins were expressed properly and would be visualized using flow cytometry and microscopy.

This cell line was plated and visualized on LSM 510 laser scanning microscopy (Zeiss) at a 40x objective using excitation and emission wavelengths that was optimized to separate of each signal. The images were subsequently analyzed on ImageJ and showed cells that expressed all three fluorescent markers (figure 2), albeit at different intensities due to the usage of a transposon-based integration technique. The PiggyBac^{TM} transposase integration technique did not control for copy number or location, so some cells had more integrated constructs than others. Furthermore, some constructs may have integrated at transcriptionally silenced sites, which contributed to additional variability in the fluorescence. Both the red fluorescent protein and green fluorescent protein had nuclear localizations tags. Thus, the microscopy images showed cells with discrete red and green nuclei (figure 2 b and c). Meanwhile, the blue

fluorescent protein was tagged with a plasma membrane tag. The microscopy images showed the characteristic HEK293T cell shape (figure 2 a) fluoresce blue.

Transient Transfections and Plasmid Copy Number Dependent Fluorescence

To test genetic circuits within mammalian cells, the DNA vector must be introduced into the cells. There were many different methods of introducing DNA into mammalian cells, each with pros and cons. We chose to transiently transfect our cells because transient transfections were quick and allowed for rapid characterization of many different constructs. There were limitations to using transient transfections, including an inability to tightly control for DNA copy number and loss of the introduced DNA through dilution and degradation. Controlling DNA copy number is particularly important as Schindelhauer and Laner (2002) demonstrated that protein expression from a transient transfection was plasmid copy number dependent, which could have affected our measurement of noise.

To ensure that the outputs of the circuits tested were not simply the effect of plasmid copy number variability, a blue fluorescent marker was used in the DNA vector backbone (figure 3). A constitutive cytomegalovirus promoter drove the expression of this blue fluorescent protein. We chose the CMV promoter due to its high activity and low variability (Dandekar et al. 2005). This fluorescent marker not only served to distinguish between cells with and without the transfected plasmid but was also used to normalize for plasmid copy number variability.

Fluorescence of un-induced transiently transfected populations was measured for every plasmid built in this study to understand plasmid copy number variability. Flow cytometry measurements of green and blue fluorescence were $\log_{10}$ transformed and plotted as scatter plots (figure 4, Appendix V: figure 4 and 5). Each circuit showed basal activity, such that some green

fluorescence was observed despite being un-induced by their respective inducer ligands. There was a positive correlation between the blue fluorescence, the plasmid copy number control, and basal green fluorescence, the terminal output for every circuit. Linear regression was also performed on all scatterplots, and p-values obtained were all $p < 0.00001$. This suggests that there was a statistically significant positive linear relationship between the log transformed blue and green fluorescence intensity. The lines obtained from the linear regression were used to distinguish between low and high basal activity of the un-induced populations. The slopes of the lines were good surrogate measures of basal activity of the synthetic transcription systems tested transiently such that the higher the slope, the higher the basal activity.

The two synthetic transcription systems characterized were the abscisic acid chemically induced proximity (ABA-CIP) and the TetOn3G. The slope of the line for the ABA-CIP construct is 0.647 (figure 4), which suggested that the transcription system had lower basal activity compared to the TetOn3G whose slope was 1.046 (Appendix V, figure 4). The basal activity of the two-node in-series construct was between that of the ABA-CIP and TetOn3G. The slope of the linear regression was 0.853 (Appendix V, figure 4). This made sense as the basal activity of the two-node construct could not be higher nor lower than the constituent parts.

Abscisic Acid One Node Induction

To build the two-node noise rheostat, different one node transcription systems were characterized to determine their suitability in terms of the engineering criterion highlighted in the introduction. The Abscisic Acid (ABA) Chemically Induced Proximity (CIP) transcription system was characterized with such goal in mind. ABA-CIP was a synthetic transcription system constructed by Liang et al. in 2011. Borrowing proteins from a plant stress signaling pathway,

Liang et al. took advantage of the natural complex formed between the pyrabactin resistance regulatory component (PYL1) of the ABA receptor, phytohormone S-(+)-abscisic acid (ABA), and an ABA insensitive 1 protein (ABI1). This PYL1-ABA-ABI1 (PAA) complex (figure 5 a) was utilized as the starting points for the construction of a synthetic transcription system as it allowed for chemically induced proximity to occur. In our genetic circuit, a Gal4 DNA binding domain (GAL4DBD), which bound to an upstream activation sequence (UAS), was linked to the ABI1 protein. A herpes simplex virus VP16 transactivation domain (VP16AD) was linked to the PYL1 protein. Both fusion proteins had a nuclear localization tag on the C-terminus. With the addition of ABA, the formation of PAA complex brought together a protein that recognizes a specific DNA sequence (GAL4DBD-ABI1) and a protein that acted as transcriptional activator (PYL1-VP16AD). A cognate promoter for the ABA-CIP was constructed using 5 UAS upstream of a minimal CMV promoter called pUAS (figure 5 b). The GAL4DBD-ABI1 protein was constitutively bound to UASs on pUAS. Upon addition of ABA, the PYL1-VP16AD protein formed a complex with the GAL4DBD-ABI1 and recruited transcriptional machinery.

The advantage of using this system was that the plant stress pathway did not have a homologous structure in mammalian cells and thus should not have crosstalk with mammalian proteins. In addition, ABA was non-toxic and inexpensive, allowing high dosages to be used. Prior characterizations of this inducible transcription system used luciferase assays, which showed that the ABA CIP system had a large fold change and ABA dosage dependent log-linear gene expression. However, direct expression of a fluorescent protein as the output was not assayed by Liang et al.

Thus, we characterized the effects of ABA dosage on the direct expression of the green fluorescent reporter in our circuit. A cytomegalovirus constitutive promoter (pCMV) was used to

drive the expression of a polycistronic GAL4DBD - ABI1 and PYL1 – VP16AD. Upon the addition of ABA to cells with functional GAL4DBD-ABI1 and PHYL1-VP16AD, the green fluorescent protein (GFP) downstream of the pUAS was expressed (figure 5 c). This genetic circuit was transiently transfected into HEK293T cells and characterized using a dosage experiment. Approximate 50,000 cells per dosage were induced with ABA for 48 hours. Cells were subsequently trypsinized and fluorescence was measured using flow cytometry. The flow cytometer collected 10,000 events and gated for cells with blue fluorescence thresholds above $1x10^3$ fluorescence units. Approximately 5140 ± 140 (mean±S.D.) cells were analyzed per dose of ABA. Histograms were plotted with the green fluorescence intensity in $\log_{10}$ scale on the x-axis and the fraction of cell population that had the intensity plotted on the y-axis (figure 6 a). By normalizing the cell numbers to obtain fraction of total cell population, we were able to compare differing population sizes. By normalizing by blue fluorescence intensity, we controlled for plasmid copy number variation.

The normalization of each cell's green fluorescence by the blue fluorescence of the cell was done by taking the ratio of the green fluorescence to the blue fluorescence (G/B). This ratio was subsequently log transformed to obtain a normal distribution for Gaussian statistics. The arithmetic means of the log transformed G/B increased as ABA dosages increased. This phenomenon was observed as an x-axis right shifting of the bell curve (figure 6 a). The mean G/B did not significantly change from the uninduced control population shown on the top-left most histogram until ABA dosages were increased to 10^2 nM. As dosage was increased, increases in the standard deviations of the histograms were also observed. One point of note while reading the histogram of the dosages equal to or greater than 10^2 nM, the distributions looked as if a second distribution began to emerge out of the un-induced distribution. This effect

was the most prominent at 10^3 nM ABA dose, such that the distribution looked like an overlap of two normal curves. This bi-modality suggested that there was two distinct "on" and "off" populations as illustrated by the cartoon green and red normal curves superimposed onto the histograms with observed bi-modality. As the ABA dosage increased, the "off" population became smaller as the "on" population became larger. At the highest ABA dosage (10^6 nM), the "off" population was still observed, but the tail was much smaller than the "on" population. A gaussian mixture model (GMM) analysis was subsequently used to quantitatively determine the effect of the bimodality on the mean of the "on" and "off" population and on the mean of the total population (Appendix V, figure 6 a). The mean of the "on" population did not differ significantly from the mean of the total population.

Although bi-modality was observed, the mean fluorescence measurement of the total population was an adequate surrogate fluorescence measurement of ligand dosage's effect on population behavior. Thus, we plotted the dosage of ABA against the mean $\log_{10}$ G/B of the cell population at that dosage (figure 6 b). The log-linear portion of the mean G/B was between 10^2 – 10^4 nM and encompassed an 8.8 ($10^{0.93}$) linear fold green fluorescence change. Such log-linearity over a 100-fold dosage change was important because it ensured that the ABA-CIP allowed for fine tune control of mean gene expression in a dosage dependent manner. This range also allowed us to determine the general range of dosages for the two node noise rheostat experiments in order to maximize data resolution. A similar plot was constructed using GMM analysis of the "on" and "off" population (Appendix V, figure 6 b). The "on" population mean data looked similar to the overall mean data. Thus, using this construct despite a bimodal output would have met the engineering criterion listed in the introduction.

Doxycycline One Node Induction

A second node must be characterized to build the two-node rheostat. Thus, we characterized the TetOn3G (ClonTech) transcription system. This system was a commercially available inducible transcription system which boasted a large fold induction, low basal activity, and high inducer sensitivity. Although doxycycline, the inducer of the system, was toxic at high concentrations - the high sensitivity of the transcription system allowed for activation concentrations below toxic levels.

To test whether this system behaved as expected in HEK293T cells, we constructed the one node circuit construct shown in figure 7 a for characterization. A pCMV constitutively expressed the transcriptional activator TetOn3G. This TetOn3G considered of a bacterial Tet repressor (TetR) fused with three minimal VP16ADs (figure 7 b). In bacteria, the TetR protein bound to the Tet operator (TetO) sites and repressed transcription unless the bacteria encounter tetracycline or related antibiotics such as doxycycline. In contrast, the synthetic TetOn3G bound to TetO if and only if it was bound to the ligand doxycycline. The cognate promoter (pTet) consisted of seven TetO sequences upstream of a modified CMV promoter. Thus, when doxycycline was added, TetOn3G bound to the TetO on pTet and promoted expression of the green fluorescent protein.

To test the dose dependently activation of the TetOn3G system, we used the one node circuit to perform a dosage experiment to characterize the induction dynamics of the TetOn3G transcription system. The 50,000 cells per dosage were induced for 48 hours prior trypsinization and flow cytometric readings. After gating for cells with greater than or equal to $1x10^3$ blue fluorescence units, 3002±2059 (mean±S.D.) cells were obtained per dosage. After normalization by both percentage of cells and by blue fluorescence, the G/B values were plotted as histogram

distributions (figure 8 a). Like the ABA-CIP one node construct, the TetOn3G circuit resulted in higher G/B ratios as the dosage of the inducer increased. This was exhibited by a right shift of the population across the x-axis. As expected, a low concentration of doxycycline was needed to turn on transcription at the pTet promoter. At only 0.1 ng/mL of doxycycline, the mean G/B increased by more than two-fold ($10^{0.43}$) from basal levels. In addition, like the ABA-CIP dynamics, there existed a concentration in which the transitioning dynamics showed bifurcation of two distinct populations. At 1 ng/mL (10^{0}) of doxycycline, the curve appeared to be two log-normal curves overlapping. For concentrations greater than 1 ng/mL, the "off" population became smaller as the "on" population became larger. The remnant of the "off" population was observed as the tail of the 10^{2} ng/mL dosage population. A GMM analysis was performed to determine the two overlapping gaussians (Appendix V, figure 7 a). The mean of the "on" population was found to be comparable to the mean of the total population, and thus only the mean of the total population was used to simplify further analysis.

Using the means of the distributions as a surrogate reading of population level changes, doxycycline dosage was plotted against mean log10 G/B (figure 8 b). The log-linear portion of the graph was observed between the dosages of 0 – 10^{1} ng/mL which encompassed over a 30 ($10^{1.5}$) fold increase from the basal activity. The low concentration of doxycycline ensured reduction cellular toxicity induced by doxycycline saturation. The plot also ensured that TetOn3G did not exhibit toggle switch like behavior but rather, dosage dependent log-linear induction. The means of just the "on" population was also plotted and appeared to show the same exact behavior as the means of the total population (Appendix V, figure 7 b).

Performing the TetOn3G characterization allowed me to determine that this TetOn3G transcription system was indeed suitable for use in the two-node in series construct as the TetOn3G appeared to be linearly inducible.

Noise Rheostat Two Node Characterization and Induction

After characterizing the inducible transcription systems, we determined that they were suitable for the use in the two-node circuit construction as each met the engineering criterion set out in the introduction. Each system was log-linearly inducible, unimodal at high concentrations, and did not appear to exhibit cross talk. The next step was to build a two node in-series circuit (figure 9) using pCMV-ABA-CIP as the first node and pUAS-TetOn3G as the second node. The reporter pTet-GFP reported for the activity of the two nodes in-series circuit as the terminal output. A pUAS-RFP was used as a reporter for ABA-CIP induction and thus a reporter of the first node. This was to ensure that induction of ABA-CIP through the addition of ABA indeed drove gene expression of genes downstream of pUAS. Differential dosages of ABA and doxycycline controlled the availability of active ABA-CIP and TetOn3G protein, which determined the gene expression levels and noise of the green fluorescent protein.

To test whether the in-series architecture of the circuit described in yeast also worked in a mammalian cell system, we tested the two-node in series circuit using differential dosages of the two inducer ligands. Using a 96 well plate, a dosage experiment was set up (figure 10). The dosage of ABA was varied across the width of the plate (twelve wells), with dosages that ranged from 0 to 10^5 nM. The doxycycline dosage was varied down the length of the plate across eight wells, with dosages that ranged from 0 to 10^3 ng/mL. The well on the top left most corner served as a control un-induced population used to measure basal activity of the reporter promoters

pUAS and pTet in the circuit. The first row was the ABA-CIP control, with the pUAS-RFP reporter giving an output reading for the first node of the in-series circuit. The first column measured the basal activity/leakiness the first node, but particularly the leakiness of the pUAS. If red fluorescence was observed despite induction of the first node using ABA, it suggested that TetOn3G was also being produced because of leakiness at pUAS. Thus, if green fluorescence increased as doxycycline dosage increased along the length of the first column, this suggested that the TetOn3G produced from the leaky pUAS was activating and driving expression of GFP.

The 96 well plate of transiently transfected cells was induced, and flow cytometric readings were obtained at 24 hours and 72 hours post induction to find steady-state readings. To measure the activity of the first node, activity of the terminal node, and plasmid copy number, the red, green, and blue fluorescence measurements, respectively were obtained for cell populations in each well. The fluorescence measurements were subsequently converted to $\log_{10}$ units and normalized to obtain the red/blue fluorescence ratios (R/B), and green/blue fluorescence ratios (G/B). Using the normalized fluorescence, the mean fluorescence and geometric coefficient of variation (CV) of each dosage combination was calculated. The geometric CV was a measurement of variability of a normal curve formed by log-transformed data and was the noise readout of the two-node circuit.

The mean G/B and R/B at 24 hours post induction and 72 hours post induction was then plotted on the x-axis and the y-axis, respectively (figure 11). The blue y = x line was also plotted and used as a reference line. This reference line allowed us to compare the two populations. If the fluorescence at 24 hours and 72 hours were equal, all the points would have laid on the line. This would have suggested that the circuit activity had reached steady-state such that the cells produced and destroyed the same amount of fluorescent protein. However, as all the points laid

above the reference line, it suggested that at 24 hours, steady state was not reached because there was more still more protein being produced at 72 hours. As the 24-hour readout was much further from steady state, all subsequent analyses were done on the 72-hour readouts

The activity of ABA-CIP was reported by a red fluorescence reporter driven by pUAS. The histogram distribution of normalized red fluorescence (log_{10} of red fluorescence/ blue fluorescence (R/B)) was plotted in the 96 well format (figure 12 a). All distributions observed were unimodal, which was surprising considering data from the one node ABA-CIP circuit. However, this was evidence to support that the bi-modality observed at 48 hours in the one node circuit (figure 6 a) and the R/B values at 24 hours in the two-node in series circuit (data not shown) were artifacts of transitioning dynamics. The unimodal distribution allowed us to calculate the mean R/B of the cells that accurately represented the average behavior of the population. A heatmap of those means was plotted (figure 12 b) and showed that the mean R/B as ABA was increased across the width of the plate. Interestingly, mean R/B also varied down the length of the plate but to a lesser extent than the width of the plate. As doxycycline concentrations increased, the mean R/B decreased by a small amount. This may have been an artifact of oversaturating and overworking the cellular system at high concentrations of both inducers and will be addressed in the discussion section. The CV of R/B measurements did not dramatically change across the different dosages of ABA or doxycycline (figure 12 c). However, a clustering of higher CVs was observed for wells on the top right 4x4 wells. The higher CV in those wells may have been dued to the long tails observed at those distribution (figure 12 a). For the most part, the CVs across the plate was consistently under 2 units. The highest CV observed from R/B values was around 6 units. The cause of such variation with the top right 4x4 well space was unknown.

The terminal node, pTet-GFP, reported for the activity of the entire two-node in series circuit. The green fluorescence was normalized by the blue fluorescence (G/B) and $\log_{10}$ transform. The histogram distribution across the 96 wells shows most populations had unimodal distributions (figure 13 a). Bimodality was observed for some populations. Specifically, as one looked across the width of the plate, most of the bi-modal populations were observed between ABA concentrations of 1×10^2 nM – 5×10^4 nM and doxycycline concentrations between 5×10^1 ng/mL – 1×10^3 ng/mL. The bi-modality was not present in higher concentrations of ABA, leaving only a tail at 1×10^5 nM ABA. Most of the bimodality for the intermediate drug dosages could have been due to simply transitioning dynamics. Meanwhile, the last row with the highest doxycycline concentration appeared to maintain a bi-modality even with a saturating ABA. It had been previously shown that a 10^2 ng/mL concentration of doxycycline (Ahler et al. 2013) was enough to drastically alter the metabolic states of cells. Above this concentration, cells were starting to exhibit cellular toxicity and other symptoms of altered metabolism. The bimodality could be explained as two clearly distinct populations of unhealthy cells and thus less so an indictment of a malfunctional circuit. Other hypothesis will be further discussed in the discussion as to why such a phenotypic presentation was observed.

The mean of the G/B was calculated and plotted as a heatmap (figure 13 b). This heatmap clearly showed that G/B values were dependent on both the ABA and doxycycline concentration. As one increased the concentration of ABA and doxycycline along the width and length of the plate, respectively, a mean G/B increase was observed. What was also observed was that different combinations of ABA and doxycycline produced very similar means which form isomean curves across the plate as highlighted by (figure 13 b).

Although these population had the same means, the population gene expression noise as represented by the geometric CVs, differed between these wells within the isomean curves (figure 13 c). This was particularly clear when plotting the individual distributions of the isomeans wells (figure 14). Although all the populations were centered at the same mean, low (blue curve), medium (yellow curve), and high (red curve) noise levels were achieved by using different drug dosages to induce or "dial" the two nodes. This behavior was particularly of note as it was similar to the theoretical noise rheostat shown in figure 1 such that differing noise levels could be achieved while maintaining the same mean. In figure 14 a, low and high noise levels were achieved by maintaining mean around -0.93. In figure 14 b, low, medium, and high noise levels were achieved while maintaining the mean at around - 0.73. In figure 14 c, medium and high noise levels were achieved while maintaining the mean at around - 0.60.

The GFP histogramic distribution showed the importance of the two node in-series architecture and its necessity for controlling noise. When the geometric CV (noise) of G/B was plotted against the CV of R/B with a y = x reference line, the graph suggested that the noise of G/B had a much larger range that R/B (figure 15). As R/B was the reporter for the first node, ABA-CIP, this suggested that ABA-CIP was not able to produce the full range of noise levels (CV range : 2-6) compared to the two node in series circuit given as G/B values (CV range: 2-14). In addition, there appeared to be a clustering trend dependent on concentrations of both ABA and doxycycline. At low concentrations of doxycycline, exhibited as the blue points in figure 15 a, the noise of R/B was more varied. Meanwhile, at higher concentration of doxycycline, the noise for the R/B was low and less varied. In addition, the higher concentrations of doxycycline, shown as points in red and orange, span the range of the noise of G/B. For those same points, the ABA concentrations were green, yellow, orange and red (figure 15 b), which

indicated intermediate concentrations (10^2– $10^{4.5}$), but no observable clustering was discerned. This makes sense as the terminal node was directly under the control of TetOn3G, and thus was more sensitive to doxycycline than ABA.

By characterizing the two-node circuitry built using the ABA-CIP and TetOn3G system, I was able to determine that the circuitry could indeed modulate noise through differential combinations of ligand dosages. Such was particularly clear in figure 14. The mean fluorescence of the cells treated with differential dosage combinations were very similar while the noise of those same cells differed. Thus, we were able to control the noise of the system through the de-convolution of the mean from the noise, suggesting that this two-node in series circuit was indeed a noise rheostat.

Landing Pad Construction and Characterization

As transient transfection requires normalization to control for DNA copy number, it would have been better to use a system where this normalization was unnecessary as normalization calculations may have overlooked important but difficult to detect behaviors of the genetic circuitry. Using recombinases with very specific recombination sites, we took advantage of targeted insertions. We created a "landing pad," defined as a specific integration site within a known locus with characterized activity. This way, we created a place within the genome for directed integration to control for construct copy number per cell and prevent location specific transcriptional silencing of our circuit of interest.

