Foundational Platform for Mammalian Synthetic Biology

by

Noah Davidsohn

SUBMITTED TO THE DEPARTMENT OF BIOLOGICAL ENGINEERING
IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF

DOCTORATE OF PHILOSOPHY
AT THE
MASSACHUSETTS INSTITUTE OF TECHNOLOGY
DECEMEBER 2012

Signature of Author:.....

Noah Davidsohn
Department of Biological Engineering
January, 2013

Certified by.....

Ron Weiss
Associate Professor of Biological Engineering
Co-Thesis Supervisor

Certified by.....

Jacob Beal
Scientist, Raytheon BBN Technologies
Co-Thesis Supervisor

Accepted by.....

Doug Lauffenburger
Ford Professor of Bioengineering
Chair, Biological Engineering Graduate Committee

Foundational Platform for Mammalian Synthetic Biology

by
Noah Davidsohn

Submitted to the Department of Biological Engineering on January 13, 2013 in Partial Fulfillment of the Requirements for the Degree of Doctor of Philosophy in Biological Engineering

Abstract

The emergent field of synthetic biology is different from many other biological engineering efforts, in that its roots, design principles, and forward engineering perspective have been adopted from electrical engineering and computer science. Synthetic biology is uniquely poised to make great contributions to numerous fields such as bio-fuel, energy production, agriculture and eco-remediation, national defense, and biomedical and tissue engineering. Considerable progress has been made in engineering novel genetic circuits in many different organisms. However, not much progress has been made toward developing a formal methodology to engineer complex genetic systems in mammalian cells. One of the most promising areas of research is the study of embryonic and adult stem cells. Synthetic biology has the potential to greatly impact the progression and development of research in this area of study. A critical impediment to the development of stem cell engineering is the innate complexity, little to no characterization of parts, and limited compositional predictive capabilities.

In this thesis, I discuss the strategies used for constructing and optimizing the performance of signaling pathways, the development of a large mammalian genetic part and circuit library, and the characterization and implementation of novel genetic parts and components aimed at developing a foundation for mammalian synthetic biology. I have designed and tested several orthogonal strategies aimed at cell-cell communication in mammalian cells. I have designed a characterization framework for the complete and proper characterization of genetic parts that allows for modular predictive composition of genetic circuits. With this characterization framework I have generated a small library of characterized parts and composite circuits that have well defined input-output relationships that can be used in novel genetic architectures. I also aided in the development of novel analysis and computational tools necessary for accurate predictive composition of these novel circuits. This work collectively provides a foundation for engineering complex intracellular transcriptional networks and intercellular signaling systems in mammalian cells.

Thesis Supervisor: Ron Weiss
Title: Associate Professor of Biological Engineering

Acknowledgments

"It is easier to go down a hill than up, but the view is from the top" – **Arnold Bennett**

My original goal in grad school was to work on nanotechnology with applications in biology. However, I didn't find any of the projects in electrical engineering to be interesting or awe-inspiring. My naïve perception was that we would be able to create little robots on the nano-scale that could interact with our cells like in cartoons. I then came across Prof. Ron Weiss's research and the unique perspective he took on biological engineering. The research he was doing treated cells as tiny computers and our job as synthetic biologist was to engineer and program them to do amazing things. After hearing his pitch, I was sold. I knew this was the field I wanted to work in. I am very thankful for all the knowledge he has imparted to me in my 6 years in his lab. He has taught me to think critically and never to settle for anything but my best. I am also privileged to have Jacob Beal Phd, Prof. Doug Lauffenburger, Prof. Jacquin Niles, and Prof. Alan Grodzinsky on my thesis committee. The guidance they have given me in the last three years has been invaluable.

I have been lucky to have had a spectacular lab environment for the majority of my graduate career. Saurabh Gupta has been more than a lab mate along the way, he has been a friend, roommate, and fellow indentured servant. I have enjoyed his company over the years and the in depth philosophical conversations we have had. Patrick Guye has been a beacon of reason and has provided amazing guidance over the time I have known him. In the time I have worked with him, he has shared with me a mere drop of the vast ocean of scientific knowledge he possesses and I have benefitted greatly from it. I am indebted to Sairam Subramanian, who mentored me at the beginning of my graduate career. Eran Bram, a recent addition to the Weiss lab, but in our short time I have enjoyed numerous coffee breaks and great intellectual debates and insights into his creative facets. I am indebted to Lila Wroblewska who listened to all my complaining. I am also thankful to our collaborators Jake Beal, Aaron Adler, and Fusun Yaman who have contributed much to my project and made the success of my research possible. I am especially appreciative of Stephen Firsing, whose tireless work and support made sure I had everything I needed, when I needed it. I am also grateful to the rest of the Weiss lab members for all their support.

My family has been on this journey with me as much as anyone. I would like to thank my sister Arielle for always boosting my ego and bragging to her friends about me. The support from my mom and dad has been endless. I am thankful to my dad for letting me bounce ideas off him and giving me the big picture perspective that's so important in science. My mom has been there for all my accomplishments and failures and has made me feel good no matter what. With out their support I might not have made it through.