We constructed the landing pad using CRISPR-Cas9 mediated homologous recombination in HEK293T cells to integrate the landing pad into the hRosa26 locus. This landing pad was illustrated in the circuit diagram in figure 16. The circuit had HS4 insulators

flanking the landing pad and prevented transcriptional activity at the landing pad from transferring activity to the genes around the landing pad and vice versa. The landing pad had a constitutive human Pgk1 promoter driving a polycistronic cassette with a resistance marker for hygromycin and a red fluorescence marker with a nuclear localization signal (mRuby). The recombination site, *attP*, which was directly downstream of Pgk1 provided a way to interrupt constitutive expression of the red fluorescence, which created an observable phenotype that signalled successful insertion at the landing pad. A corresponding insertion vector was constructed with another polycistronic cassette with a resistance marker for blasticidin and blue fluorescent protein. This cassette did not have a promoter upstream and thus would be transcriptional inactive unless integration of the insertion vector placed the cassette downstream of the Pgk1 promoter. Landing pad cells were red when there was no insert, and blue when the insertion vector was properly inserted. Two *attP*/*attB* recombinase pairs (BxBI and PhiC31) were tested to look at the efficiency of integration.

The landing pad cells were single cell sorted using flow cytometry assisted cell sorting (FACS) and clonally expanded. The landing pad recombination was verified using PCR. To test whether the landing pad/insertion vector pair effectively created stably integrated cell lines, a constitutive green fluorescence reporter was subcloned into each insertion vector. The reporter constructs were then co-transfected with their respective recombinase expression vectors. Cells were subsequently imaged seven days after transfection using confocal microscopy (figure 17). Unfortunately, the blue fluorescence was not observed, suggesting that the *attP* site was in the wrong orientation and thus recombinase assisted recombination was also in the wrong orientation. Thus, the promoter-less polycistronic blasticidin blue fluorescent cassette did not gain the transcriptional activity of the Pgk1 constitutive promoter of the landing pad. Both BxBI

and PhiC31 *attB* landing pads appeared to accept insertion vectors (figure 17 a,b), such that the cells with the insert no longer expressed the red fluorescence protein of the landing pad. Some cells appeared to only express the green fluorescent reporter, suggesting that the insertion vector properly inserted at the *attB* site. Some cells, particularly in the *PhiC31attB* landing pad cells, appeared to express both red and green fluorescence (figure 17 b). This suggestsed that the insertion vector was inserted into another location in the genome as the cells appeared to retain red fluorescence despite gaining the green reporter. There was a possibility that those cells had un-integrated green reporter construct, but this was highly unlike at day seven due to cell division and degradation of foreign DNA. In the *BxBIattB* land pad cells, there was a cluster of green cells surrounded by red cells. It was very possibly that the cluster of green cells originated from one cell that had a successful transfection.

DISCUSSION

Characterizing reporter constructs

To build the noise rheostat circuit, the parts of the circuit must be characterized within the cell line prior to the full construction. In the beginning, the quantitative tool used for the project was flow cytometry and fluorescence microscopy, thus different fluorescent proteins reporters were characterized within HEK293T cells to ensure the gene sequence not only produced functional protein within the cells but to ensure that there were no aberrations that may form such as inclusion bodies consisted of misfolded proteins. Certain fluorescent proteins, particularly mRuby, have been shown to have a higher rate of forming aggregates, seen as localized foci through the cell body. In addition, localization tags were tested to ensure proper localization for masking purposes during quantification of fluorescence using microscopy. The fluorescent proteins and localization tags for this study have shown proper function and were subsequently used for cloning the larger circuits.

Transient Transfections

Transient transfections are not particularly ideal for studying non-genetic noise but provides a way in which one can quickly prototype genetic circuits within cellular systems. Because transient transfections don't allow for one to control the number of copies a cell may receive, the availability of the vector has been shown be directly correlated to the level of proteins that vector expresses. In addition, transient transfections require a reporter construct that allows for the differentiation between cells with and without vectors. As we are using flow cytometry, the usage of a constitutively expressed fluorescent marker like the blue fluorescent protein, can be utilized on the vector backbone to ensure that the cells analyzed, and the

behaviors observed are due to the genetic circuit transfect and not due to other factors unrelated to the circuit.

There are ways to control for copy number dependent gene expression levels. One way that has been used is by normalizing one's experimental measurements with expected measurements within that same cell. Using the same blue fluorescent protein controlled by a constitutive high activity promoter like the cytomegalovirus promoter (CMV), the other fluorescent proteins expressed on that same vector may be normalized by the blue fluorescence levels. Although the CMV promoter has previously been shown to be at risk of transcriptional silencing through histone modifications and methylation when integrated into the genome, and thus may not be a reliable promoter for reporting copy number when integrated, such is not the case for transient transfections.

The assumption that a high blue fluorescence level is the direct result of high copy number and this the other fluorescence levels are similarly due to the copy number is deceptive and oversimplified. Noise of the system is still observed, albeit at a worse resolution and based on oversimplified assumptions, transient transfection nonetheless provides a quick and easy way to rapidly prototype. If the circuit does vaguely what one might expect, a stable and clonally expanded cell line must be characterized. Unfortunately, for circuits that are not particularly high activity or robust, it's behavior may be undetectable.

Characterizing Synthetic Transcription Systems in HEK293T Cells

A consistent problem within synthetic biology has been the transpose of synthetic constructs and circuits from one cell type to another (Cameron et al. 2016). The dynamic properties within a cell line like HeLa differ from HEK293T because of their cell state (i.e.

cancerous vs. immortalized stem cell), contributing to much of the extrinsic noise of a circuit. Even the same cell line across different labs differ tremendously. Thus, although all the synthetic transcription systems used in this study have been previously characterized – the dynamics of this proteins must be characterized in my particular cell line to allow for accurate understanding of the dynamics in the full synthetic circuit of the two-node system.

During the characterization process, flow cytometry data for the highest dosages of doxycycline was not plotted in the dosage experiment figures because of aberrant behavior. It appeared that measure fluorescence at 10^3 ng/mL was overall much lower than the 10^2 ng/mL, and even of 10^1 ng/mL. This may be due to a shift in the gating by the flow cytometer during the measurement itself, a phenomenon observed previously for some other measurements obtained, but those were recorded and correctly removed. The dramatic dip in the fluorescence may be also due to the toxicity of doxycycline at such a high concentration. In had previously been demonstrated that concentrations above 10^2 ng/mL of doxycycline cause alter metabolic states in HEK293Ts (Ahler et al. 2013). The cells induced with doxycycline concentrations above 10^2 ng/mL may be simply too unhealthy to allocate resources to production of the transcription factor and fluorophore reporter.

The ABA-CIP transcription system showed a similar dip observed at the highest concentration of abscisic acid (10^6 nM). The dip at 10^6 nM concentration is also observed in the paper by Liang et al. (2011), which the characterization of this transcription system was done in Chinese Hamster Ovary cells (CHOs). The scientists who originally synthetically constructed this transcription system did not make a comment on the dip. This may suggest that although abscisic acid is comparably less toxic than doxycycline and thus can be used in high quantities, high concentrations of abscisic acid may be altering cellular metabolism.

Addition when one looks at the ABA-CIP transcription system (Figure 6 a), it almost looks like the distributions have a bimodal distribution after 10^3 nM abscisic acid concentration. Cartoon log-normal curves were overlaid to demonstrate overlapping bimodal curves that form the histogram distributions observed. More interestingly, although such bimodality is observed at 48 hours in the one node construct, the same induction for a subsequent experiment with the ABA-CIP controlling red fluorescent protein levels and TetOn3G transcription factor levels, the flow cytometric reading shows that the bimodality no longer appear. This suggests that at 48 hours, the one node construct had not yet reached steady state, and thus the bimodality observed is an artifact of transitioning dynamics. A time course experiment to determine steady state for ABA-CIP was too materials costly – unfortunately there were not enough cells at the times of the experiments to do multiple time points. Thus, the bi-modality for the ABA-CIP, which was originally problematic, appeared to simply be some transitioning dynamics issues rather than an inherent property of the system itself, and thus allows its usage in the two-node noise rheostat construct.

Characterizing the Two Node Noise Rheostat Construct

The data from the two-node noise rheostat is complicated and difficult to process. The first inducible node, the ABA-CIP drives the expression of a red fluorescent protein and the inactive TetOn3G. The experiment for the 96 well plate (figure 10) increase the concentration of abscisic acid across the width of the plate while doxycycline is increased down the length of the plate. For the first node, the measured output, red fluorescence should only vary across the width of the plate. While one does see that mean red fluorescence increase across the width of the plate (figure 12 b), one also sees the red fluorescence decreases as the doxycycline concentrations

increases down the length of the plate. Two possible explanations I could think of is (1) altered metabolism after 10^2 ng/mL doxycycline due to cellular toxicity induced by the high concentration of doxycycline (2) over expression of transcription factors and fluorescent proteins at high concentrations of drugs causes exhaustion of transcription and translation machinery and may divert the metabolic capacity from doing normal cellular functions. In both cases, the cells are not particularly healthy at these high expression levels. Thus, a stable cell line with a known copy number must be constructed to ensure that the experimental data observed in the transient and overexpression model is conserved.

A second layer of complication is the time in which the circuits reach steady state. I don't think that the construct has reached steady state at 72 hours because although the first node no longer exhibits transitioning dynamics, the second node that is controlled by the TetOn3G shows some bimodal distributions at the highest levels of doxycycline. This bimodal distribution is being not observed in the one node TetOn3G, which readings obtained were at 48 hours. Thus, I am unsure that the bi-modality is a transition state dynamic or whether the bi-modality is a real feature of the circuit. A reading at 96 hours would have been much needed, but unfortunately, I ran out of both materials and time.

The green fluorescence does indeed increases as both the concentration of abscisic acid and doxycycline increases. The geometric co-efficient of variation, (geo CV) was used as a surrogate measurement of noise, normally simply the linear co-efficient of variation. Because of the normalization of the data using the copy number data obtained from the blue fluorescence, the green fluorescence is no longer a real fluorescence count but a ratio between the green and blue fluorescence. This ratio unfortunately does not allow for the usage of a linear CV because

the means obtained from the ratios are very close to zero in the log linear data set, making the CV incredibly large and not particularly reflective of the dataset itself.

For what the data is right now, it nonetheless produces a promising first prototype of the noise rheostat in mammalian cells. The next step is produce a singularly insert stable cell line for characterization. It would be nice for the next characterization to allow for cell cycle synchrony and thus reduce the role of extrinsic noise from contributing to total noise levels. Ultimately further circuit design changes may be needed, in particular with the promoter of the first node (figure 9) being CMV may produce problems during integration due to high rates of transcriptional silencing of the viral promoter due to methylation after integration (Brooks et al. 2004).

REFERENCES

Ahler, E., Sullivan, W., Cass, A., Braas, D., York., Bensinger, S., Graeber, T., Christofk, H. Doxycycline alters metabolism and proliferation of human cell lines. *PLOS one* **8**(5), e64561. (2013).

Ajikumar, P., Xiao, W., Tyo, K., Wang, Y., Simeon, F., Leonard, E., Mucha, O., Phon, T., Pfeifer, B., Stephanopoulos, G. Isoprenoid pathway optimization for Taxol precursor overproduction in *Escherichia coli*. *Science* **330**(6000), 70 -74 (2010).

Anderson, J., Clarke, E., Arkin, A., Voigt, C. Environmentally controlled invasion of cancer cells by engineered bacteria. *J. Mol. Biol.* **355**(1), 619 – 627 (2006).

Anon. Meeting the malaria treatment challenge: effective introduction of new technologies for sustainable supply of ACTs. *Report of 2008 Artemisinin Enterprise Conference* [online], https://www.york.ac.uk/org/cnap/artemisiaproject/pdfs/AEconference-report-web.pdf (2008).

Anon. Stabilizing the antimalarial drug supply: semisynthetic artemisinin could meet up to one-third of global need. *Path: Drug Development* [online], https://www.path.org/publications/files/DRG_ssart_fs.pdf (2014).

Aranda-Diaz, 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**(3), 545-554 (2017).

Biteen, J. et al. Tools for the microbiome: Nano and beyond. *ACS Nano* **10**(1), 6 – 37 (2015).

Blake, W., Kaern, M., Cantor, C., Collins, J. Noise in eukaryotic gene expression. *Nature* **422**(1), 633-637 (2003).

Blake, W., Balazsi, G., Kohanski, M., Isaacs, F., Murphy, K., Kuang, Y., Cantor, C., Walt, D., Collins, J. Phenotypic consequences of promoter mediated transcriptional noise. *Mol. Cell* **24**(6), 853-865 (2006).

Brooks, A., Harkins, R., Wang, P., Qian, H., Liu, P., Rubanyi, G. Transcriptional silencing is associated with extensive methylation of the CMV promoter following adenoviral gene delivery to muscle. *J. Gene. Med.* **6**(4), 395 – 404 (2004).

Cameron, E., Bashor, C., Collins, J. A brief history of synthetic biology. *Nat. Rev. Microbio.* **12**(1), 381-390 (2014).

Chang, H., Hemberg, M., Barahona, M., Ingber, D., Huang, S. Transcriptome-wide noise controls lineage choice is mammalian progenitor cells. *Nature* **453**(1)**,** 544 – 547 (2008).

Cohen, S., Chang, A., Boyer, H., Hellings, R. Construction of biologically function bacterial plasmid *in vitro*. *Proc. Nat. Acad Sci.* **70**(11), 3240 – 3244 (1973).

Cong, L. et al. Multiplex genome engineering using CRISPR/Cas systems. Science **339**(6121), 819-823 (2013).

Dandekar, D., Kumar, M., Ladha, J., Ganesh., K., Mitra, D. A quantitative method for normalization of transfection efficiency using enhanced green fluorescent protein. *Anal. Biochem*. **342**(1), 341-344 (2005).

Dar, R., Hosmane, N., Arkin, M., Siliciano, R., Weinberger, L. Screening for noise in gene expression identifies drug synergies. Science **344**(6190), 1392-1396 (2014).

Eldar, A., Elowitz, M. Functional roles of noise in genetic circuits. *Nature* **467**(1), 167-173 (2010).

Elowitz, M., Leibler, S. A synthetic oscillatory network of transcriptional regulators. *Nature* **403**(1), 335-338 (2000).

Engler, C., Kandzia, R., Marillonnet, S. A one pot, one step, precision cloning method with high throughput capacity. *PLoS One* **3**(11), e3647 (2008).

Esensten, J., Bluestone, J., Lim, W. Engineering therapeutic T cells: from synthetic biology to clinical trials. *Annu. Rev. of Path. Mechanisms of Disease* **12**(1), 305 – 330 (2017).

Gardener, T, Cantor, C., Collins, J., Construction of a genetic toggle switch in *Escherichia coli*. *Nature* **403**(1)**,** 339-342 (2000).

Gibson, D., Young, L., Chuang, R., Venter, C., Hutchison, C., Smith, H. Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nat. Methods.* **6**(5), 343-345 (2009).

Grote, J., Kyrsciak, D., Streit W. Phenotypic heterogeneity, a phenomenon that may explain why quorum sensing does not always result in truly homogeneous cell behavior. *Appl. Environ. Microbiol.* **81**(16), 5280 – 5289 (2015).

Guet, C., Elowitz, M., Hsing, W., Leibler, S. Combinatorial synthesis of genetic network. *Science* **296**(1), 1466 – 1470 (2002).

Guye, P., Li, Y., Wroblewska, L., Duportet, X., Weiss, R. Rapid, modular, and reliable construction of complex mammalian gene circuits. *Nucleic Acids Res.* **41**(16), e156 (2013).

Jacob, F., Monod, J. One the regulation of gene activity. *Cold Spring Harb. Symp. Quant. Biol.* **26**(1), 193-211 (1961).

Jinek, M., Chylinski, K., Fonfara, I., Hauer, M., Doudna, J., Charpentier, E. A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science* **337**(6096), 816 – 821 (2012).

Kalmar, T. et al. Regulated fluctuations in *nanog* expression mediate cell fate decisions in embryonic stem cells. *PLoS Biol.* **7**(7), e1000149 (2009).

Kelly, D., and Kelly, S. Reviewing yeast for drug synthesis. *Nat. Biotech.* **21**(1), 133 – 134 (2003).

Khalil, A., Collins, J. Synthetic biology: applications come of age. *Nat. Rev. Genet*. **11**(5), 367-379 (2010).

Knight, T. Idempotent vector design for standard assembly of BioBricks. *MIT Synthetic Biology Working Group Technical Reports* [online], http://web.mit.edu/synbio/release/docs/biobricks.pdf (2003).

Knuuttila, T. Loettgers, A. Varieties of noise: analogical reasoning in synthetic biology. *Studies in History and Philosophy of Science Part A.* **48**, 76-88 (2014).

Lee, M., DeLoache, W., Cervantes, B., Dueber, J. A highly-characterized yeast toolkit for modular, multi-part assembly. *ACS Synthetic Biology* **4**(9)**,** 975 – 986 (2015).

Liang, F., Ho, W., Crabtree, G. Engineering the ABA plant stress pathway for regulation of induced proximity. *Sci. Signal.* **4**(164), rs2 (2011).

Loew, R., Heinz, N., Hampf, M., Bujard, H., Gossen, M. Improved Tet-responsive promoter with minimized background expression. *BMC Biotechnol.* **10**(81), [online] (2010).

Lyubimova, A. et al. Single-molecule mRNA detection and counting in mammalian tissue. *Nat. Prot*. **8**(9), 1743-1758 (2013).

McAdams, H., Shapiro, L. Circuit simulation of genetic networks. *Science* **269**(1), 650-656 (1995).

McAdams, H., Arkin, A. Towards a circuit engineering principle. *Curr. Biol.* **10**(1), R318-R320 (2000).

McCullagh, E., Seshan, A., El-Samad, H., Madhani, H. Coordinate control of gene expression noise and interchromosomal interactions in a MAP kinase pathway. *Nature Cell Biology* **12**(1), 954 – 962 (2010).

Moris, N., Pina, C., Martinez Arias, A. Transition states and cell fate decisions in epigenetic landscapes. *Nature Reviews Genetics* **17**(1), 693-703 (2016).

Ozbudak, E., Thattai, M., Kurtser, I., Grossman, A., van Oudenaarden, A. Regulation of noise in the expression of a single gene. *Nature Genetics* **44**(1), 69 – 73 (2002).

Paddon, C., Westfall, P., Newman, J. High-level semi-synthetic production of the potent antimalarial artemisinin. *Nature* **496**(1), 528-532 (2013).

Raj, A., Oudenaarden, A. Nature, nurture, or chance: stochastic gene expression and its consequences. *Cell* **135**(1), 212- 226 (2008).

Raser, J., O'Shea, E. Control of stochasticity in eukaryotic gene expression. *Science* **304**(5678), 1811 – 1814 (2004).

Raser, J., O'Shea, E. Noise in gene expression: origins, consequences, and control. *Science* **309**(5743), 2010 – 2013 (2005).

Richard, M., Yvert, G. How does evolution tune biological noise? *Frontiers in Genetics* **5**(1), 374 doi:10.3389/fgene.2014.00374 (2014).

Ro, D. et al. Production of antimalarial drug precursor artemisinic acid in engineered yeast. *Nature* **440**(1), 940-943 (2006).

Sarrion-Perdigones, A., Falconi, E., Zandalines, S., Juarez, P., Fernandez-del-Carmen, A., Grannel, A., Orzaez, D. GoldenBraid: an iterative cloning system for standardize assembly of reusable genetic modules. *PLoS One* **6**(7), e21622 (2011).

Schmid, G., Hofheinz, W. Total synthesis of qinghaosu. *J. Am. Chem. Soc.* **105**(3), 624-625 (2014).

Sigal et al. Variability and meory of protein levels in human cells. *Nature* **444**(1)**,** 643 – 646 (2006).

Singh, A., Weinberger, L. Stochastic gene expression as a molecular switch for viral latency. *Curr. Opin. Microbiol.* **12**(4)**,** 460 – 466 (2009).

Smanski, M., Zhou, H., Claesen, J., Shen, B., Fischbach, M., Voight, C. Synthetic biology to access and expand nature's chemical diversity. *Nat. Rev. Microbio.* **14**(1), 135 – 149 (2016).

Swain, P., Elowitz, M. Siggia, E. Intrinsic and extrinsic contributions to stochasticity in gene expression. *Proc. Nat. Acad. Sci.* **99**(20), 12795- 12800 (2002).

Trapnell, C. et al. Pseudo-temporal ordering of individual cells reveals dynamics and regulators for cell fate decisions. *Nat. Biotechnol.* **32**(4)**,** 381-386 (2014).

Wang, H et al. Programming cells by multiplex genome engineering and accelerated evolution. *Nature* **460**(1), 894 – 898 (2009).

Weber, E., Engler, C., Gruetzner, R., Werner, S., Marillonnet, S. A modular cloning system for standardized assembly for multigene constructs. *PLoS One* **6**(2), e16765 (2011).

Weiss, R., Basu, S. The device physics of cellular logic gates. *First Workshop on Non-Silicon Computing* [online], http://hpcaconf.org/hpca8/nsc.pdf (2002).

Wu, M., Singh, A. Single-cell protein analysis. *Curr. Opin. Biotechnol.* 23(1), 83 – 88 (2012).

Yim, H. et al. Metabolic engineering of *Escherichia coli* for direct production of 1,4-butanediol. *Nat. Chem. Biol.* **7**(7), 445 – 452 (2011).

ACKNOWLEDGEMENTS

I want to thank my friends and family for supporting me through the writing process, reading through my notes, and bouncing ideas with me. I want to thank Dr. Joao Pedro Fonseca and Alain Bonny their mentorship and constant support throughout the past two years. My development as a young scientist would not have been possible if they had not taken me under their wing. I want to also thank Lindsey Osmiri for the encouragement throughout my time in lab as well as friendship and camaraderie over failed cloning experiments. I want to thank Professor Hana El-Samad for allowing me to work in the lab for two summers and cultivating an open space for learning and growth. I want to thank Professor Phoebe Lostroh for the years of mentorship and support as both my major advisor and thesis advisor. I want to thank Professor Jennifer Garcia for her mentorship and advice throughout the thesis process, fellowship writing, and graduate school applications. Lastly, I want to thank the various funding sources that have made this research possible: The National Science Foundation Research Experiences for Undergraduates, Colorado College Unpaid Internship Funding through the Career Center, and lastly to Jean-Paul Bonny.

APPENDICES

Appendix 1: Recipes and Protocols

Luria Broth (LB)

10g	Tryptone
10g	NaCl
5g	Yeast Extract
Up to 1L	dH_2O

Adjust volume up to 1 L using water. Adjust pH to 7.0 with NaOH and autoclave prior to usage.

Transformation Buffer (TB)

1.18g	HEPES
9.3g	KCl
1.1g	$CaCl_2$
5.4g	$MnCl_2 \cdot 4H_2O$
Up to 500 mL	dH_2O

Add HEPES, KCl, and $CaCl_2$ into 500 mL of water and mix to dissolve all solid. Modulate pH to 6.7 with KOH. Add $MnCl_2 \cdot 4H_2O$. Filter solution through 0.2µm filter and store solution at 4^{o} C.

SOB Media

20g	Tryptone
5g	Yeast Extract
0.5g	NaCl
10 mL	250 mM KCl
Up to 1L	dH_2O

Adjust volume up to 1 L using water. Adjust pH to 7.0 with NaOH and autoclave the solution. Prior to usage, add 5mL of sterile 2M $MgCl_2$.

DH5α Chemically Competent Cells

Inoculate *E. coli* DH5α into 5 mL of LB. Grow overnight at 30° C. Inoculate SOB media with overnight culture at a 1:1000 dilution and grow at 20° C until $OD_{600} = 0.6$. Chill cells rapidly in an ice bath for 10 minutes, and swirl occasionally. Pellet the cells at 3,800 rpm for 20 minutes. Resuspend pellets completely with 10 mL ice cold TB and transfer cells into 50 mL falcon tubes. Chill tubes for 10 minutes in ice bath. Spin down cells at 3800 rpm for 15 minutes and decant supernatant. Resuspend the pellet in ice cold TB at a final volume of 41.5 mL. Add 3 mL of DMSO per falcon tube and gently mix the tube. Leave the contents on ice for 10 minutes prior to aliquoting cells into pre-chilled 1.5 mL centrifuge tubes. Freeze tubes immediately using liquid nitrogen.

Appendix 2: Plasmid Bank

All plasmids built and used for this project are found on Benchling.com in shared Summer 2017 Noise Rheostat Project Folder.

Appendix 3: Data Analysis

Flow Cytometry Fluorescence Gating Parameters

Filter	**What the reading represents**	**Gated Lower Range**	**Gated Higher Range**

FCS-A	**Forward Scatter ***	**$0.5x10^5$**	**$2.7x10^5$**
SSC-A	**Side Scatter ****	**N/A**	**$2.6x10^5$**
FITC-A	**Green Fluorescence**	**0**	**N/A**
PETxR-A	**Red Fluorescence**	**0**	**N/A**
DAPI-A	**Blue Fluorescence**	**$1.5x10^3$**	**N/A**

* Forward scatter is generally a measurement for the size of the cell. Particularly, it is the light that is scattered forward into a detector.

** Side scatter is a measurement for granularity of a cell. Generally, this is the light that is scattered by granules and other particulates within the cell towards side detectors.

Statistical Analysis and Equations

Log_{10} Data Formatting

All fluorescence data will be presented in Log_{10} to better visualize variance in the context of a bell curve distribution.

Mean Calculation

$$\bar{x} = \frac{\sum_{i=1}^{n} x_i \, n}{n}$$

Standard Deviation

$$\sigma_n = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2}$$

Geometric Coefficient of Variation

$$Cv = \sqrt{e^{\sigma_n \ln 10} - 1}$$

Appendix 4: Engineering vocabulary

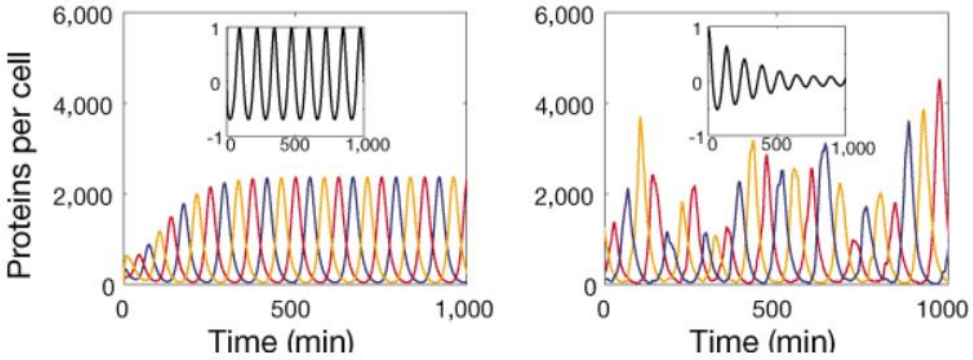
Analogue Circuitry: Circuits with continuous variable signals. These types of circuits are sensitive to noise as it takes continuous signals. In contrast, digital circuits take quantized signals, which is not affected by small fluctuations within the circuit.

Logic Gate: A circuit that can perform simple logical commands. These simple commands include: and, or, nand (not and), nor (not or), and not commands can process two signals to determine output in binary. These simple gates can be linked into complex logic circuits that can perform bio-computing problems.

Metabolic Engineering: The science of synthetic construction of artificial metabolic pathways through rational design principles to improve product processing *in vivo*.

Moore's Law: First described in 1965, the phenomenon that describes the doubling of the density of transistors that fit into the same space every year since the early 1970s. This is a surrogate measurement of the increasing computing power of computers and other computer like devices.

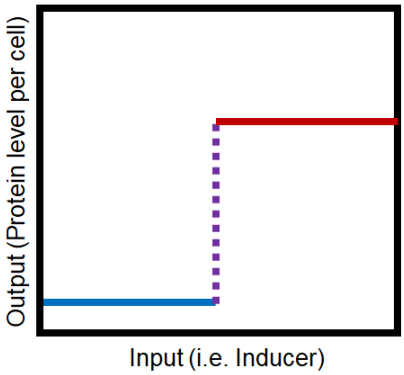
Oscillator: a device generating oscillating electric currents or voltage by nonmechanical means.



A biological adaption of an oscillator involves oscillating of protein levels per cell in a predictable pattern.

Adapted from Elowitz and Leibler 2000.

Toggle Switch: a physical switch that could open and close a circuit by interrupting current flow. This is literally an on and off switch.



Vector Decomposition: the decomposition of vector in R^n into n number of linearly independent vectors

Appendix 5: Additional Figures

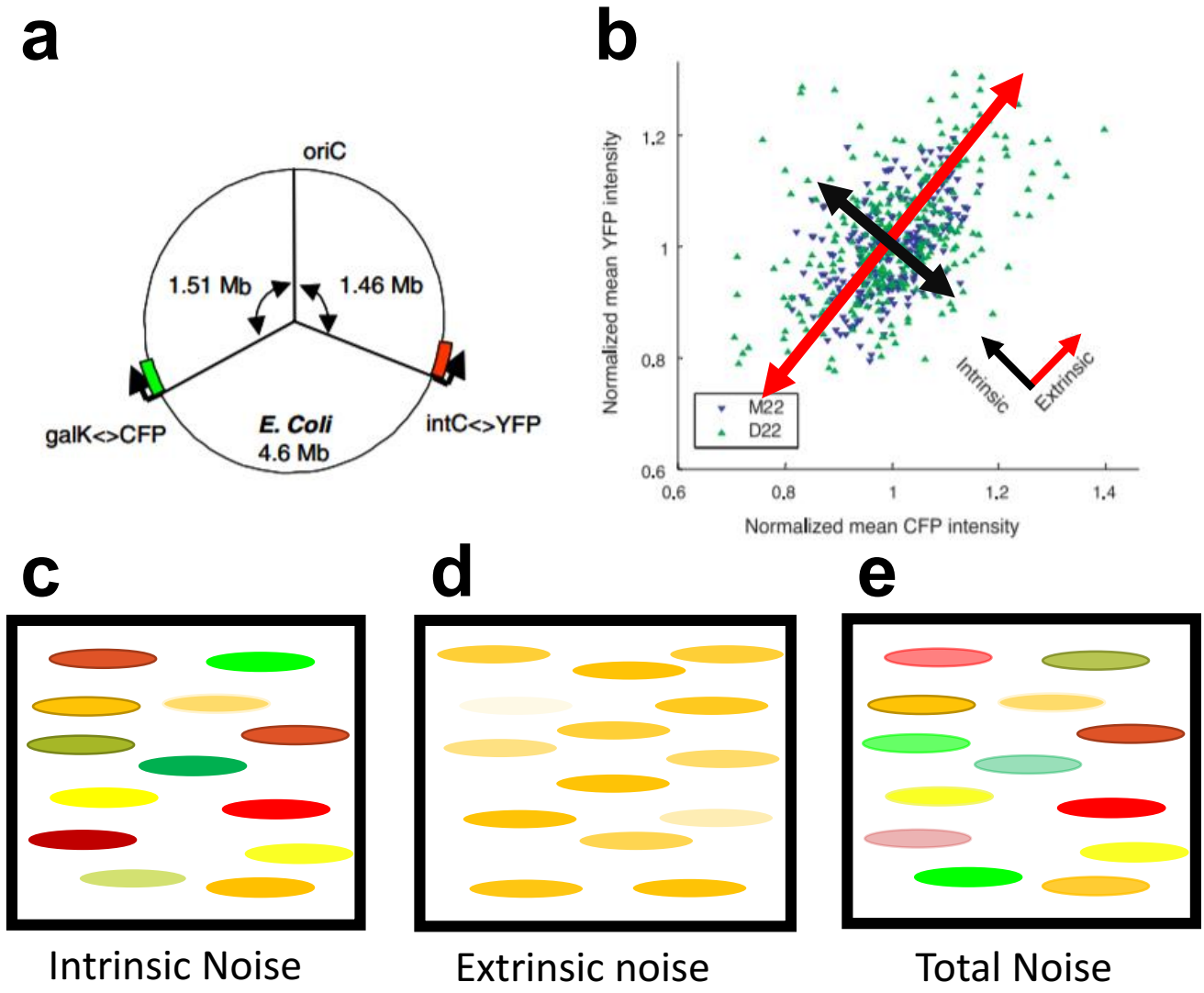


Figure 1. Intrinsic and extrinsic noise de-convolution
This image was adapted from Elowitz et al. (2003). (a) Two reporter constructs were stably integrated into *E. coli* genome. The CFP and YFP proteins were integrated into galK and intC loci respectively. Each protein is under the transcriptional control of identical promoters. (b) Flow cytometry measurements were taken for M22 and D22 strains, each with a different promoter driving the reporters. The intrinsic and extrinsic noise vectors are superimposed onto the dataset with black and red arrows, respectively. The different promoter reporters appear to have different levels of noise. D22 appears to have more intrinsic noise compared to M22, suggesting that the promoter used is inherently noisier. (c) The intrinsic noise of the reporters is shown as a cartoon. If intrinsic noise varies perpendicular to the best fit line, then the amount of protein total is consistent but the ratio of each individual reporter is different. Thus,

Figure 1. Intrinsic and extrinsic noise de-convolution (*continued*)
the cells observed would exhibit different colors as differing amounts of green and red fluorescence are mixed. (d) The extrinsic noise of the reporters is shown as a cartoon. If the extrinsic noise varies along the best fit line, this suggest that the ratio of each reporter should be the same, but the amount per cell differs. Thus, the color of each cell in the population is identical, but the intensity differs. (e) The total noise is a combination of the intrinsic and extrinsic such that the ratio of the reporters as well as the amount of proteins between cells differ.

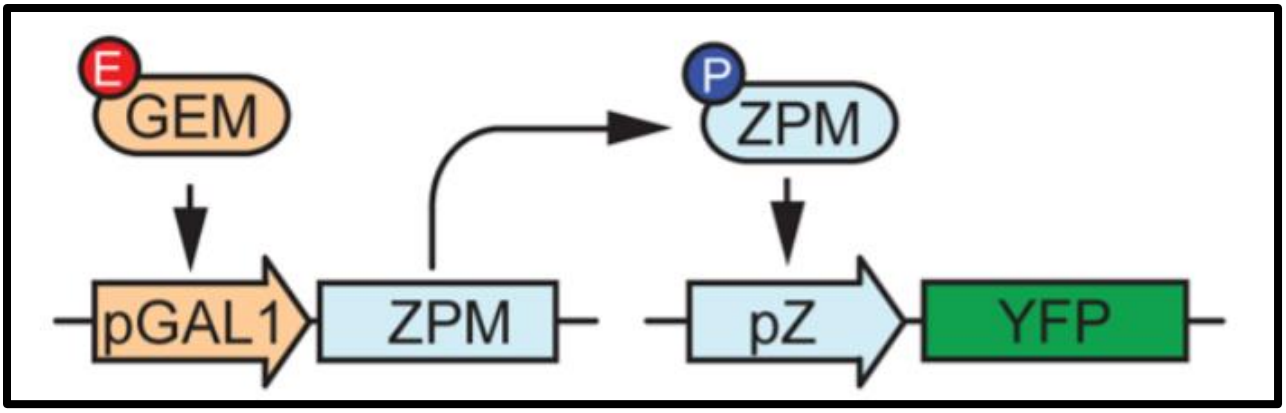


Figure 2. Two-node in-series construct produces noise control in *S. cerevisiae*
This image was adapted from Andras-Diaz et al. (2017). The two-node in-series construct consists of two inducible transcription systems GEM and ZPM. GEM is constitutively expressed but is only active with the addition of estradiol. GEM promotes the expression of ZPM under the promoter pGAL1. ZPM is the second transcription factor. It is induced by progesterone, and promotes the expression of YFP, a fluorescent protein, under the control of pZ. This in-series architecture is what is used to build the noise rheostat in Andras-Diaz et al.

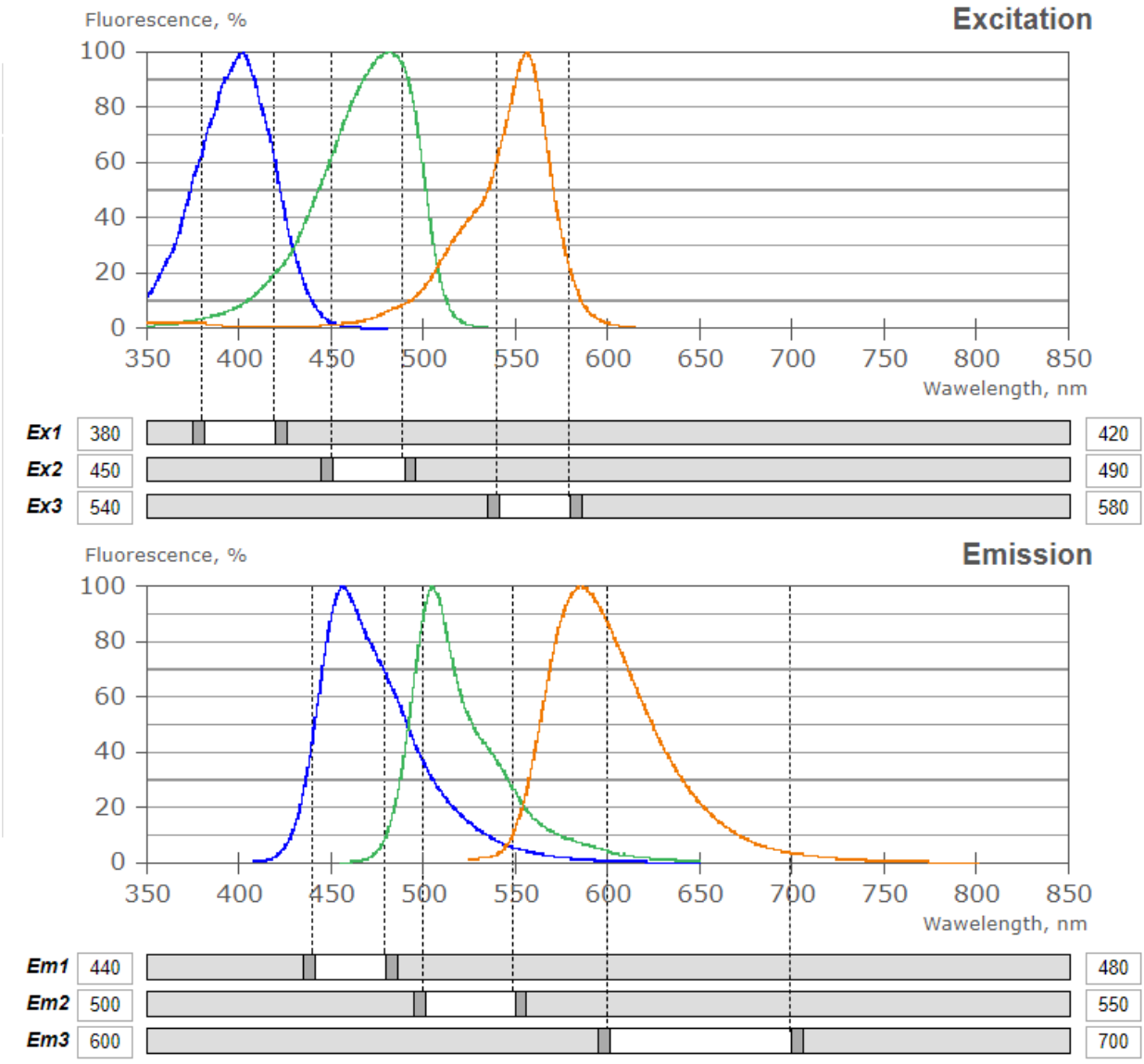


Figure 3. Excitation and emission spectra of blue fluorescent protein, green florescent protein, and red fluorescent protein (tagBFP, GFP, RFP)
This image was adapted from MicroscopyU. The excitation spectra and emission spectra of these fluorescent proteins are the most separated and distinct peaks of all other fluorescent protein combinations. The blue peaks represents tagBFP, with peak excitation at 400 nm and peak emission at 460 nm. The green peak represents EGFP, with peak excitation at 480 nm and peak emission at 510 nm. The red peak represents RFP, with peak excitation at 560 nm and peak emission at 590 nm.

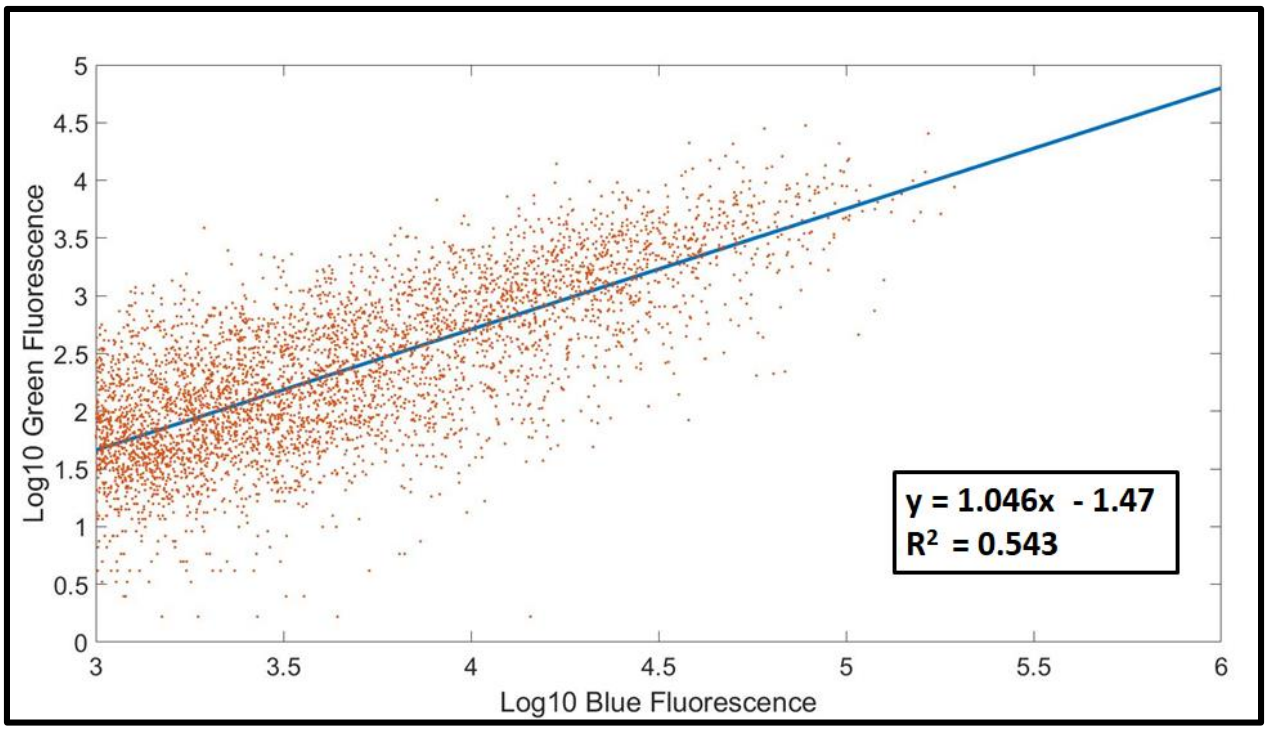


Figure 4. TetOn3G blue fluorescence plotted against green fluorescence shows plasmid copy number variability

The blue fluorescence plotted against the green fluorescence shows that there is a positive relationship between blue and green fluorescence. A linear regression was fit onto the data, with a $p<0.00001$. The slope of the line is 1.046, which is slightly higher than the 0.647 of the abscisic acid one node construct suggesting that TetOn3G has a comparably higher basal fluorescence.