I am also thankful to DARPA for the funding of my project, without this support my research would not have been possible.

List Of Figures

Table Of Contents

Chapter 1

1. Introduction

Synthetic gene networks are at the forefront of systems biology and provide a framework for understanding and engineering life. The field of synthetic biology offers tremendous potential to both understand natural biological phenomena and to re-engineer existing cells and organisms to equip them with synthetic capabilities that are useful for predetermined functionalities[1].

The origins of synthetic biology are seen in the initial efforts of the field. The first devices built were biological analogues of simple electrical engineering components. They created gene networks to create basic 'devices' such as oscillators, memory elements and transcriptional cascades aimed at signal amplification[2-4]. Other success stories include a transcriptional 'pulse generator' and engineered cell-cell communication in prokaryotes resulting in controlled population control or predictable pattern formation[5, 6]. The full and well-defined characterization of the input-output relationships of these basic electrical devices allowed the successive engineering of a plethora of more complex systems. However, a sufficient and well-defined method of characterization of the biological analogues has not yet been achieved. With the knowledge obtained from these initial engineered circuits the field began to focus on more advanced application-oriented projects such as gene therapy to cure illness and prevent disease, metabolic engineering, drug production, and tissue engineering.

One such health problem where synthetic biology can hope to contribute is Diabetes Mellitus, which is characterized by an immune-mediated loss of pancreatic β cells (cells that produce insulin). It is a devastating, currently incurable disease that affects over 8% of the population in the United States alone[1]. Recent developments in genome technologies, tissue engineering and synthetic biology offer exciting possibilities to establish highly accurate and robust approaches for predictable and controllable cell fate regulation which can be used to address the root causes of diabetes.

Synthetic biology holds the promise for one such cure, engineered tissue homeostasis. Artificial tissue homeostasis involves engineering an isogenic population of human embryonic stem cells (hESC) or adult (e.g. iPS) stem cells to have the capability to produce a stable population of insulin producing β cells. In the engineered system, cells are not simply exogenously induced to differentiate, but rather are programmed to sense and respond to changes in their environment and the state of other cells, allowing them to coordinate their collective behavior based on the needs of the system. In this system, a growing population of engineered mammalian embryonic stem cells will communicate using an artificial signaling pathway. They must be able to detect the size of their own population (quorum sensing I), detect the size of insulin producing β cell population (quorum sensing II), be able to decide when it is appropriate to produce more β cells (differentiate), when it is necessary to divide or stop dividing (proliferation and quiescence), and each cell must know which state it is in (hESC or β cell). A cell in this system must

constantly be making calculations based on inputs from its surrounding environment. A genetic representation of this system is seen in chapter 4 Figure 6a.

This system is extremely complex and would be intractable with out the development of computer aided design tools to properly predict the resulting function of large genetic circuits. Computational design and modeling is needed to guide the experimental construction of the proposed complex synthetic system.

The goal of this thesis is to create foundational tools for synthetic biology in mammalian cells by creating new orthogonal signaling pathways, designing a characterization framework for transcription factors, developing analytical and modeling tools to aid in the design of more complex genetic circuits and constructing and verifying the functionality of novel genetic circuits. The ability to obtain sufficient information about individual genetic parts and develop a computer model that will use this information to quantitatively predict novel interactions of the individual parts will *significantly* advance the field of synthetic biology. It will allow for the design and construction of much larger and more complex systems then have been seen thus far.

1.1. Thesis Statement

The predictive composition of genetic circuits from well-characterized parts has been the central dogma of synthetic biology since its inception. However, the innate complexity of biological systems and lack of knowledge has thus far prevented the development of a standard characterization technique that will yield enough high quality information about individual parts to be used in computer-aided design. Mammalian cells provide additional difficulties and complications owing to the heterogeneity of individual cells, the methods used to introduce foreign DNA, and the lack of intercellular communication systems to coordinate population behavior.

Having a method to coordinate population wide behavior is necessary if we wish to create complex systems that can operate in diverse environments and respond spatially and temporally to extracellular cues. Developing intercellular communication systems that are orthogonal to innate signaling pathways has proved somewhat difficult in mammalian cells. To date, a completely orthogonal extensible, intercellular communication pathways has not been developed that is robust and can coordinate multicellular behavior for the purpose of tissue engineering. Also, there have not been any examples of predictable composition of intracellular genetic circuits in mammalian cells.

A synthetic orthogonal communication system combined with a novel characterization process can be used to coordinate population behavior and predictively compose novel intracellular genetic circuits to create a foundation for mammalian synthetic biology.

The approach described below outlines methods to engineer cell-cell communication networks and to create a characterization framework for the predictive composition of genetic circuits. This approach was used to construct and test 4 different communication systems, characterize 3 transcription factors, and test 6 different transcriptional cascades. The ability to coordinate population behavior in mammalian cells and design and predictively engineer genetic circuits will create a foundation for the development of autonomously regulated tissues and a myriad of other applications.