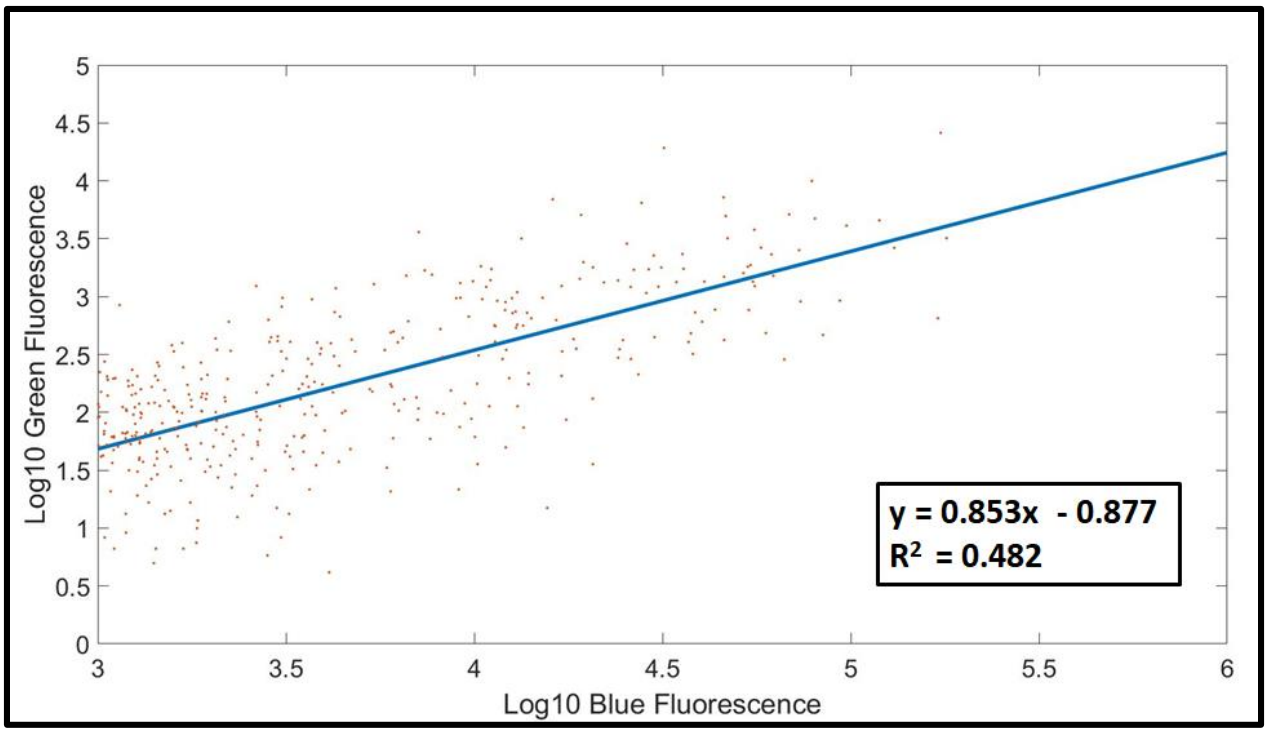


Figure 5. Two-node rheostat blue fluorescence plotted against green fluorescence shows copy number variability
The blue fluorescence plotted against the green fluorescence of the two node rheostat shows a positive correlation. A linear regression was done with $p < 0.000001$, which suggest that the linear fit is significant. The slope of 0.853 is an intermediate value between the TetOn3G system and ABA-CIP system. The basal fluorescence is in between the one-node constructs.

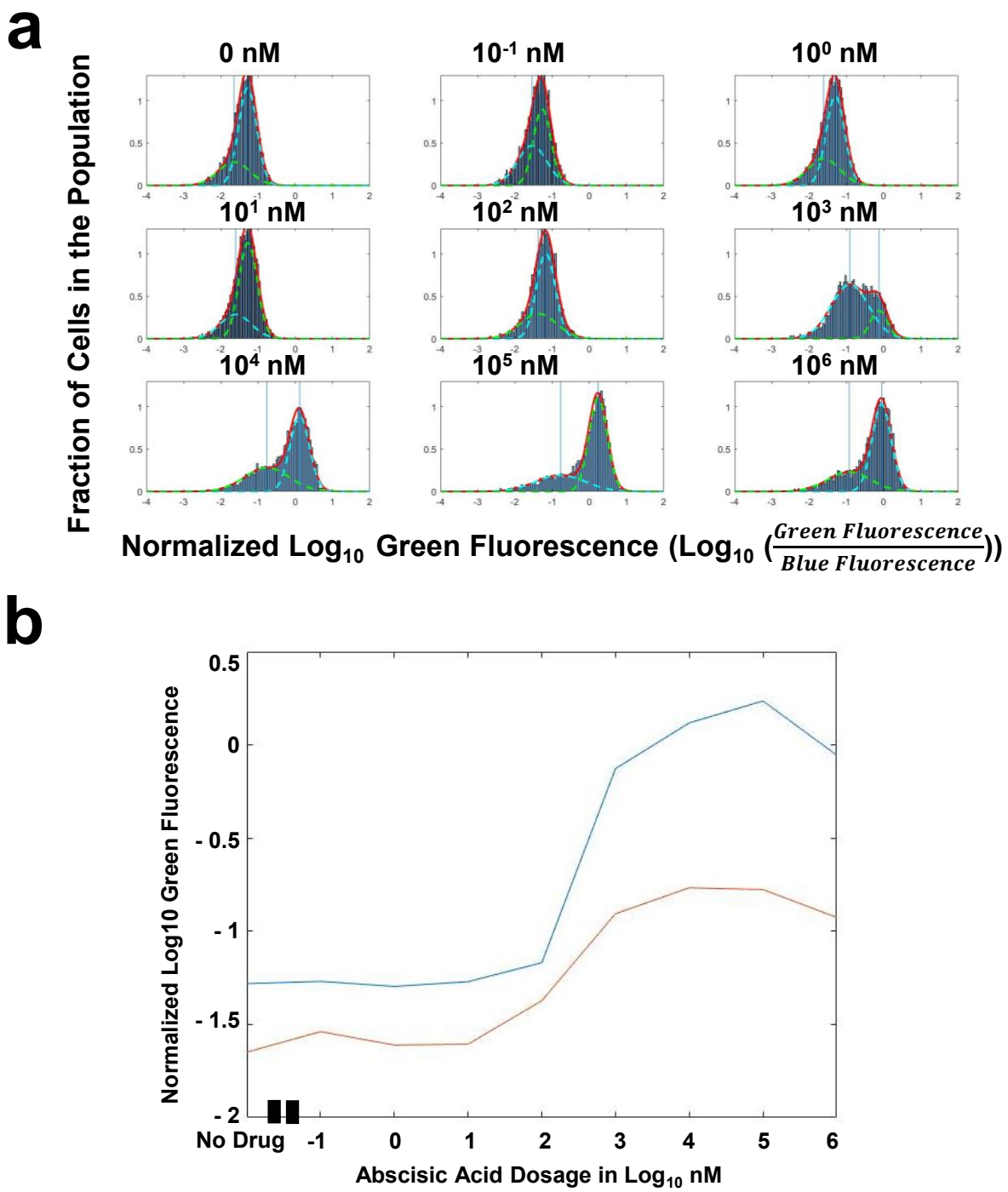


Figure 6. Gaussian mixture model analysis on ABA-CIP one-node circuit induction data set

Gaussian mixture model analysis was done to separate gaussian peaks from the dataset to determine “on” and “off” populations. (a) The blue vertical lines represent the mean of the individual peaks. The “on” population is the right peak while the “off” population is the left peak. As one increases the dosage, the right peak begins to

Figure 6. Gaussian mixture model analysis on ABA-CIP one-node circuit induction data set (*continued*)
move away from the left peak in a dose dependent manner. The mean of the “on” population peak is similar to the mean of the total population. (b) The mean of the “on” and “off” population is plotted in blue and red respectively. The “on” population mean curve looks almost identical to the total population mean curve.

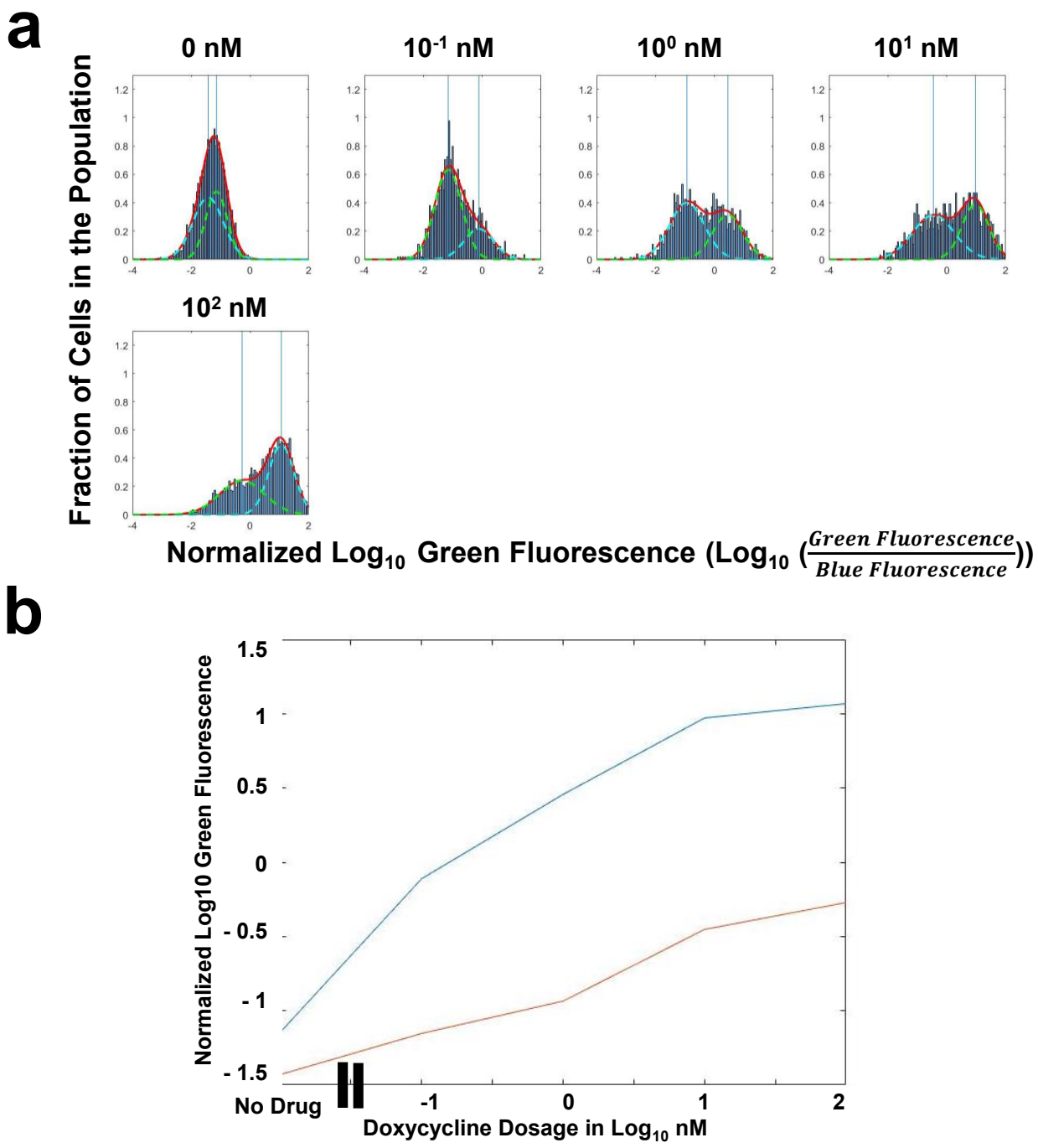


Figure 7. Gaussian mixture model analysis on TetOn3G one-node induction dataset.

Gaussian mixture model analysis was performed on the dosage experiment for the TetOn3G one-node construct to separate the on and off population. (a) The blue vertical lines mark the mean of the individual peaks. The curve on the right is the “on” population while the curve on the left is the “off” population. As the dosage of

Figure 7. Gaussian mixture model analysis on TetOn3G one-node induction dataset (*continued)*
doxycycline increased, the two curves move away from one another. The mean of the "on" population is comparable to the mean of the total population. (b) The mean of the "on" and "off" population are plotted in blue and red respectively. The dosage curve of the "on" population is reminiscent in shape and magnitude to the dosage curve of the total population.

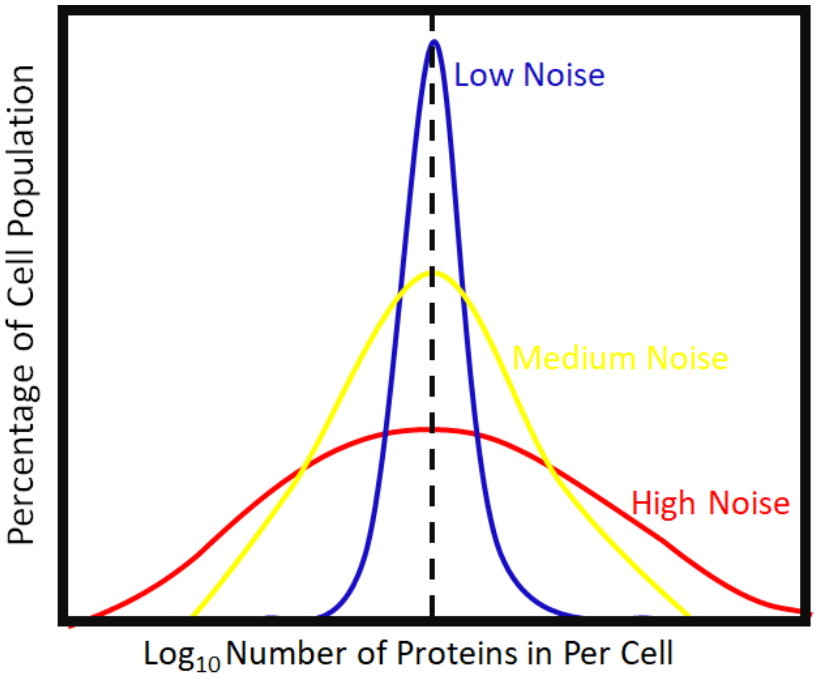


Figure 1. Theoretical model of gene expression noise control
Noise of gene expression is described by the width of a normal distribution. Low gene expression noise (in blue) describes a narrow distribution such that variance of phenotypic presentation is low. High gene expression noise (in red) describes a wide distribution such that the variance of phenotypic presentation is high. Current methods of changing the noise of gene expression also happens to change the mean of gene expression. Biological readouts obtain from such methods are difficult to interpret as deconvolution of the effects of noise and the effects of mean gene expression is nearly impossible. We must be able to control for the mean of gene expression while changing the variance to be able to understand the biological role of noise in various systems.

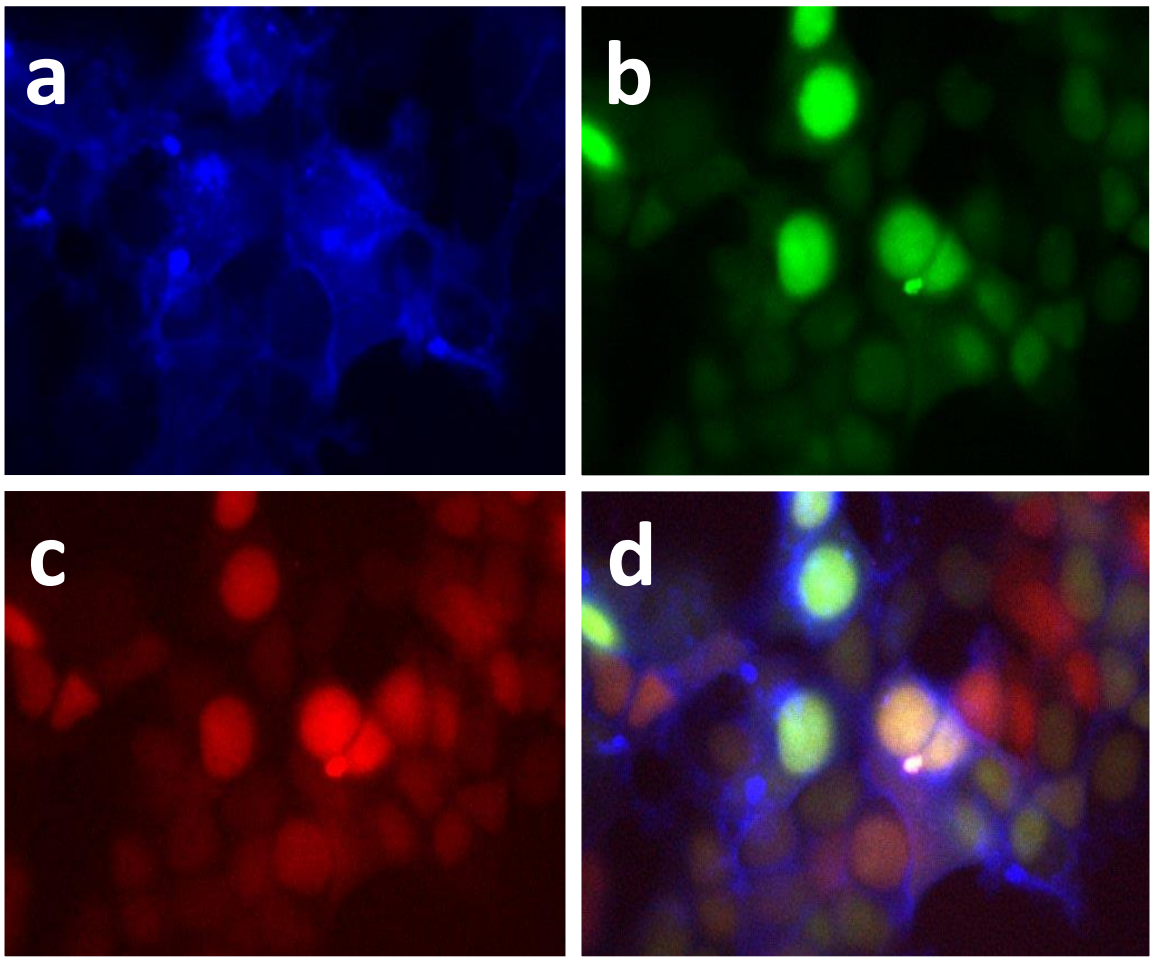


Figure 2. Verification of correct expression of three fluorescent proteins with localization signals
HEK293T cells had stably integrated red, blue, and green fluorescence proteins tagged with localization signals attached to the protein. The cells were subsequently imaged using a Zeiss Laser Scanning Microscope (LSM) and ensure that the fluorescent proteins properly folded and fluoresced. (a) The blue fluorescence protein (tagBFP) was tagged with a plasma membrane tag (Lyn) that allowed them to localize on the intracellular side of the plasma membrane. (b) The green fluorescent protein (EGFP) was tagged with a nuclear localization signal (NLS) that allowed the protein to localize in the nucleus. (c) The red fluorescent protein (RFP) was tagged with a NLS and allowed for localization at the nucleus. (d) Images a, b, and c were overlayed on top of each other and showed co-localization of the green fluorescent protein and red fluorescent protein.

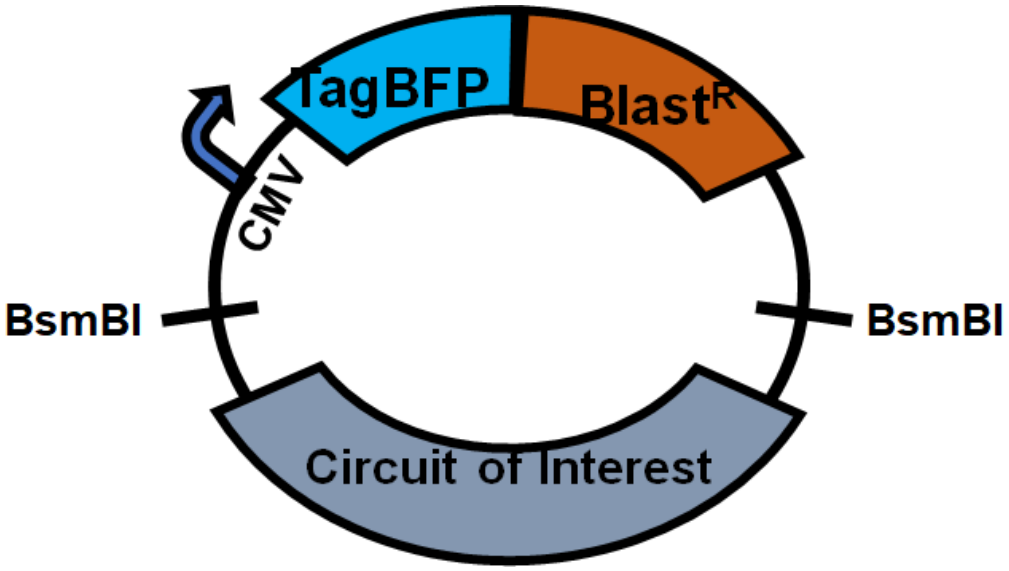


Figure 3. Transient Expression Vector Diagram
Transient expression experiments were done using a plasmid with a fluorescent reporter being constitutively expressed. The tagged blue fluorescent protein (tagBFP) and blasticidin restriance gene (Blast^{R}) is under the control of a cytomegalovirus (CMV) constitutive promoter. Our circuit of interest is inserted into this backbone through golden gate assembly. Using the tagBFP, we can differentiate cells with and without the transient expression vector based on the presence or absence of blue fluorescence through flow cytometry. The blue fluorescence may also be used to control for plasmid copy number differences and thus provide a way for normalization of subsequent data.

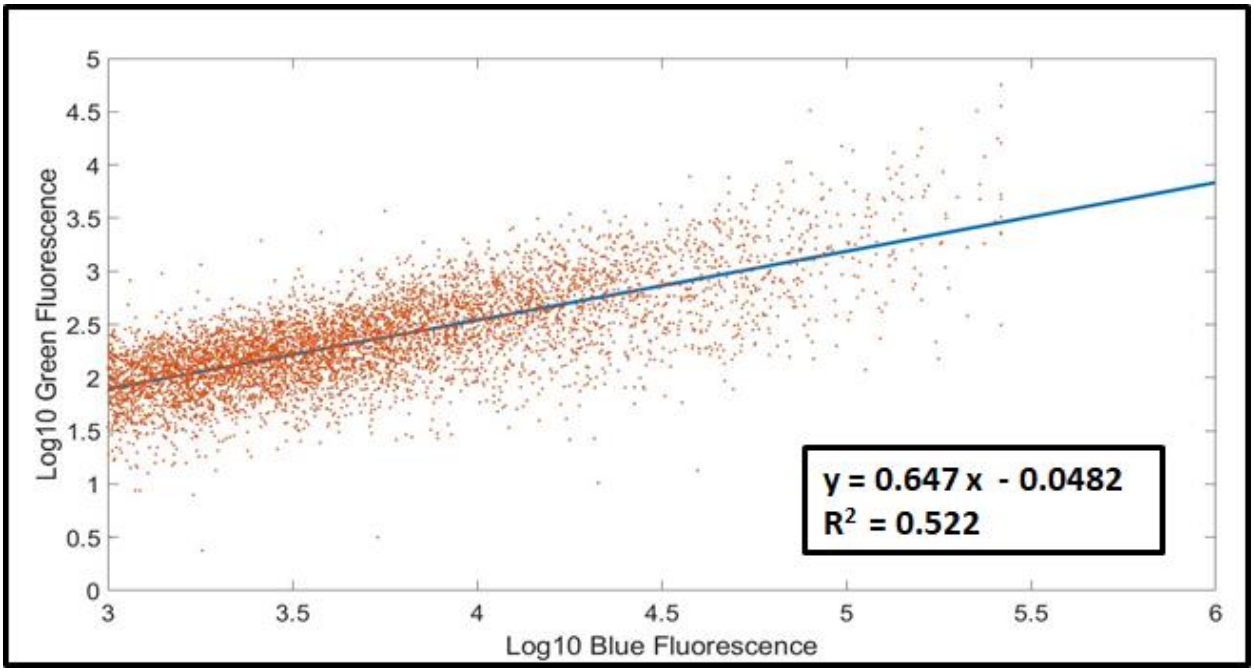


Figure 4. Plasmid copy number dependent gene expression requires normalization of green fluorescence output using blue fluorescence output. Although there was only one copy of green fluorescent protein (*gfp*) and one copy of blue fluorescent protein (*bfp*) in each plasmid, the transient transfection does not allow for the control of plasmid copy number. Using a CMV-BFP on the circuit plasmid backbone, a consistent level of blue fluorescence protein is produced in cells that received the plasmid. This provides a good way to distinguish cells with and without the plasmid using flow cytometry and gating for blue fluorescence intensity. In this plot, the blue fluorescence intensity is plotted against the green fluorescence intensity and exhibits a clear positive correlation between the two intensities.

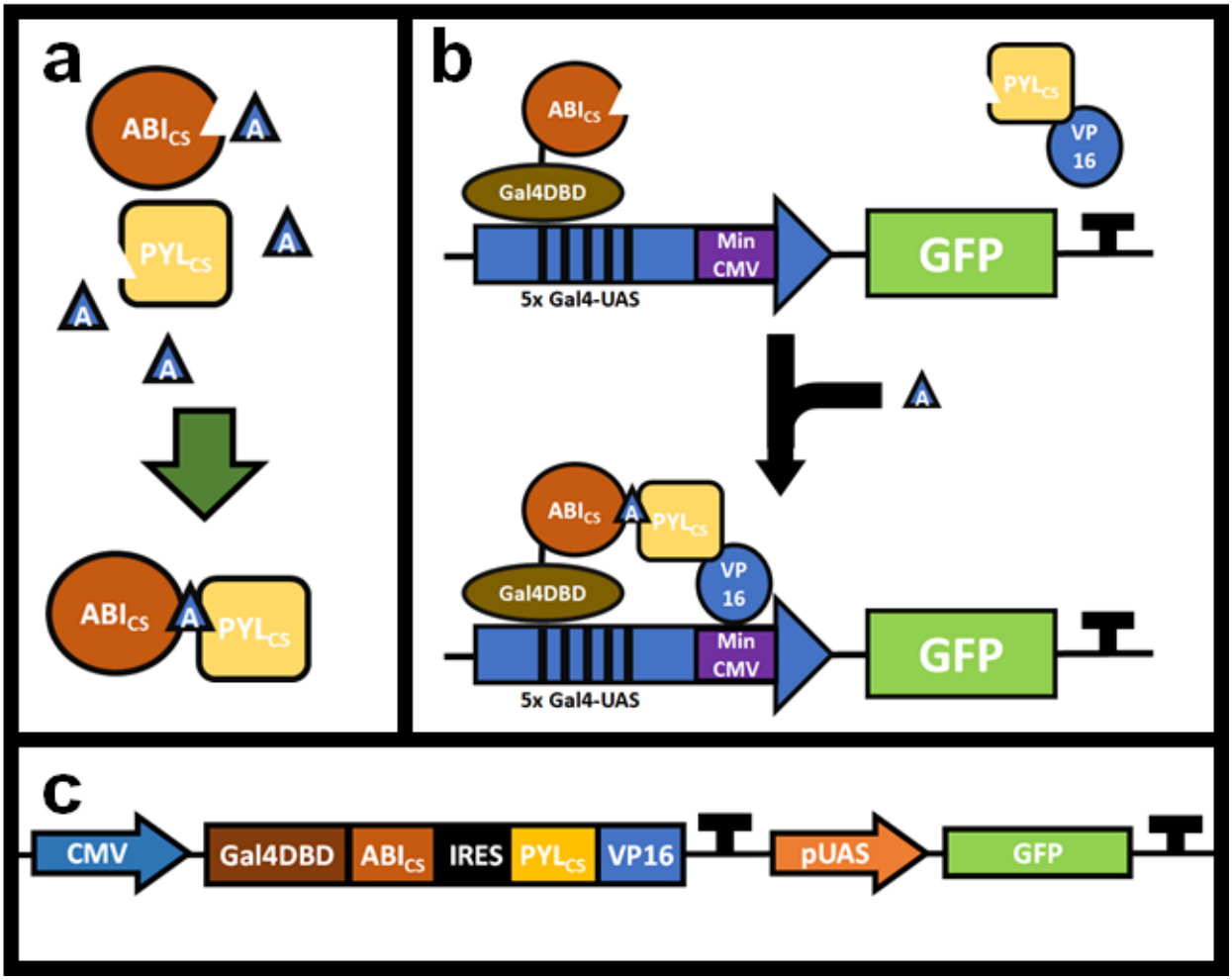


Figure 5. Abscisic acid chemically induced proximity transcription system schematic adapted from Liang et al. 2011

Abscisic acid chemically induced proximity (ABA-CIP) transcription system is composed of a synthetic transcription factor and promoter pair that controls transcription activation using a phytohormone S (+) - abscisic acid (ABA). The transcription factor system has components involved in the plant stress pathway and thus is responsive to the phytohormone (a) Abscisic acid insensitive protein (ABI) and pyrobactin resistance regulatory component (PYL) are free floating and monomers. With the addition of abscisic acid, the triangle with the A, the two proteins form a heterodimer. (b) Liang et al. took advantage of this dimerization upon addition of an inducer, and built around ABI and PYL. Fusion proteins are constructed such that a Gal 4 DNA binding domain (GAL4DBD) is fused to the N-terminus of ABI and VP16, a

Figure 5. Abscisic acid chemically induced proximity transcription system schematic adapted from Liang et al. 2011 (*continued)*
herpes simplex virus transactivation domain, is fused to the C-terminus of PYL. The GAL4DBD is able to bind to an upstream activation sequence (UAS) on a synthetic promoter pUAS, that consist of 5 UAS motifs upstream of a minimal cytomegalovirus (CMV) promoter. This binding localizes, the ABI protein that is fused to the GAL4DBD close to the pUAS. Upon addition of ABA, the GAL4DBD-ABI and PYL-VP16 dimerize. The VP16 is now localized near pUAS and recruit transcriptional machinery to activate transcription of the GFP protein downstream of pUAS. (c) To characterize the transcription system in HEK293T cells, we constructed a genetic construct that will be transfected into the cells. The construct consist of a CMV promoter, which constitutively produce the ABA-CIP synthetic transcription factor components. There is a terminator downstream of the coding sequence as well as two HS4 insulators to make a discrete transcriptional unit. A pUAS promoter driving a green fluorescent protein (GFP) is used as reporter for the activity of the ABA-CIP transcription factors. This construct is subcloned into a backbone with a constitutive blue fluorescence protein reporter as shown in figure 3.

a

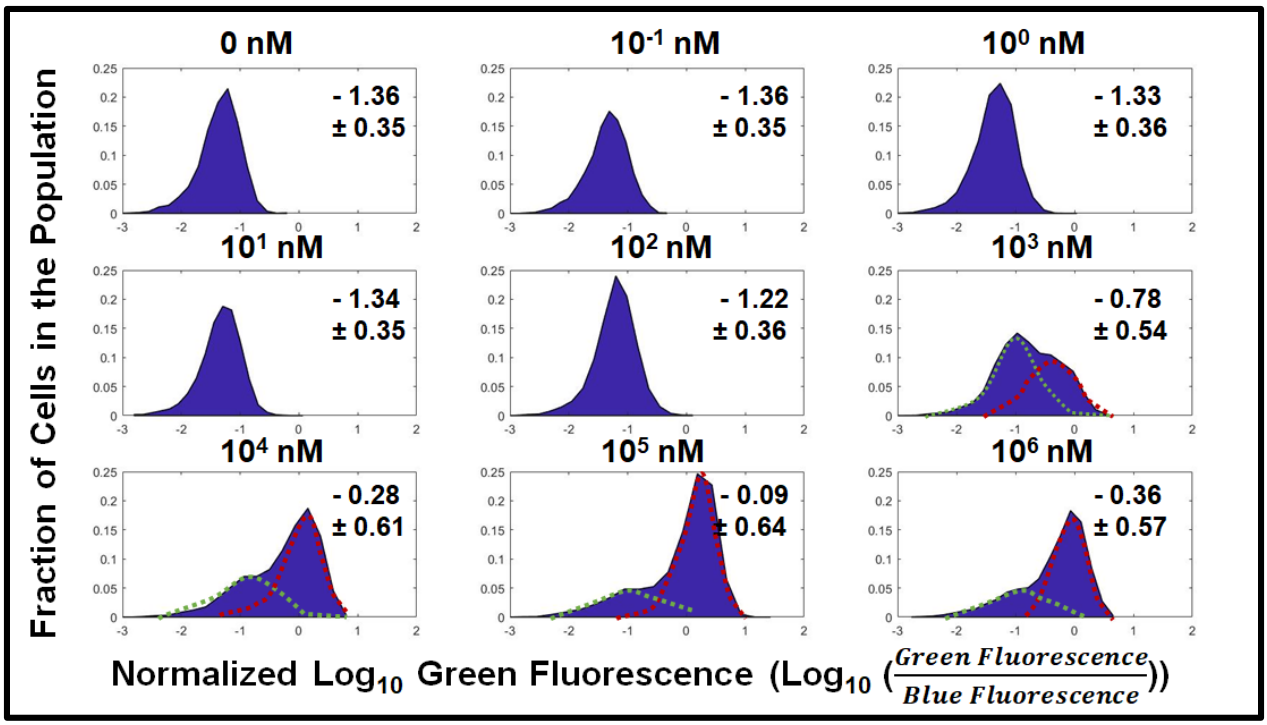


Figure 6. Abscisic acid chemically induced proximity transcription system gene expression characterization

Nine wells with 5 x 10^4 cells per well were induced with their respective dosages of abscisic acid. After 48 hours of induction, flow cytometry readings were obtained for each dose. 10,000 events were collected and subsequently gated for the presence of tagBFP (≥1 x 10^3 fluorescence units through DAPIA filter). The mean ± standard deviation (mean ± S.D.) of cells per well was 5190±140 cells. The green fluorescence reading was subsequently normalized by dividing the green fluorescence by the blue fluorescence (G/B) and $\log_{10}$ transformed. (a) the log-G/B measurements per well were plotted as histograms. These histograms were normalized using a probability normalization such that the integration under the curve for each curve is equal to 1. This makes the histogram distributions more easily comparable if the population numbers differ. The mean ± S.D. log-G/B of each dosage is noted on each histogram. As the dosage of ABA increases, the mean G/B is also increasing. The histogram on the bottom right corner, representing the 10^6 nM ABA dosage shows a decrease in the mean which may be a real phenomenon as it is also observed by Liang et al. in a luciferase assay in Chinese hamster ovary cells. Not only does the mean G/B change, but so does the standard deviation. This change in standard deviation is

b

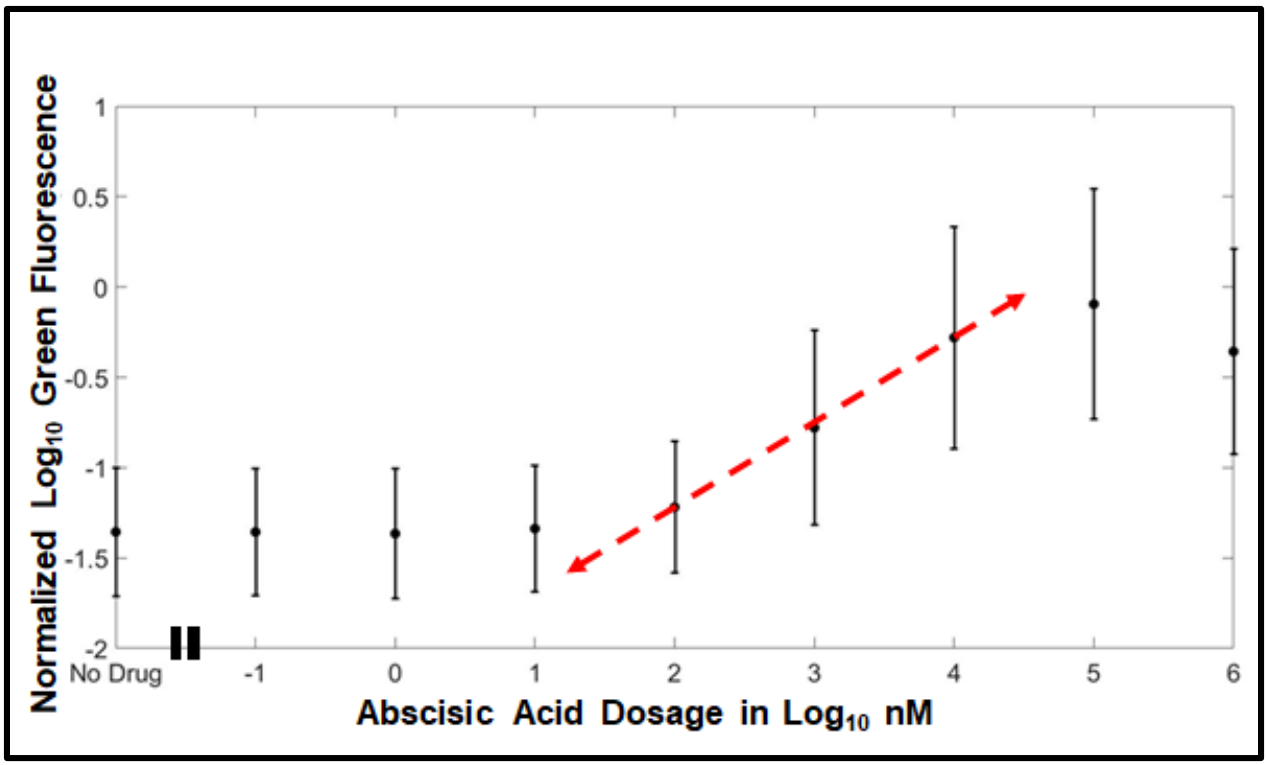


Figure 6. Abscisic acid chemically induced proximity transcription system gene expression characterization *(continued)*
capturing the dynamics of a bi-modal distribution construct from the separation of the on and off cell populations. This phenomenon was first observed at 10^2 nM ABA dosage where we begin to observe first dramatic change in mean fluorescence from the previous dosage. The bi-modality is maintain for the rest of the dosages, but the off population, represented by the cartoon green curve becomes smaller as a larger portion of the cell population turns on. Although this bi-modality exist, the histograms still exhibit a right shifting as the dosages increase. (b) The means of log-G/B for each dosage was plotted. The error bars for each mean is the standard deviation. This plot shows that dosages below 2 log nM ABA are too low to activate ABA-CIP. The construct appears to exhibit log-linear induction between 2 log and 4 log nM ABA concentrations and account for a 8.8 - fold ($10^{0.93}$) increase in G/B. The log-linear dosages allow us to determine preliminary concentrations to be used in the two node noise rheostat dosage experiment to obtain optimal resolution.

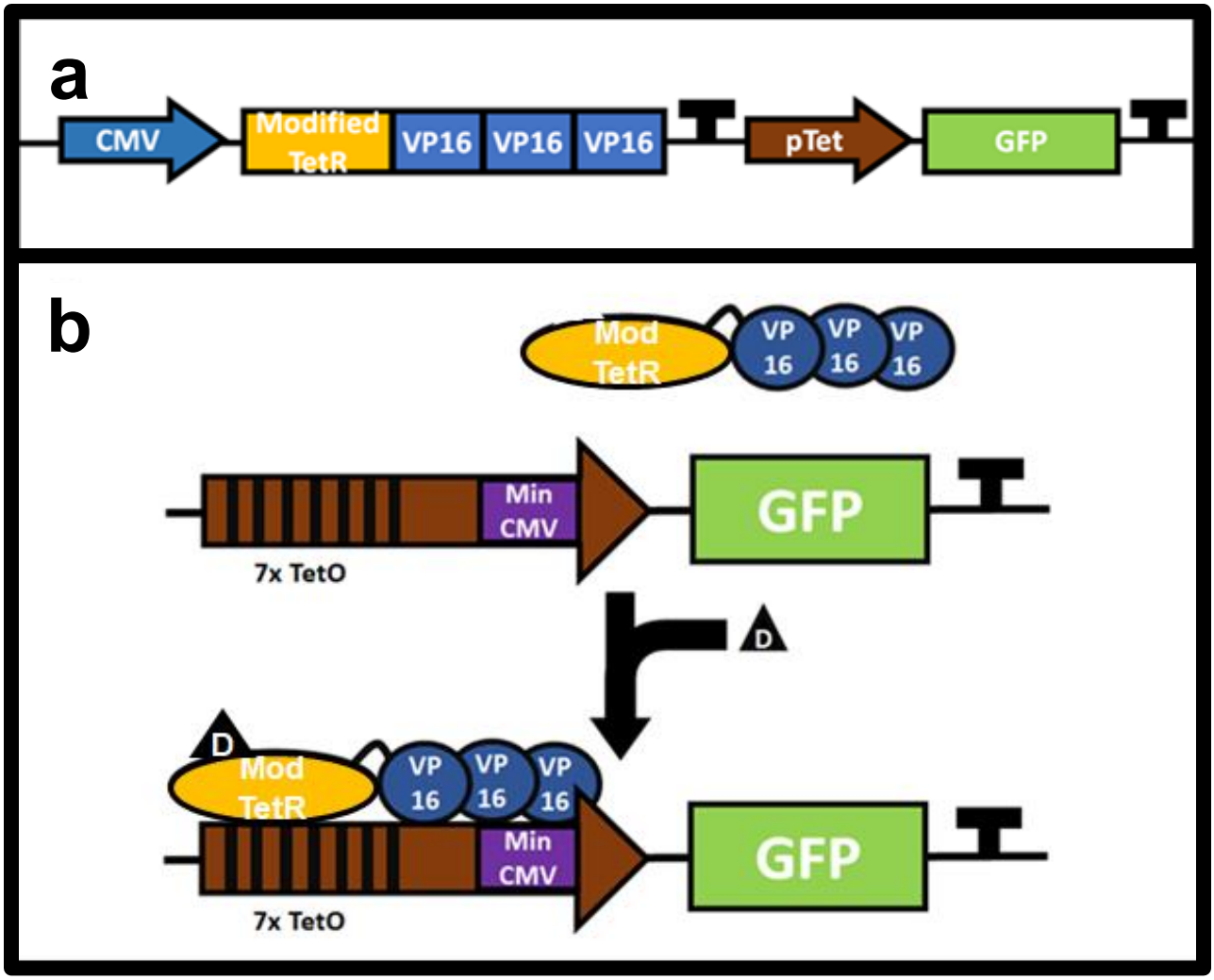


Figure 7. TetOn3G doxycycline inducible transcription system schematic
TetOn3G (Clontech) is another synthetic transcription system which includes an inducible synthetic transcription factor and synthetic cognate promoter. (a) TetR, a tetracycline responsive protein is of bacterial origins. It's been modified bind to Tet operator sequences (TetO) after a doxycycline molecule induces conformational changes. The TetOn3G synthetic transcription factor is composed of a modified TetR (Mod TetR) with three C-terminus fused VP16 transactivation domains. A synthetic promoter (pTet) made up of 7 TetO sequences upstream of a minimal CMV promoter drives the expression of a green fluorescence protein (GFP). Upon addition of doxycycline, labeled as the black triangle with a D, the TetOn3G transcription factor binds to pTet and recruits transcriptional machinery to activate transcription of the green fluorescent protein.