1.2. Approach and Summary of Contributions

A major thrust of synthetic biology has been on developing modular libraries of parts with well-characterized operations. A direct result of having such libraries is being able to combine these parts in novel architectures to carry out complex logic functions. Synthetic biology in mammalian cells has the potential for direct applications in humans but lacks the necessary development of tools and paradigms required to engineer mammalian synthetic systems.

The better understood part libraries have thus far been composed of transcription factors that directly regulate the expression of one or more proteins at the transcriptional level. A large amount of work has gone into creating libraries of orthogonal transcription factors in prokaryotes (Zinc Finger proteins in the Collins lab, TetR homologs in the Voigt lab) [108,109] . Also, numerous tools have been developed along these lines to help synthetic biologists quickly and

combinatorically create large libraries of parts and components for prokaryotes (i.e. Bio Bricks, BglBricks, Golden Gate, Gibson etc...)[7, 8].

Engineering complex systems becomes intractable without a fundamental infrastructure. Mammalian synthetic biology is no exception; unfortunately, it does not have a standard parts library yet. Several standard prokaryotic parts libraries exist each with their own limitations, but they provide a framework on which to build upon. A quick efficient method is needed that can create large combinatorial libraries of promoter-gene pairs for use in mammalian systems. These libraries of parts also need a well-defined method of characterization in order to be used in novel networks and circuits.

Another challenge when working with mammalian cells is our inability to control large multicellular systems with precise spatio-temporal control. Although independently operating engineered cells can perform tasks of varying complexity, more sophisticated coordinated tasks are possible with populations of communicating cells. However, a number of challenges confront the successful development of intercellular signaling components and the utilization of these components to form multicellular systems. When a new signaling pathway is initially incorporated into a host, signal synthesis and response elements often require considerable modification for proper function. Furthermore, once a functional set of devices is constructed to enable intercellular communication, these communication devices must be interfaced with intracellular signal processing modules that determine how the cell responds to the intercellular messages received [1].

1.2.1. DNA Part Library

As part of my thesis, I supported the development of an infrastructure for mammalian part construction. In short, I constructed and tested over 70 plasmids for 'part' libraries and expanded the construction paradigm from a two part system to include a third part, a 3' UTR. I have also contributed over 300 promoter-gene pairs to our expression library. The new construction paradigm will allow for the quick generation of a part and component library similar to what has been seen in prokaryotes. It is dependent on Gateway® technology from Invitrogen and Gibson assembly [8]. Gateway® is an *in vitro* recombination based method of cloning that has extremely high efficiencies. There is a library of promoters and a similar library of genes. Any promoter can be combined with any gene to create a promoter-gene pair that can be directly expressed in mammalian cells. The first stage of construction creates a promoter-gene pair, or expression vector, and these vectors can also be combined with other expression vectors to create large genetic circuits contained on a single plasmid. The second stage construction involves using Gibson assembly with a dedicated region of 40bp overhangs on either side of the expression cassette. Populating these libraries with as many known parts as possible, allows for characterization and future use in larger more complex systems.

1.2.2. Cell-cell Communication Systems

One of the traits of multi-cellular systems is coordinated cell behavior in a population. To realize this in a synthetic system, one needs to have a way for the

cells to communicate with each other in order to synchronize their states or exchange other information that enables control of a population. Engineering a cell-cell communication system presents significant challenges. There needs to be the generation of a signal inside a cell, reliable transmission of the signal across cell membranes, and recognition and decoding of the signal in the recipients triggering the appropriate response. Other pre-requisites are that the signal be non-toxic, have a reasonable half-life that ensures its stable production and detection, and that the signal does not have crosstalk with other endogenous pathways[1].

I initially tried engineering a cell-cell communication system based on acyl-homoserine lactone (AHL) that had promising results from a previous graduate student[1]. I tried implementing some improvements that should have produced the desired behavior, but unfortunately this system did not yield any fruitful results. Based on my experience in the lab with lenti virus particle production, I sought to engineer a virus-like-particle (VLP) based communication system that would carry any reasonably sized protein inside a protected particle. Cell type specific targeting could be achieved by pseudotyping the virus with various membrane receptors that would make it selectively infect certain cells. The immunological concerns combined with other challenges discussed in chapter 4 deterred continued effort along this path.

Another group had developed an intercellular communication system based on the TEV (tobacco etch virus) protease. It was initially used in conjunction with endogenous pathways to produce novel actuation from existing signals [9]. The TEV system works by anchoring a transcription factor (TF) to the membrane and

selectively cleaving it with the TEV protease to activate or repress transcription. Dr. Patrick Guye and myself worked on developing an orthogonal version of this communication system with promising preliminary results.

I also engineered a two-component system that demonstrated promising preliminary results. This system is based on PhoB, a transcription factor found in prokaryotes, that is combined with the signaling network of *Arabidopsis thaliana* and responds to the signaling molecule IP (isopentenyl adenine) [10]. A sender cell contains the IP synthase and the receiver cells contain a membrane bound receptor for IP, AHK4, a phospho-transferase, AHP3, and the transcription factor PhoB that regulates expression based on its phosphorylation state. Once IP binds AHK4, a phospho relay system would be activated ending in transcription from the PhoB cognate promoter.