Figure 7. TetOn3G doxycycline inducible transcription system schematic (*continued)*

(b) To characterize the TetOn3G system in HEK293T cells, this construct was built for transfection into the mammalian cells. The constitutive CMV promoter drives expression of the TetOn3G transcription factor . There is a terminator and two HS4 insulators downstream of the coding sequence. pTet-GFP reports of the activity of the TetOn3G transcription factor and the effect of doxycycline on the transcriptional activity of the transcription factor.

a

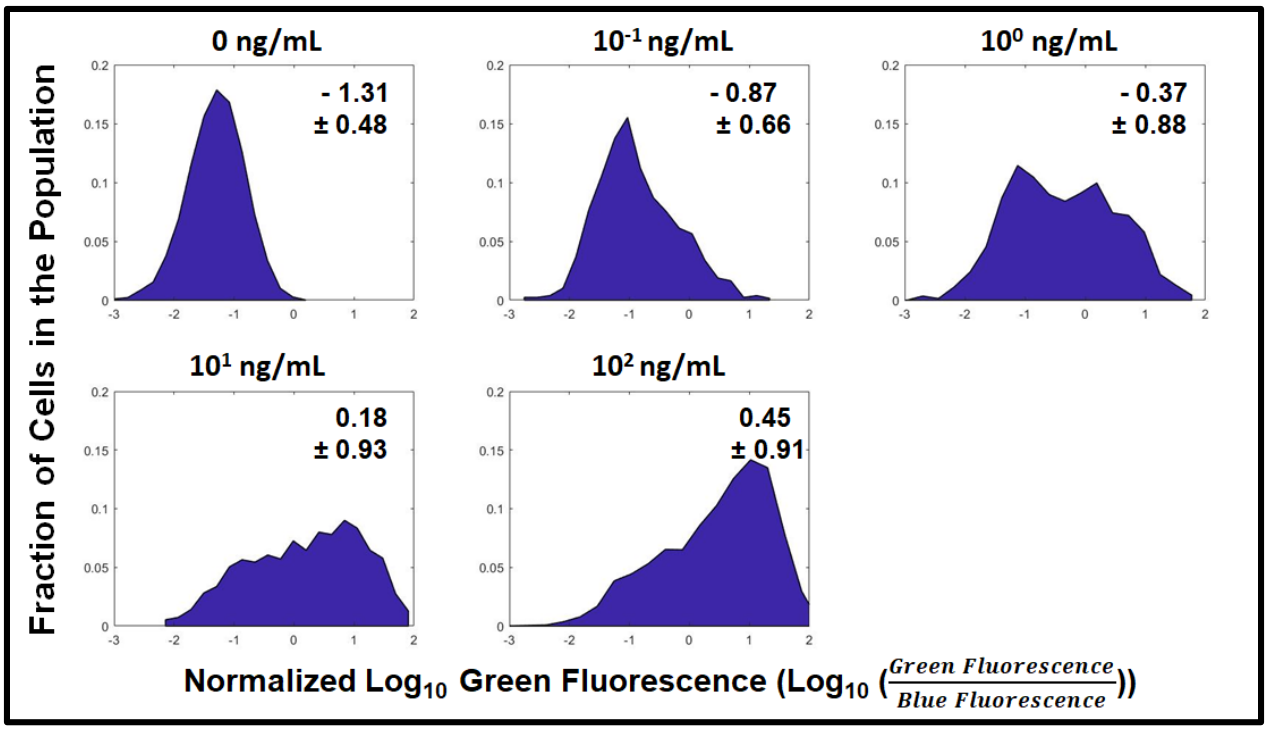


Figure 8. TetOn3G transcription system characterization
Seven wells with 5 x 10^4 cells were induced with their respective dosages of doxycycline. After 48 hours of induction, flow cytometry readings were obtained for each dose. 10,000 events were collected and subsequently gated for the presence of tagBFP. Two wells were eliminated because of gating shift by the flow cytometer at the time of the readings, leaving five populations with valid data. After gating by blue fluorescence, there were 3002 ± 2059 (mean ± S.D.) cells per dosage. The green fluorescence was normalized by dividing by the blue fluorescence (G/B) control marker and $\log_{10}$ transformed. (a) The log-G/B measurements were plotted as histograms. To more easily compare populations of differing sizes, the histograms were normalized by probability normalization such that the integrations under the curves are equal to 1. The mean ± S.D of log-G/B was labeled on each histogram. As the dosage of doxycycline increased, the mean G/B also increased. The induction had occurred at a very low concentration of doxycycline, such that the change in the mean G/B of the population from uninduced to 10^{-1} ng/mL doxycycline had produced a 2.8($10^{0.43}$) – fold change in the mean G/B. The curves do appear to exhibit right shifting with for the most part unimodal distributions with some skew occurring. Some histograms can be interpreted to have bi-modal distributions, but this is less clearly exhibited compared to the ABA-CIP's bimodal.

b

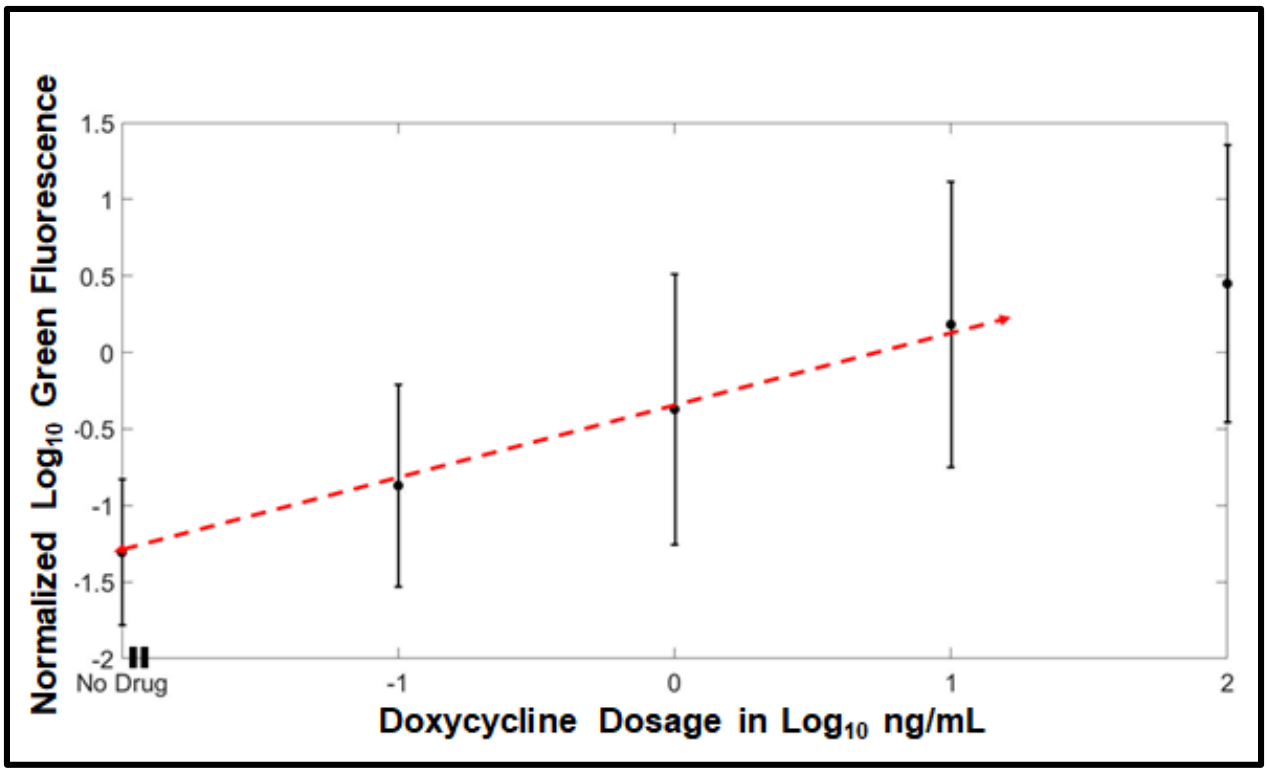


Figure 8. TetOn3G transcription system characterization *(continued)*
(b) The means of the distributions were plotted against the concentration of the inducer. The TetOn3G system appear to exhibit log linearity between 0 – 1 log ng/mL doxycycline concentrations. This log-linear portion encompasses a 30 ($10^{1.5}$) –fold G/B increase from basal levels. The low drug dosage required for transcriptional activation allows prevents cellular toxicity. The log-linear dosages allow us to determine preliminary concentrations to be used in the two node noise rheostat dosage experiment to obtain optimal resolution.

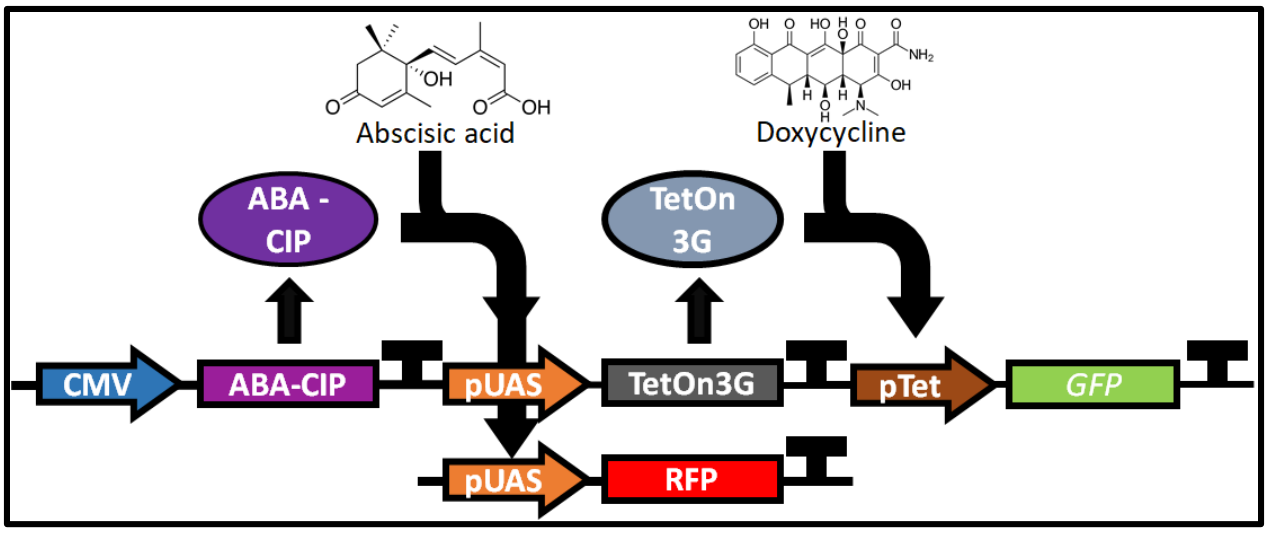


Figure 9. Two-node noise rheostat circuit schematic
The two-node noise rheostat is composed of the two synthetic transcription systems characterized in this paper. These two transcription systems are linked in an in-series architecture. What this means is that first transcription factor controls the expression of the second transcription factor which then controls the expression of terminal reporter, the green fluorescent protein (GFP). Both transcription factors are induced by their respective ligands and are dialable, meaning they exhibit dosage dependent activity. ABA-CIP drives the gene expression at the pUAS promoters. One pUAS controls TetOn3G expression while the other controls a red fluorescent protein (RFP) reporter. The red fluorescent protein allows us to keep track of the ABA-CIP activity and may tell us a relative number of TetOn3G transcription factors are in the cell. TetOn3G controls pTet – GFP and represents the output of the whole circuit. The levels of GFP are determined by dialing the ligands for the inducible transcription factors, and hence, a two node system.

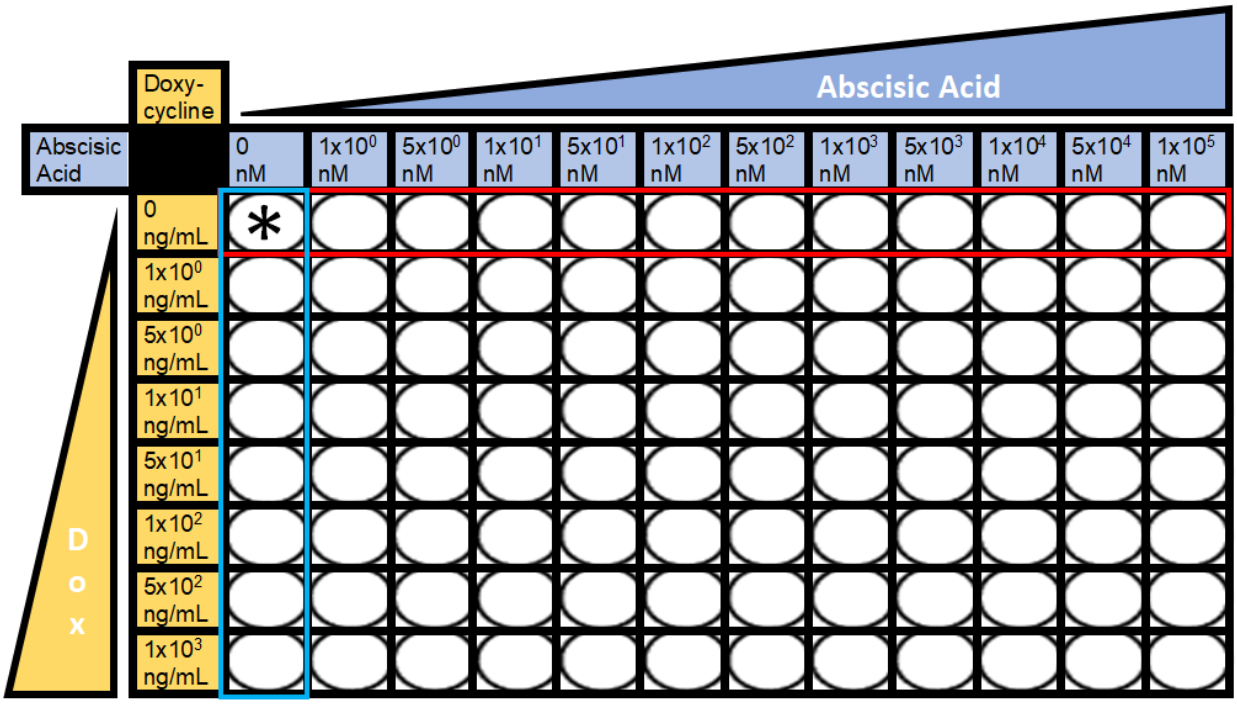


Figure 10. Two-node noise rheostat 96 well dosage experiment set up
20,000 transient transfected cells were plated into each well. Taking advantage of the 96 well format, abscisic acid and doxycycline concentrations were systematically graded along the width and length of the plate. The * well is a negative control for the plate, which represents the basal fluorescence the circuit. The wells highlighted in red is a control for abscisic acid and is a dosage experiment in itself for the ABA-CIP first node. The reporter output is the red fluorescence. The wells highlighted in blue is a control for the second node. Theoretically because ABA-CIP is not induced, TetOn3G is not produced and pTet would therefore be inactive. However, due to leakiness of the transcription systems, this may not always be the case. These wells are controls for understand the leakiness of the first node.

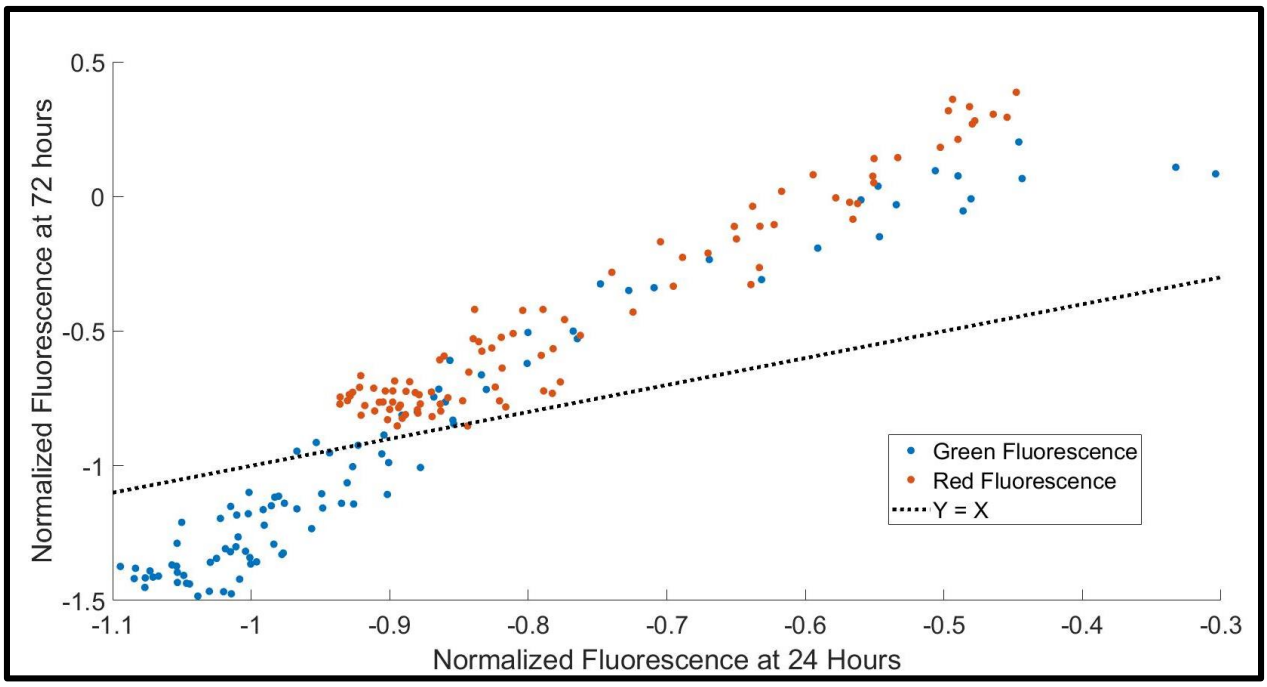


Figure 11. Two-node noise rheostat dynamics characterization
The normalized mean fluorescence of the cell populations at 24 hours and 72 hours were plotted against each other to observe the dynamics of induction. The green fluorescence was plotted in blue and the red fluorescence was plotted in red. The y = x is a reference line. If the fluorescence at 72 hours had not changed from the 24 hour fluorescence, then the points should lie along the y = x line. However, because many of the points lie above this y = x line, which suggests that after 24 hours, the fluorescence of the cells continued to grow as it had not reached steady state. This was the case for both the red fluorescence, which is the output of the first node, and green fluorescence, the terminal output. More importantly, this graph suggests that we should only be doing further data analysis with the 72 hour dataset as the 24 hour data set had not reached steady state.