1.2.3. Characterization Framework

Synthetic biology creates novel circuits from existing parts to solve important biological problems. A good characterization system does not yet exist that can produce predictable behavior of parts in novel contexts. I have participated in developing a characterization framework to elucidate the input/output relationship (also called a transfer function) for each part. The information obtained from such characterizations would be used to modularly compose circuits independent of the part original characterization circuit. The characterization method is able to garner an input-output relationship from a transcription factor

that is extensible over an entire library of parts. These parts are sufficiently characterized for their use in predictable engineering of novel genetic circuits.

1.2.4. Predictive composition

A major thrust of synthetic biology has been the predictive design of engineered circuits from well-characterized parts; similar to what has been seen in the electronics industry. However, the goal of quantitative predictive engineering of biological circuits has not previously been achieved in mammalian cells. High quality characterization data combined with a tight coupling to high-level design tools, developed during the TASBE (A Tool-Chain to Accelerate Synthetic Biological Engineering) project [110], has allowed us to achieve this long sought after goal. I have characterized 3 biological parts using an innovative characterization framework described previously. I also created 6 different genetic transcriptional cascades based on these 3 parts to verify our quantitative predictive capabilities. Predictions with <1.6 fold mean squared error have been achieved.

1.3. Thesis Outline

In the remainder of the thesis, Chapter 3 discuses the DNA assembly technology used to create the libraries of parts and circuits. One application of these foundational technologies (i.e., characterization and cell-cell communication) is the design of an artificial tissue homeostasis system, described in Chapter 4. Chapter 5 discusses the intercellular communication systems I pursued to coordinate population level behavior. Chapter 6 presents the contributions towards a

characterization framework and the quantitative predictive capabilities achieved from this novel method. Finally Chapter 7 looks at the future of the exciting field of synthetic biology and presents a few applications that can be pursued based on efforts and results found in this thesis.

Chapter 2

2. Background and Significance

Mammalian synthetic biology has made great progress in recent years. Most efforts were seen in replicating the accomplishments achieved in lower eukaryotes and prokaryotes [1,9,10,11,16,23,24]. However, these successes should not be diminished because of their previous accomplishment in lower organisms. Mammalian cells present new and difficult challenges that must be overcome in order to attain comparable results to these lower organisms. The increased complexity and compartmentalization of mammalian cells into organelles presents a further obstacle, as proteins introduced must be trafficked to the appropriate target sites for proper functionality unlike the bacterium where the cytoplasm is a bag of enzymes, proteins and DNA which interact with one another through probabilistic collisions and mutual attractions [26]. For example, in mammalian cells, DNA binding proteins need to be targeted to the nucleus. Also, owing to the complexity of the mammalian transcription machinery, activation and repression domains may need to be fused to such DNA binding proteins without compromising the functionality of their native domains, to ensure transcriptional activation and repression. These are but a few of the complications that need to be considered when engineering circuits in mammalian cells[1].

2.1. Engineering and characterization efforts

Some of the earliest efforts in eukaryotic cells were to re-create simple systems from lower organism such as synthetic cascades, toggle switches, and oscillators. Kramer et al. in 2003[11], created a three-stage cascade similar to one created in bacteria by Hooshangi et al. [5]. They are similar in that they are both three-stage cascades, but the one in mammalian cells is composed of activators where as the bacterial one is composed of repressors. In the mammalian cascade they were able to tune the output by repressing any of the activators with small molecule inhibitors. By disrupting the cascade at different places they were able to precisely tune the behavior of the output through the degree of leaky expression that carried through the cascade [11].

Some of the earliest efforts in bacteria were seen by the creation of genetic toggle switches that operated through cross repression [3]. A similar epigenetic toggle switch in mammalian cells was created by Kramer et al. 2004b in which a DBD (DNA binding domain) was fused to a KRAB (kruppel associated box) domain [12] and through cross repression created a toggle switch memory element[13]. The toggle switch consisted of PIP-KRAB and E-KRAB both cross-repressing each other. The system was switchable by the addition of a small molecule inducer that repressed its cognate repressor and allowed expression of the other arm of the memory element.

Oscillators are ubiquitous in natural systems. Cell cycle regulation and circadian rhythms are just some functions governed by consistent robust oscillators.

Most efforts to understand natural oscillators are seen in reverse engineering efforts to determine functional importance of various elements [14]. Synthetic biology aims to forward engineer systems and allows for in depth understanding of the systems created. One of the first and simplest oscillators constructed was the Goodwin oscillator where the expression of a repressor stopped its own production [15]. A more complex oscillator was engineered by Elowitz et al, called the "repressillator". It consisted of three repressors where each acted on its successor and created a closed loop of repression [4].

The next logical step for oscillator construction was to include not only negative feedback but also have positive feedback. The combination of positive and negative feedback systems is routinely seen in endogenous oscillators and has been shown to increase the stability and robustness of oscillations. Hasty et al. created such a system in bacteria that was fast, robust, and tunable and included both positive and negative feedback [16]. One of the first mammalian oscillators also contained both positive and negative feedback and was created by Fussenager et al. [17]. Its network topology was very similar to the Hasty oscillator.

Of all these devices that were created and characterized, none were done so in a predictive manner. Each oscillator mentioned above had an accompanying computer model, however, the models were not used for quantitative predictions. The computer models and simulations were able to provide qualitative information about the device (i.e. which parts were more sensitive to change and which were more robust) and this helped determine areas to focus their engineering efforts. The oscillators in the end were developed through years of painstaking trial-and-

error and researchers' intuition. **A long-standing goal of synthetic biology is to rapidly engineer new regulatory circuits from simpler regulatory elements whose properties have previously been characterized individually [101].**

There have been numerous efforts at characterization but they have not yet yielded a generalizable method to characterize an individual part such that a quantitative prediction can be made from the connection of numerous individual parts. The Collins lab characterized a large number of variable strength promoters but only for maximum and minimum expression levels and never obtained the full transfer curve [20]. Similarly Imperial CSynBI does not obtain a full dynamic range for their characterized parts. Both the BioBricks specification sheet and the BIOFAB project collects data across a full dynamic range, however, the data collected is of the population averages and does not give information about the individual cells or cell-to-cell variability [18] [19]. Elowitz was able to predict an integrated feedback circuit but it is uncertain if this method generalizable to non-integrated circuits without feedback [21]. Prior work by Weiss did not collect single cell information and did not calculate the transfer functions on a per plasmid copy basis and was unable to make quantitative predictions [5].

A recent publication tried to achieve prediction based composition of genetic circuits. Two repressors quantitatively characterized and then connected together to create an 'AND' gate. Quantitative characterization of the 'AND' and 'NOT' allowed predictable composition of these two parts [22]. However, this prediction methodology still relies on parameter estimation and fitting and determining biochemical rate constants. As soon as the context of the system becomes more

complicated it is unclear if their prediction method will hold, where as the prediction method described later in this thesis proposes a method that implicitly accounts for contextual changes by phenotypically characterization and composition of new genetic circuits.

2.2. Cell-cell Communication systems

Coordinated behavior of a population of cells allows for complicated logic functions that are not possible without cell-cell communication. Mammalian cells have numerous mechanisms to relay information about their surrounding environment inside the cell. Many of these mechanisms are specialized and have very particular responses associated with them. In order to implement new communication systems in mammalian cells one would need to introduce completely new orthogonal systems or reappropriate endogenous pathways for new functionality, or some combination of the two.

One of the first systems introduced into mammalian cells used acetaldehyde as the communication signal [23]. Sender cells comprised an alcohol dehydrogenase that converted traces of spiked ethanol into acetaldehyde. Acetaldehyde has a boiling point of 21°C and is therefore in its gaseous form at 37°C in which the experiment was carried out. The signal would diffuse to the receiver cells through the air and turn on expression of a reporter gene. The receiver cells were able to detect the signal through an acetaldehyde-inducible expression system that used AlcR to activate a chimeric promoter in the presence of acetaldehyde[24].

Weber et al. were able to engineer a pathway that did not use endogenous receptors or reporter proteins. Their senders however, used a protein derived from the liver. Another communication system that was purported into mammalian cells was based on the Tobacco Etch Virus (TEV) protease. Barnea et al. knew that β-arrestin gets recruited to GPCR's (G-protein coupled receptors) in order to down regulate a GPCR response after activation. In order to redirect the response of a GPCR, a transcription factor (TF) was fused to the cytoplasmic tail of the GPCR with a TEV cleavage site in between. The TEV protease was also fused to the β-arrestin so that when the GPCR is activated the TEV protease will be recruited and release the TF from the membrane to enter the nucleus and activate transcription [9]. Barnea et al. similarly used a receptor tyrosine kinases (RTK) and fused the TEV protease to the SH2 domain protein that gets recruited upon activation of the RTK to release a TF to the nucleus. This is the basis for one of the communication systems I am working on and discuss later in my thesis. In Barnea's system there is cross talk since they are using an endogenous pathway to activate their signal. In my design it is completely orthogonal.

Chen and Weiss et al. had previously engineered a histidine kinase (HKs) or two-component system. A two-component system from plants was engineered to function in yeast creating a novel communication system in eukaryotes [25]. Sairam Subramanian tried to implement this same system in mammalian cells but was unable to get the receivers cells to function[1].

Antunes et al. was able to take a bacterial HK and implement it in plant cells [10]. A bacterial TF PhoB was fused to a VP64 domain and was used in conjunction

with a specifically created minimal promoter with PhoB binding sites upstream. When PhoBVP64 bound the promoter it activated the reporter gene. Some promiscuity was discovered in HK phosphorylation insofar as activation of a plant hormone pathway phosphorylated PhoB. Localization differences of PhoBVP64 were seen upon activation of AHK4(plant receptor that responds to isopentenyladenine (IP) and trans-zeatin). The change in localization was used to create a reporter system that responded to the addition of a small molecule, trans-zeatin [10]. This system seemed attractive to purport to mammalian cells since Sairam Subramanian had already created mammalian cells that could produce IP[1]. The use of the PhoB system could possibly overcome the problems seen in previous attempts at creating a mammalian two-component system (discussed in more detail in chapter 5). This system also adapted to detect other ligands, which would prove useful for extensibility of this design [26].