a

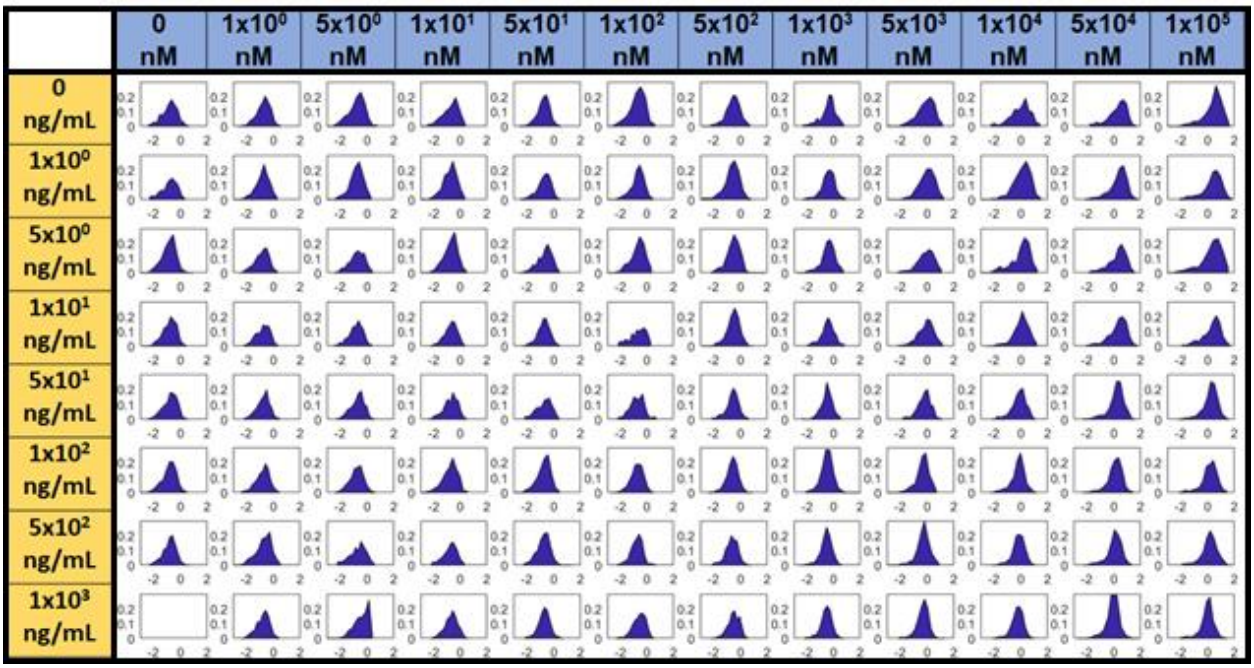
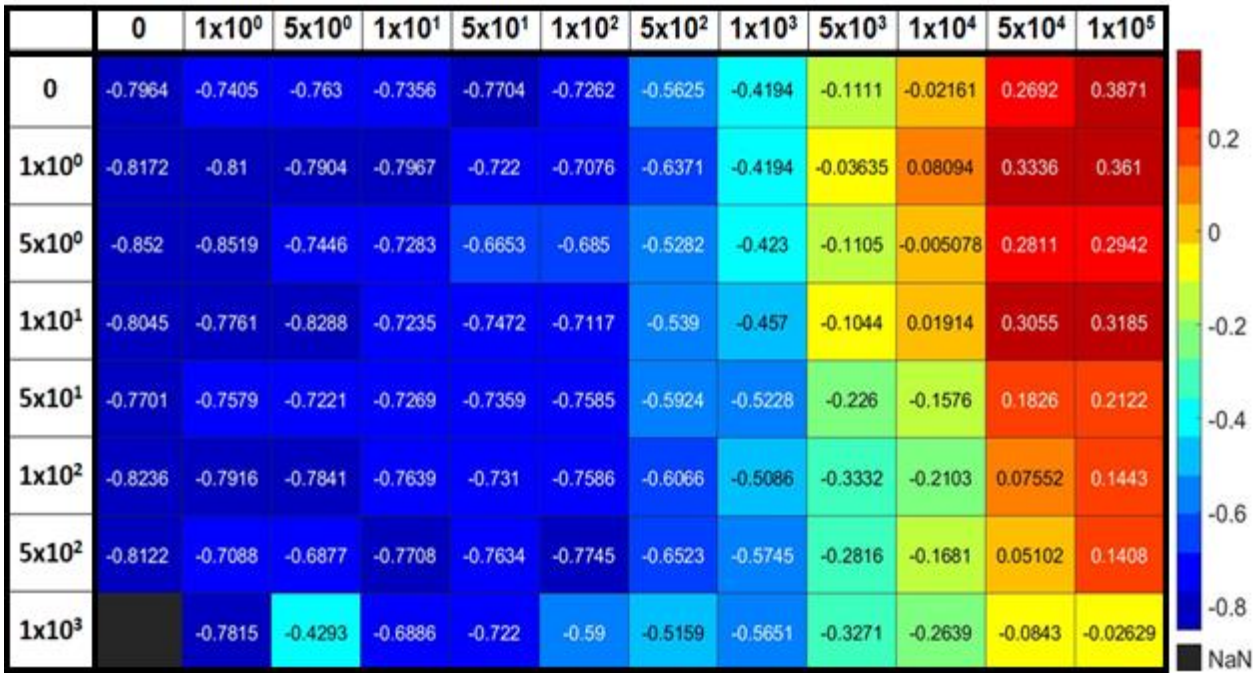


Figure 12. Red fluorescence output of 96 well dosage experiment for the two-node noise rheostat

The red fluorescent protein is the output of the first node and is dependent on the concentration of ABA which controls the availability of active ABA-CIP complexes. The abscisic acid concentrations is varied across width of the 96 well plate, which allows us to test 12 different concentrations of ABA in a graded manner. The normalized red fluorescence should then change across the width of the plate, but not across the length of the plate as red fluorescence should not be affected by doxycycline concentrations. (a) The histograms of normalized red fluorescence was plotted in the 96 format. All populations observed are all unimodal and right shifting as ABA increases in concentration. Interestingly, a similar dosage experiment done with the one node ABA-CIP system showed bimodality, but that may be an artifact of dynamics when the readings were obtained at 48 hours in the first experiment, compared to the readings at 72 hours in this experiment.

b



	0	$1x10^0$	$5x10^0$	$1x10^1$	$5x10^1$	$1x10^2$	$5x10^2$	$1x10^3$	$5x10^3$	$1x10^4$	$5x10^4$	$1x10^5$
0	-0.7964	-0.7405	-0.763	-0.7356	-0.7704	-0.7262	-0.5625	-0.4194	-0.1111	-0.02161	0.2692	0.3871
$1x10^0$	-0.8172	-0.81	-0.7904	-0.7967	-0.722	-0.7076	-0.6371	-0.4194	-0.03635	0.08094	0.3336	0.361
$5x10^0$	-0.852	-0.8519	-0.7446	-0.7283	-0.6653	-0.685	-0.5282	-0.423	-0.1105	-0.005078	0.2811	0.2942
$1x10^1$	-0.8045	-0.7761	-0.8288	-0.7235	-0.7472	-0.7117	-0.539	-0.457	-0.1044	0.01914	0.3055	0.3185
$5x10^1$	-0.7701	-0.7579	-0.7221	-0.7269	-0.7359	-0.7585	-0.5924	-0.5228	-0.226	-0.1576	0.1826	0.2122
$1x10^2$	-0.8236	-0.7916	-0.7841	-0.7639	-0.731	-0.7586	-0.6066	-0.5086	-0.3332	-0.2103	0.07552	0.1443
$5x10^2$	-0.8122	-0.7088	-0.6877	-0.7708	-0.7634	-0.7745	-0.6523	-0.5745	-0.2816	-0.1681	0.05102	0.1408
$1x10^3$		-0.7815	-0.4293	-0.6886	-0.722	-0.59	-0.5159	-0.5651	-0.3271	-0.2639	-0.0843	-0.02629

Figure 12. Red fluorescence output of 96 well dosage experiment for the two-node noise rheostat *(continued)*

(b) The mean red fluorescence was calculated and plotted as a heatmap in the same 96 well format. The gradation of colors range from blue for lower means, to red for higher means. It is clear that the means are increasing across the length of the plate, going from dark blue to dark red from left to right. Interestingly, the mean red fluorescence also appears to vary down the length of the plate, which was unexpected. As doxycycline concentrations increased, the red fluorescence decreased. The effect of that decreased is not as dramatic as the increased caused ABA concentrations.

c

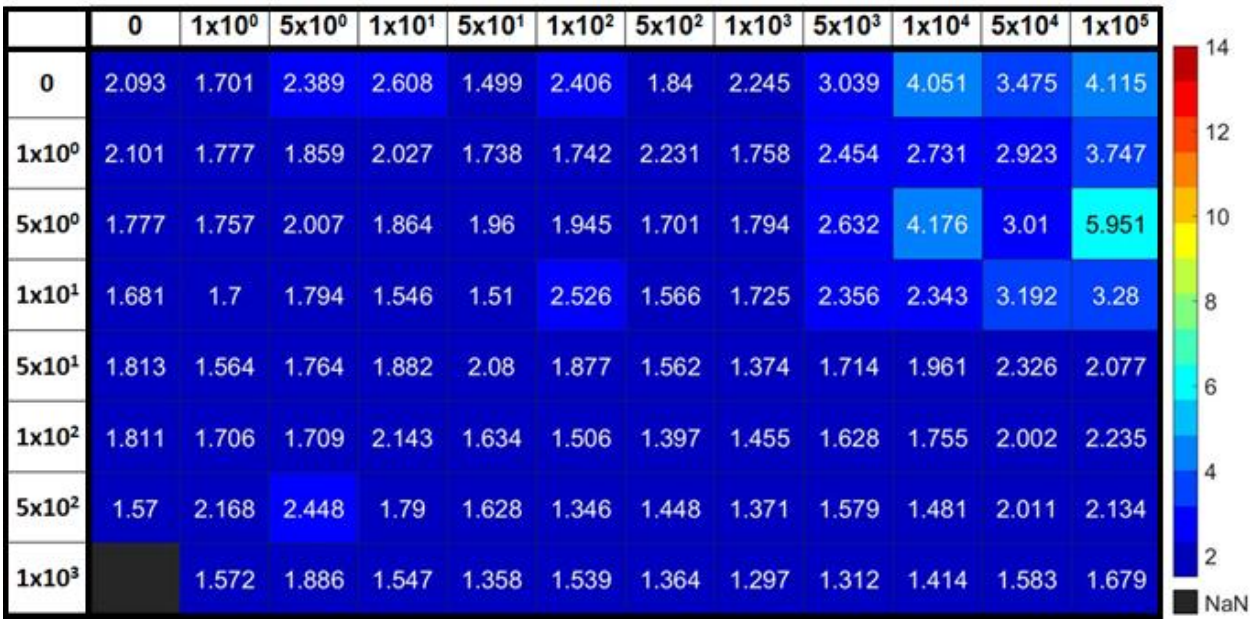


Figure 12. Red fluorescence output of 96 well dosage experiment for the two-node noise rheostat *(continued)*

(c) The geometric coefficient of variation (CV) was plotted as a heatmap in the 96 well format. The gradation of colors are the same as the mean heatmap. For the most part, the CV for the heatmap are low and consistent across the plate. The exception to this rule lies in the 4x4 well cluster on the top right hand corner of the plate. There are higher CVs in this cluster, possible due to the longer tail observed in those populations.

a

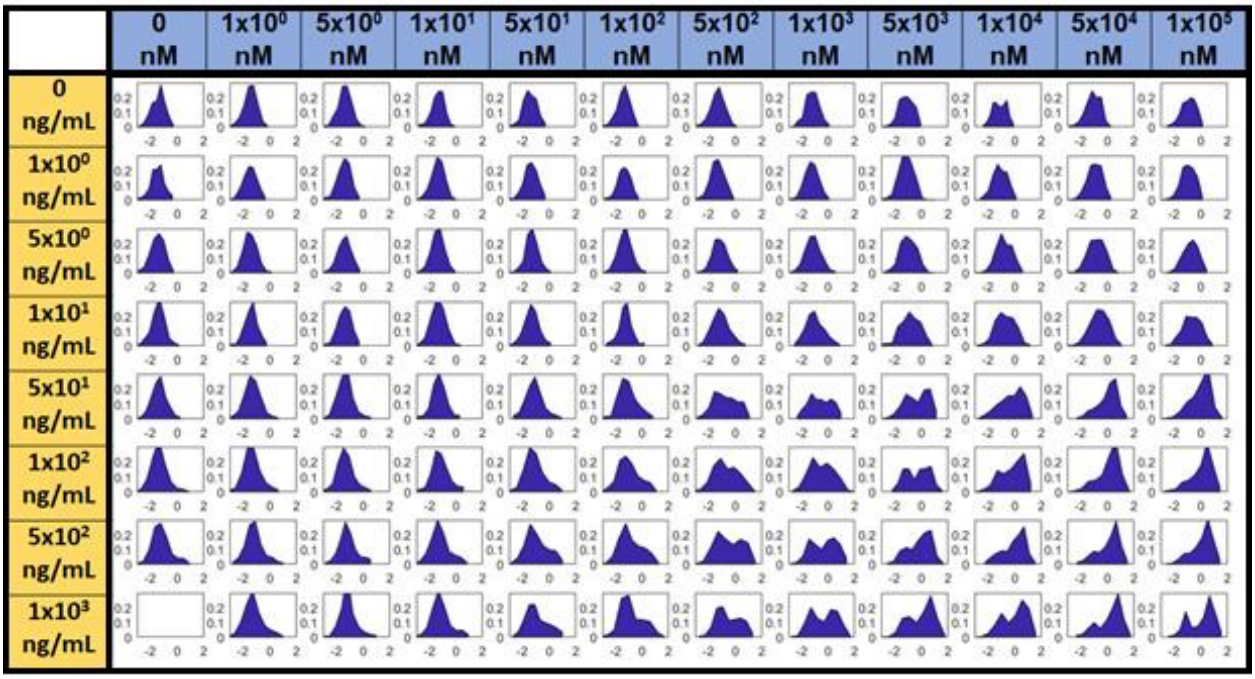


Figure 13. Green fluorescence output as the terminal node of the noise rheostat construct

The green fluorescent protein is the terminal output and is dependent on the concentrations of both doxycycline and ABA. Hence, the fluorescence of the population should vary both down and across the plate. (a) The histograms of normalized green fluorescence for each dosage is plotted in a 96 well format. For most of the populations, the histograms have unimodal distributions. However, for a cluster of wells with high concentrations of doxycycline ≥ 5 x 10^1 ng/mL and high concentrations of ABA ≥ 5x10^2 nM, is bi-modality observed. Interestingly, the bi-modality disappears at the highest concentration of ABA, except for the well with the highest ABA and doxycycline concentration.

b

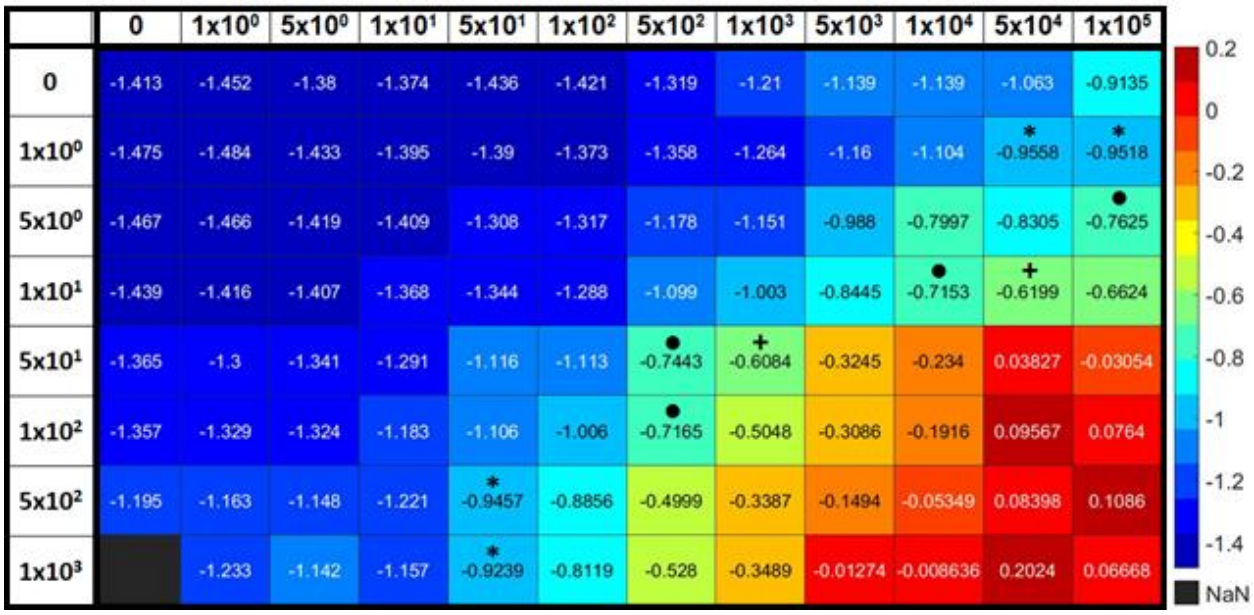


Figure 13. Green fluorescence output as the terminal node of the noise rheostat construct *(continued)*

(b) The means of the normalized green fluorescence was plotted as a heatmap in the 96 well plate format. The gradation goes from lower means represented by the dark blue, to the higher means represented by the dark red. The mean green fluorescence increases as both doxycycline and ABA concentrations increase. It also appears that different dosages of doxycycline and ABA can achieve the same mean green fluorescence. This phenomenon forms isomean curves as shown by looking across the plate for the same color box. A few isomean curves are selected to further inspect. They are marked with a symbol. Specifically we are looking at green fluorescence means - 0.940±0.015, marked with the symbol *; - 0.745±0.015, marked with the symbol ● ; and - 0.0605±0.015 marked with the symbol +. There are many other isomean curves that exist on this plate, but we will be observing this selection.

c

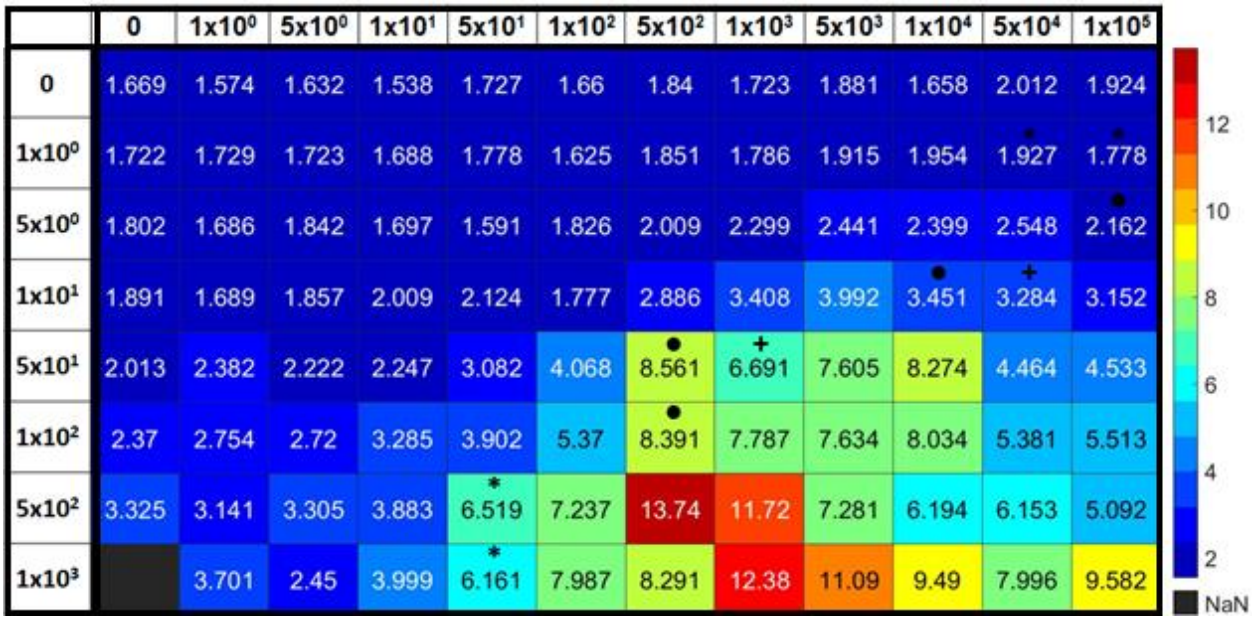
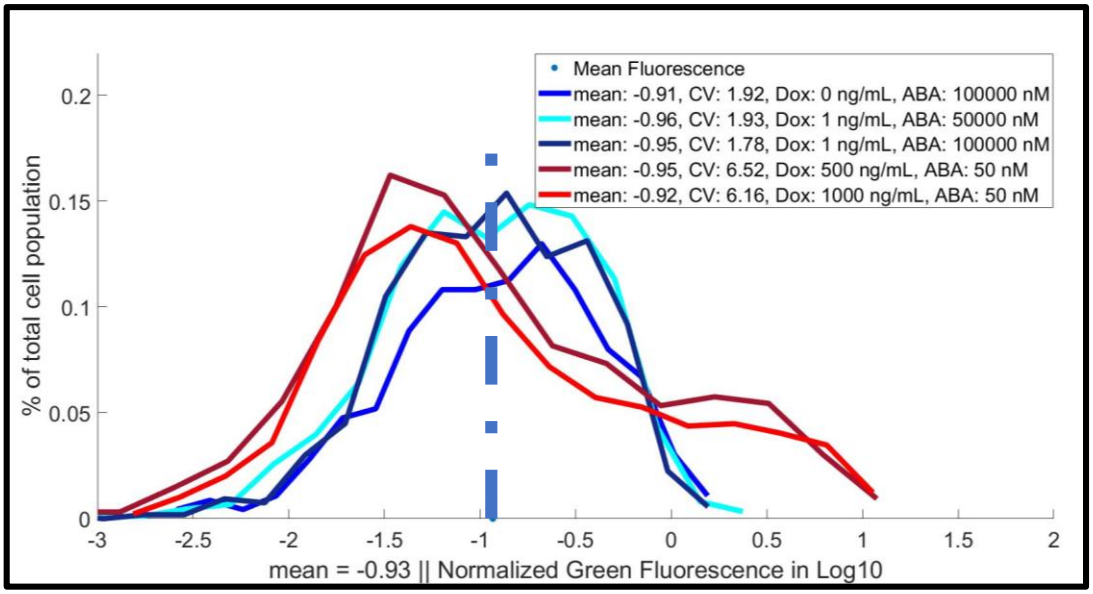