Efforts at developing an orthogonal communication system in mammalian cells have been met with much resistance. The existing systems have evolved to carry out a specific function and the more that is understood about endogenous pathways, the more likely it is that synthetic biologists can engineer a system to have the capabilities necessary to coordinate population level control.

Chapter 3

3. DNA assembly Technology

Mammalian synthetic biology is less developed than its prokaryotic counterpart, such that there are several bacterial standards (i.e. BioBrick, BglBrick, etc...)[7, 27]for assembling different modular components, however, in mammalian cells there is no standard or easy way to create these modules. This problem affects the speed, versatility, and expandability of mammalian synthetic biology. We sought to develop a standard to allow us to use principles of synthetic biology for the creation of mammalian parts. In bacterial assembly methods, standard cloning techniques would utilize enzymes that were not commonly found in prokaryotic proteins. But often in mammalian systems the constructs are so large and varied, one can usually find one of these restriction sites in the genes of interest. The techniques developed in the Weiss lab can overcome the shortcomings in systems designed without regard for future expansion into mammalian cells.

DNA assembly technology has recently seen a large gain in technical achievement through the advent of such techniques as golden gate and Gibson assembly. Initially, the main technique for creating mammalian genetic parts was still standard restriction enzyme cloning. Through painstaking efforts to build genetic constructs we realized that in order to build large modular mammalian genetic circuits quickly and efficiently, a new DNA assembly platform needed to be created. Two people in my lab, Yinqing Li and Patrick Guye, started to develop a new

system that combines Gateway® technology (Invitrogen; Carlsbad, CA) and Gibson assembly[8].

3.1. Gateway® Cloning

The Gateway® system from Invitrogen relies on homologous recombination. The entry vectors are created by first PCRing the gene or promoter of interest and adding unique flanking regions called attB sites. These PCR products can be used for homologous recombination in a BP reaction. The BP clonase enzyme recognizes attB and attP sites and performs the recombination event to create an entry vector (Figure 1). Once entry vectors have been created and sequenced it is extremely unlikely for any mutations to occur to the gene parts because all the following steps rely on homologous recombination; allowing for quick and cheap validation of successful construction of expression vectors using restriction mapping.

Once the library of entry vectors is created the next step is to create an expression vector. An expression vector is a combination of two or more entry vectors and a mammalian expression backbone or destination vector. A similar protocol to entry vector creation is used when making expression vectors. The difference is that numerous parts will be combined together to create a functional expression unit. In one reaction a destination vector and up to four entry vectors can be combined to create an expression vector (Figure 1a,b).

3.1.1. Gateway® Entry vectors

The Gateway® system from Invitrogen (Carlsbad, CA) was used to create the library of parts. This system relies on homologous recombination of sites that vary from 20-250bp in length making it extremely unlikely to find these sequences in any gene used. This system has 3 levels of constructs used to make the final genetic module. The first level is the donor vector, the backbone used for the library of parts (promoters and genes). There is a donor vector for each element (i.e. promoter, gene, 3'UTR). Once a homologous recombination event is performed *in vitro* between a donor vector and either a promoter, gene, or 3' UTR (which normally was PCR'd), a entry vector is created. These entry vectors make the library of basic parts used in combination with each other to create functional parts (The "entry clones" level in Figure 1).

This system has 3 libraries of parts, one is a library of promoters, the second, a library of genes, and the third is a library of 3' un-translated regions (UTR's) such as micro RNAs or their binding sequence and poly adenylation (polyA) sequences. The system was initially designed with only promoters and genes by Li and Guye, and I later expanded it to include a third entry vector library of 3'UTR's.

3.1.2. Gateway® Destination vectors

The other component to the Gateway® system is the destination vectors. These are the final backbones for the Gateway® system and the location in which a promoter-gene pair will end up. The initial destination vectors designed had the

components necessary to create lenti virus particles for integration into mammalian genomes. Once the library of parts is created, one can combine any promoter, gene, and 3'UTR with any destination vector to create a novel expression vector. This is the second level of the 3-part assembly scheme. A schematic of the process can be seen in Figure 1b. This provides the versatility and interchangeability to combine any promoter with any gene, and the speed necessary to test many different combinations of parts. This also allows for the creation of huge libraries of parts with the flexibility to change the destination vector for use in different situations (i.e. instead of lenti-viruses one can use a different method for integration and the only change was creating a new destination vector).

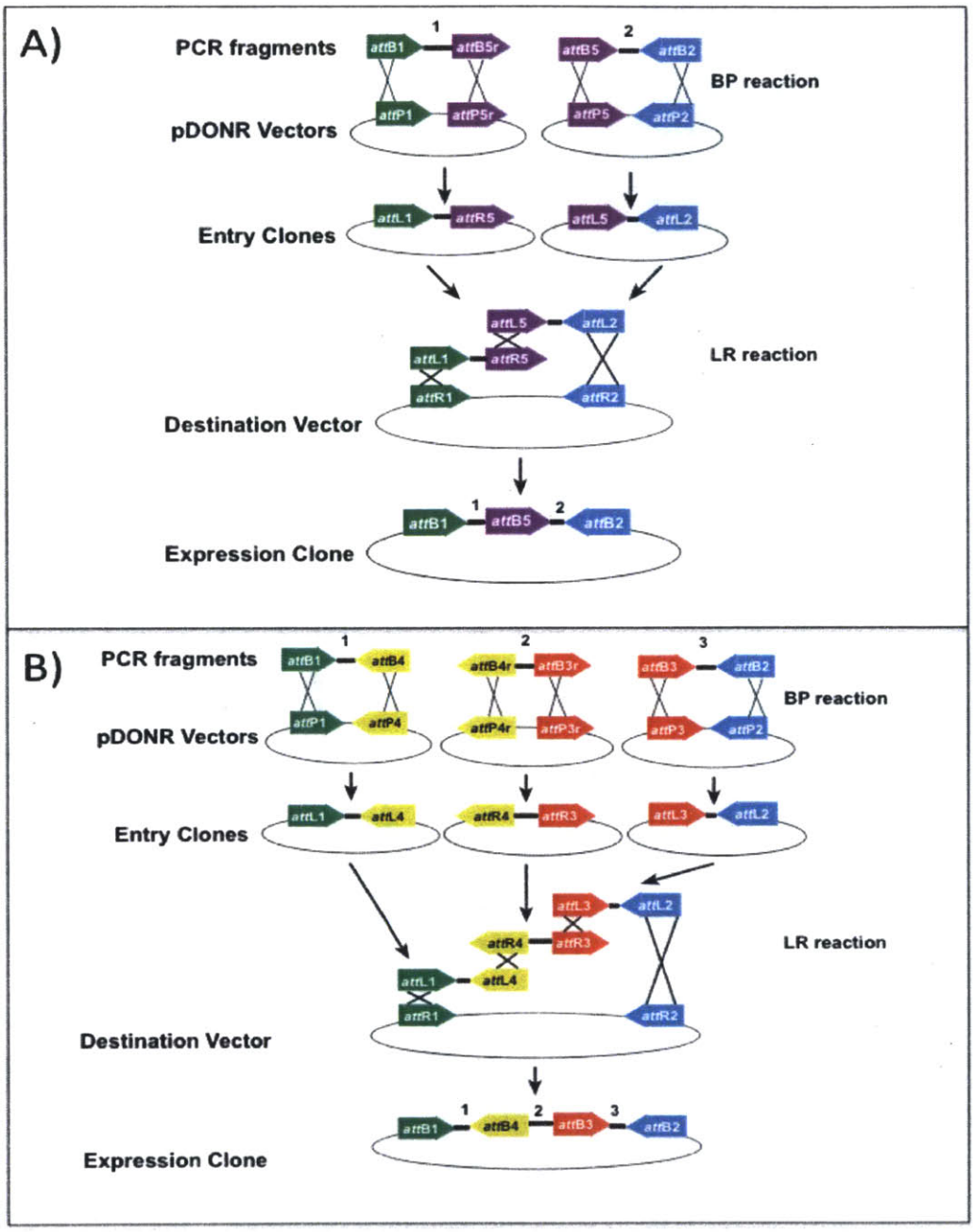


Figure 1 Gateway Assembly
A) The gateway method to create an expression vector starting from PCR fragments of the desired parts. This shows a two-part gateway reaction that utilizes two entry vectors. B) This shows a gateway reaction that involves 3 entry vectors and is called multi-site gateway cloning. Up to a 4-way multisite cloning strategy can be used with decent efficiency. (Figure adapted from Invitrogen; Carlsbad, CA)