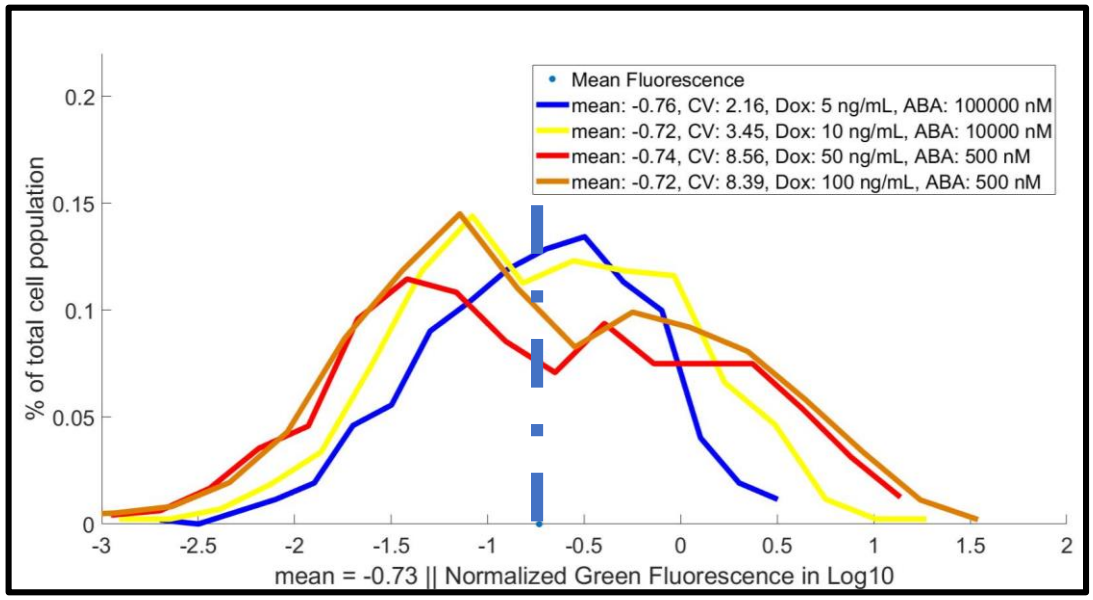
Figure 13. Green fluorescence output as the terminal node of the noise rheostat construct *(continued)*

(c) The geometric coefficient of variation (CV) of the normalized green fluorescence was plotted as a heatmap in the 96 well plate format. The gradation is similar to the mean heatmap and follow the similar color scheme. The CV seems to be lower for lower concentrations of doxycycline and abscisic acid. This is likely because the circuit is simply not turning at those dosages. There is a clustering of a higher CVs on the lower right hand side of the plate, with CV ranging between 1.7 – 13.7. The isomean curves that were marked with different symbols are marked here on the CV map as well for the same wells. It is clear that for wells with similar means, the CVs differ across those isomean wells. Wells marked with the * symbol with -0.940 ± 0.015 (mean ± S.D.) have differing CVs ranging from 1.7 – 6.5 CV. Similarly, wells marked with the ● symbol have mean G/B of 0.745±0.015 have different CV's ranging from 2.2 – 8.5.

a


Mean Fluorescence
mean: -0.91, CV: 1.92, Dox: 0 ng/mL, ABA: 100000 nM
mean: -0.96, CV: 1.93, Dox: 1 ng/mL, ABA: 50000 nM
mean: -0.95, CV: 1.78, Dox: 1 ng/mL, ABA: 100000 nM
mean: -0.95, CV: 6.52, Dox: 500 ng/mL, ABA: 50 nM
mean: -0.92, CV: 6.16, Dox: 1000 ng/mL, ABA: 50 nM
% of total cell population
0.2
0.15
0.1
0.05
0
-3
-2.5
-2
-1.5
-1
-0.5
0
0.5
1
1.5
2
mean = -0.93 || Normalized Green Fluorescence in Log10

b


Mean Fluorescence
mean: -0.76, CV: 2.16, Dox: 5 ng/mL, ABA: 100000 nM
mean: -0.72, CV: 3.45, Dox: 10 ng/mL, ABA: 10000 nM
mean: -0.74, CV: 8.56, Dox: 50 ng/mL, ABA: 500 nM
mean: -0.72, CV: 8.39, Dox: 100 ng/mL, ABA: 500 nM
% of total cell population
0.2
0.15
0.1
0.05
0
-3
-2.5
-2
-1.5
-1
-0.5
0
0.5
1
1.5
2
mean = -0.73 || Normalized Green Fluorescence in Log10

c

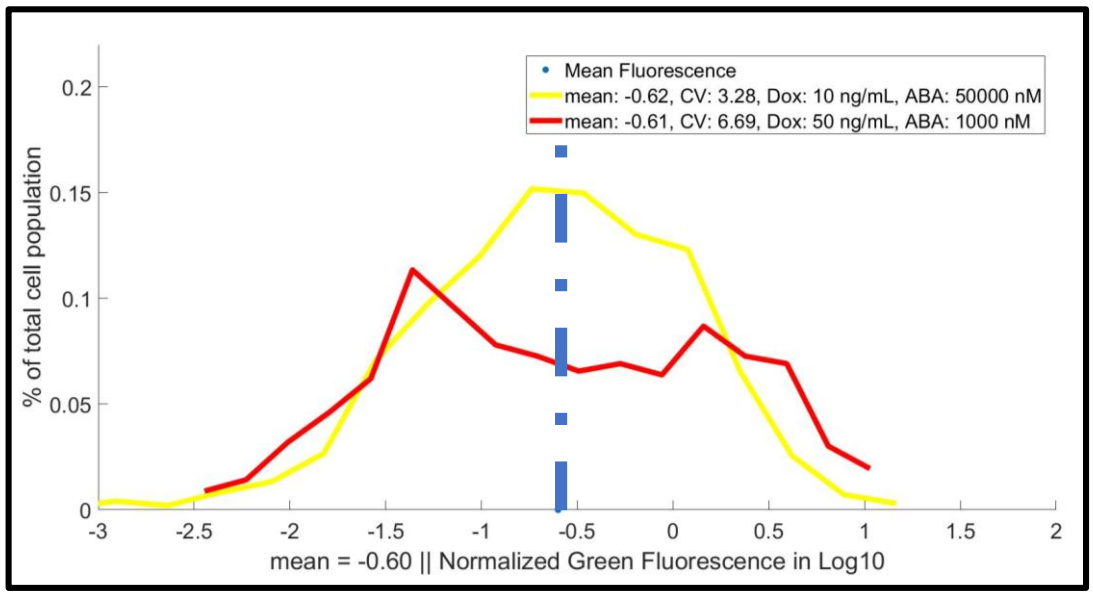

Mean Fluorescence
mean: -0.62, CV: 3.28, Dox: 10 ng/mL, ABA: 50000 nM
mean: -0.61, CV: 6.69, Dox: 50 ng/mL, ABA: 1000 nM
% of total cell population
0.2
0.15
0.1
0.05
0
-3
-2.5
-2
-1.5
-1
-0.5
0
0.5
1
1.5
2
mean = -0.60 || Normalized Green Fluorescence in Log10

Figure 14. Histograms of iso-mean wells show differing geometric CVs
The histograms of normalized green fluorescence shows that despite the fact these wells have the same mean fluorescence, the shape of the curves are very different. This reflects the schematic from the theoretical noise rheostat in figure 1 that shows that the noise level can be adjusted while the mean maintains steady. This is important for the study of noise, as we can now separate the mean gene expression from the noise of the gene expression. Low noise curves are colored with blue, while medium noise curves are colored by yellow. High noise curves are colored by red. (a) The wells with the mean fluorescence -0.940 ± 0.015 (mean ± S.D.) can be achieved with different combinations of ABA and doxycycline (ABA nM|Dox ng/mL) : 10^5 | 0, $5x10^4$ | 1, 10^5 | 1, $5x10^1$| $5x10^2$, $5x10^1$ | 10^3. At this mean, these different drug combinations are able to achieve low and high noise levels, exhibited by CVs ranging from 1.78 – 6.52. (b) The wells with the mean fluorescence 0.745±0.015 can be achieved with different combinations of ABA and doxycycline: 10^5 | 5, 10^4 | 10, $5x10^2$ | $5x10^1$, $5x10^2$ | $1x10^2$. At this mean, these different drug combinations can achieve the entire range of noise with low noise of 2.16 shown by the blue curve, achieved with 10^5 | 5 drug combination. While a medium noise level of 3.45 shown by the yellow curve is achieved with 10^4 | 10. The high noise curves are shown in orange and red with CVs of 8.56 and 8.39 are achieved using drug combinations of $5x10^2$ | $5x10^1$ and $5x10^2$ | 10^2 respectively. (c) The wells with the mean fluorescence - 0.60 ± 0.015 can be achieved with different drug combinations of ABA and doxycycline: 5 x 10^4 | $1x10^1$ and $1x10^3$ | $5x10^1$. At this mean, a medium and high noise level is achieved with CV of 3.28 and 6.69 labeled in yellow and red respectively.

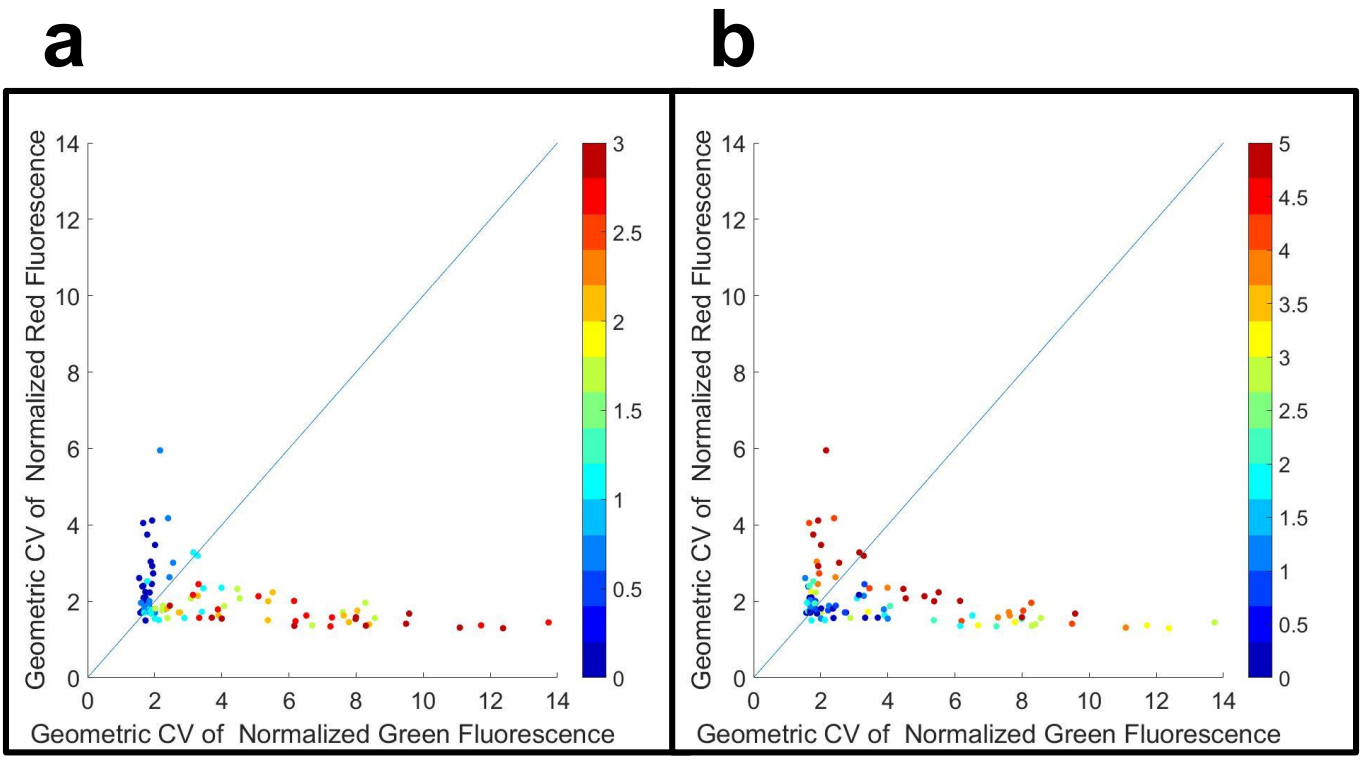


Figure 15. Geometric CV of red fluorescence plotted against geometric CV of green fluorescence colored by concentration of doxycycline and abscisic acid Geometric CV of normalized red fluorescence is plotted against normalized green fluorescence. The color of the dots represents the drug concentrations that produce those CVs with blue representing lower concentrations while red represents high concentration. The y = x line is a reference line to compare the CVs. Points above this line has higher noise in the red fluorescence that the green fluorescence, while points below this line has higher noise in the green fluorescence than the red fluorescence. The green fluorescence noise has a higher range than the red fluorescence suggesting that the ABA-CIP one node can not produce the high noise range alone. (a) The points are colored by the $\log_{10}$ concentration of doxycycline. There appears to be a cluster of high noise levels in red fluorescence and low noise in green fluorescence at low concentrations of doxycycline as shown as the cluster of blue points. At high concentrations of doxycycline as colored by the red points, the noise of red fluorescence is low, but the green fluorescence appear to span the range of noise levels from 2-14 CV units. (b) The points are colored by $\log_{10}$ concentrations of ABA. There are less clustering that is observed for ABA dosages compared to doxycycline dosages. However, it makes sense that at low concentrations of ABA, colored by the blue, there is low noise in the green and red fluorescence. Meanwhile, the high concentrations of ABA, colored by the red dots, the noise of red fluorescence and green fluorescence are inversely related.

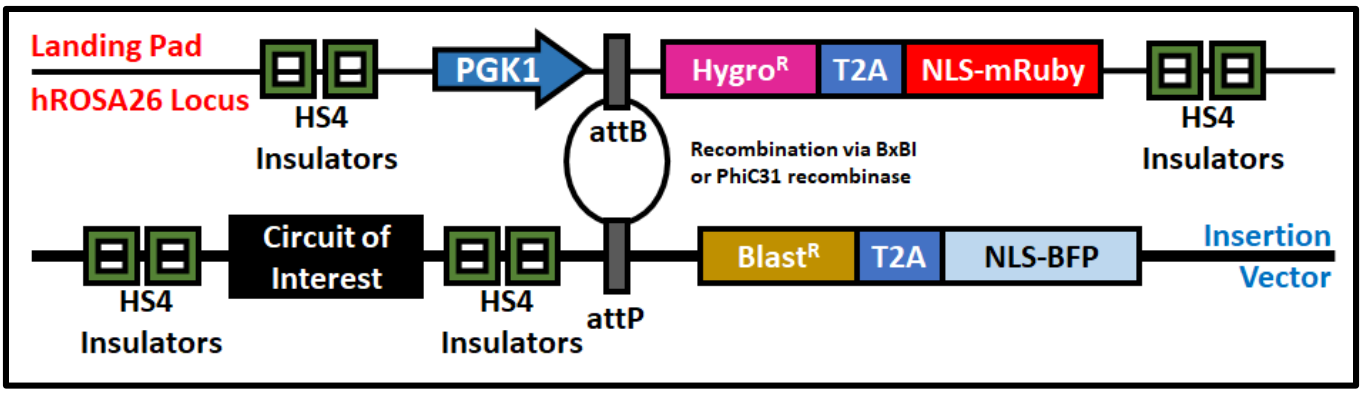


Figure 16. Landing pad and insertion vector schematic
A genetic construct landing pad refers to a known integration site within a host genome. Landing pads can control for copy number and transcriptional silencing. This landing pad was constructed using CRISPR-cas9 mediated homologous recombination. The landing pad construct was first integrated in the hROSA26 locus and single cell sorted to create clonal cell populations. The landing pad consist of two HS4 insulator on both sides of the transcriptional unit. The human Pgk1 constitutive promoter is upstream of an *attB* site and polycistronic cassette with antibiotic resistance to hygromycin and a red fluorescent protein mRuby with a nuclear localization signal. The insertion vector, which integrates the circuit of interest into the landing pad, uses *attB/attP* responsive recombinases (PhiC31 and BxBI). If a construct was inserted in the correct orientation, the polycistronic cassette on the insertion vector gains the activity of Pgk1 while the polycistronic cassette on the landing pad becomes promoterless. If the circuit was inserted in the correct place and in the correct orientation, the cell goes from constitutively producing mRuby to producing blue fluorescent protein.

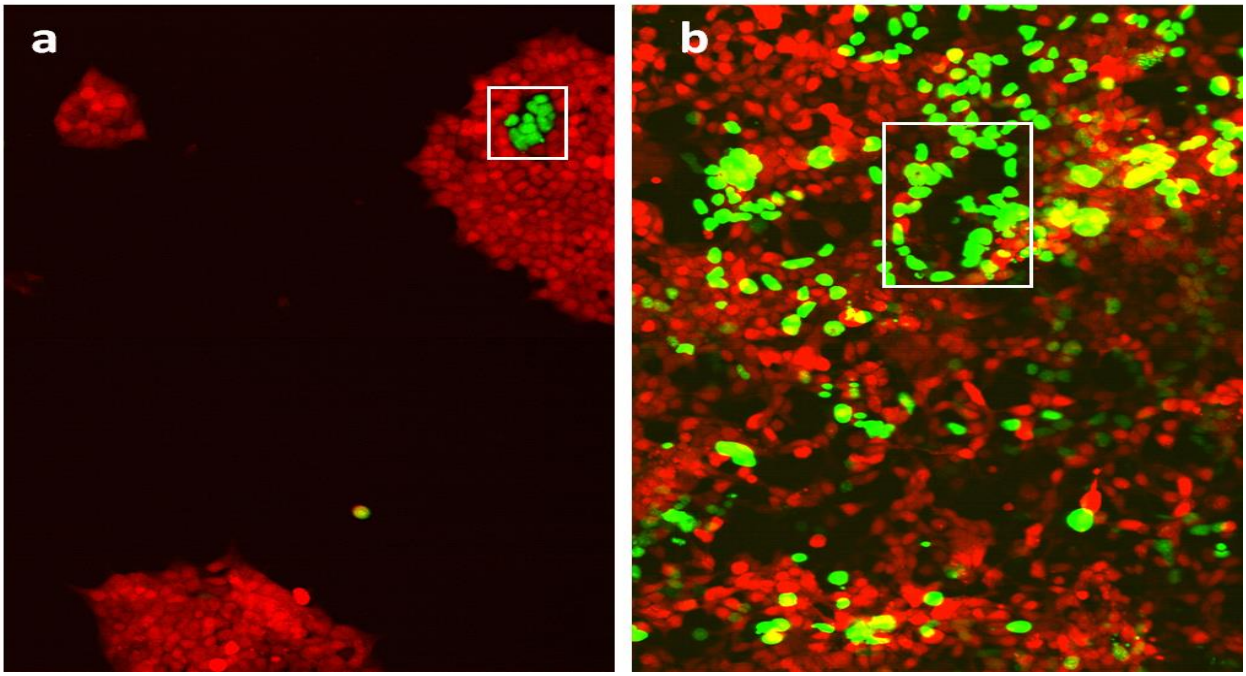


Figure 17. Clonal landing pad cell lines tested with PhiC31 and BxBI recombinase *attB/attP* pairs.
Landing pad cell lines were co-transfected with a constitutive promoter driving a green fluorescent protein construct within the insertion vector and an expression vector containing PhiC31 or BxBI recombinase. These cells were imaged seven days post transfection using confocal microscopy. No blue fluorescence was observed suggesting that the insertion vector *attP* sequence is in the wrong orientation. (a) The landing pad cell line has a *BxBIattB* and the insertion vector used here is the *BxBIattP.* A cluster of cells that are green but not red are observed in the white rectangle. Those cells have integrated the green fluorescent reporter construct within the landing pad and thus have interrupted the expression of the red fluorescent protein in the polycistronic cassette. (b) The landing pad cell line has a *PhiC31attP* site and the insertion vector has a corresponding *PhiCattB* site. A lot of cells are observed to have integrated into the landing pad, exhibiting only green but not red (as shown in the white box). There are also some cells that exhibit both red fluorescence and green fluorescence, as shown in by the yellow cells. This suggests that the green fluorescent reporter was integrated into the wrong location of the genome, as the construct did not disrupt red fluorescent protein expression at the landing pad. It is highly unlikely plasmid DNA from transfection still exist as few floating transcriptional active DNA in cells that have gone through seven cell divisions without a lot of dilution and foreign DNA degradation to prevent continued transient expression of the green fluorescent marker.