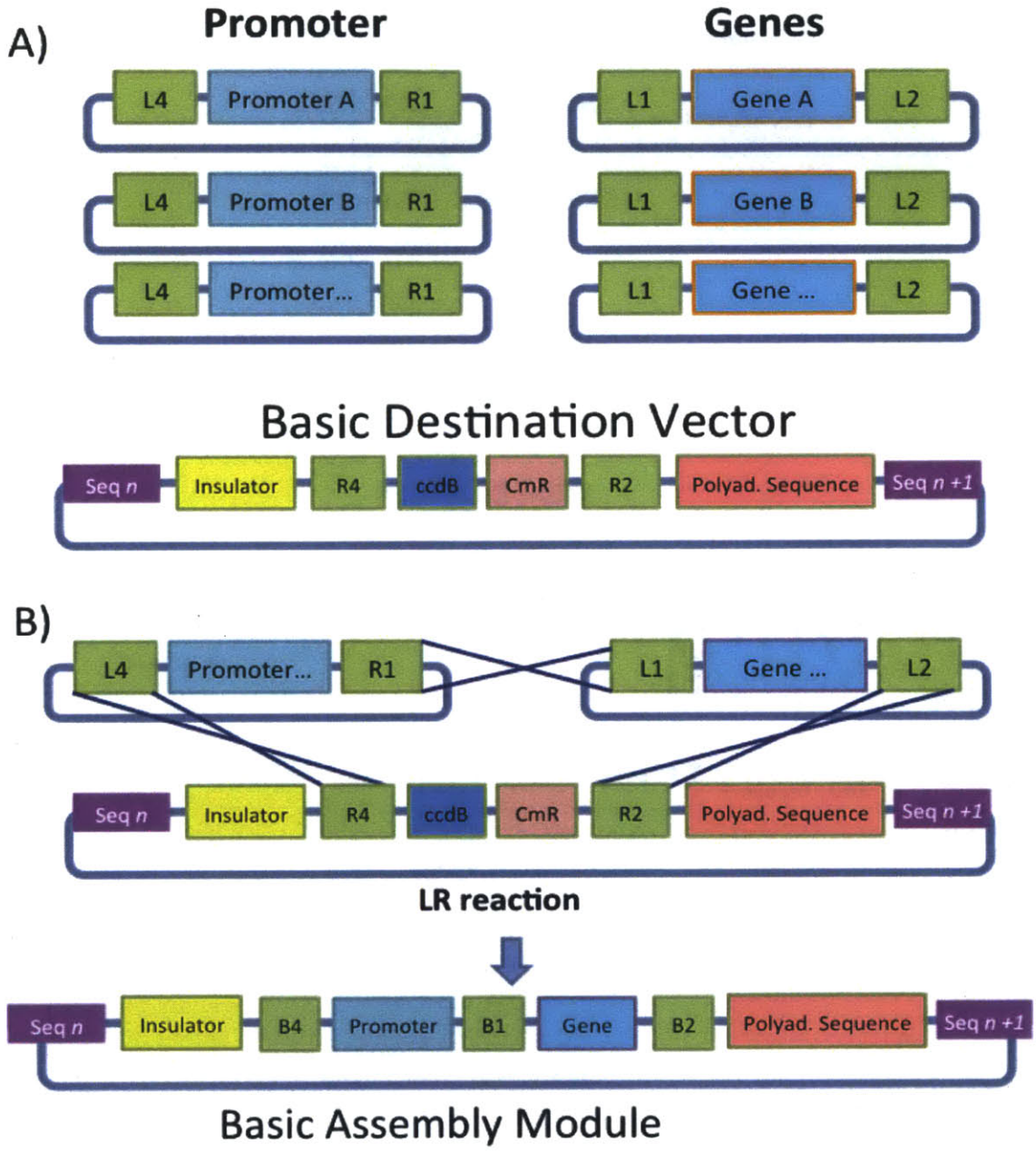


Figure 2 Assembly System Developed in the Weiss Lab[28]
A) The promoter and gene entry vectors along with the destination and important features added to the specific destination vectors used. Also shows which L and R sites used for recombination. Seq n and n+1 represent 40bp overhangs used for Gibson assembly at a later stage in assembly. Insulators are used to transcriptionally isolate each expression unit from each other. The ccdb and CmR are used for counter selection after performing an LR reaction and the Poly A. sequence is used for mRNA stability. B) An LR reaction with our entry vectors and the resultant expression vector [Figure courtesy of Patrick Guye and Yinqing Li].

3.1.3. Gibson Assembly

This system provided a lot of versatility and speed although the lenti viral vectors restricted the size of the circuit to be integrated to one promoter-gene pair per virus. In order to create most imagined mammalian circuits, one would need to make several rounds of infection to test one configuration. The amount of time needed per round of infection and selection is approximately 2 weeks. It would take 5-6 weeks to test out a three promoter-gene pair circuit. All the circuits that we are interested in are much larger than 3 promoter-gene pairs and would therefore require several months to a year just to test one configuration. If any part needed to be altered, one would have to start from the beginning. Also, cells that were exposed to multiple rounds of infection had reduced fitness from all the random integrations. Using lenti-viruses as a delivery vector seemed intractable for our purposes.

Integrating DNA permanently into mammalian cells is desirable for long-term experiments as well as to simplify the characterization models for simple circuits. Since lenti viruses seemed impractical for this purpose a new method for integrating large genetic cassettes was developed. For this purpose, we turned to an isothermal single step reaction using enzymatic assembly of multiple pieces of DNA (called from here on as 'Gibson assembly') (Figure 4). The Gibson method allowed for the construction of very large genetic circuits on a single plasmid. The assembly method can be seen in Figure 3. It uses regions of DNA that are 40bp in length that are homologous to only one other identical region on another expression vector. By using these unique DNA regions DNA can be combined in a specific order. The

destination vector used to create expression vectors are unique in that they contain two 'Gibson sites'. Each Gibson site on a destination vector matches one other Gibson site on another destination vector. This creates an ordered assembly of the expression vectors (Figure 3ab)

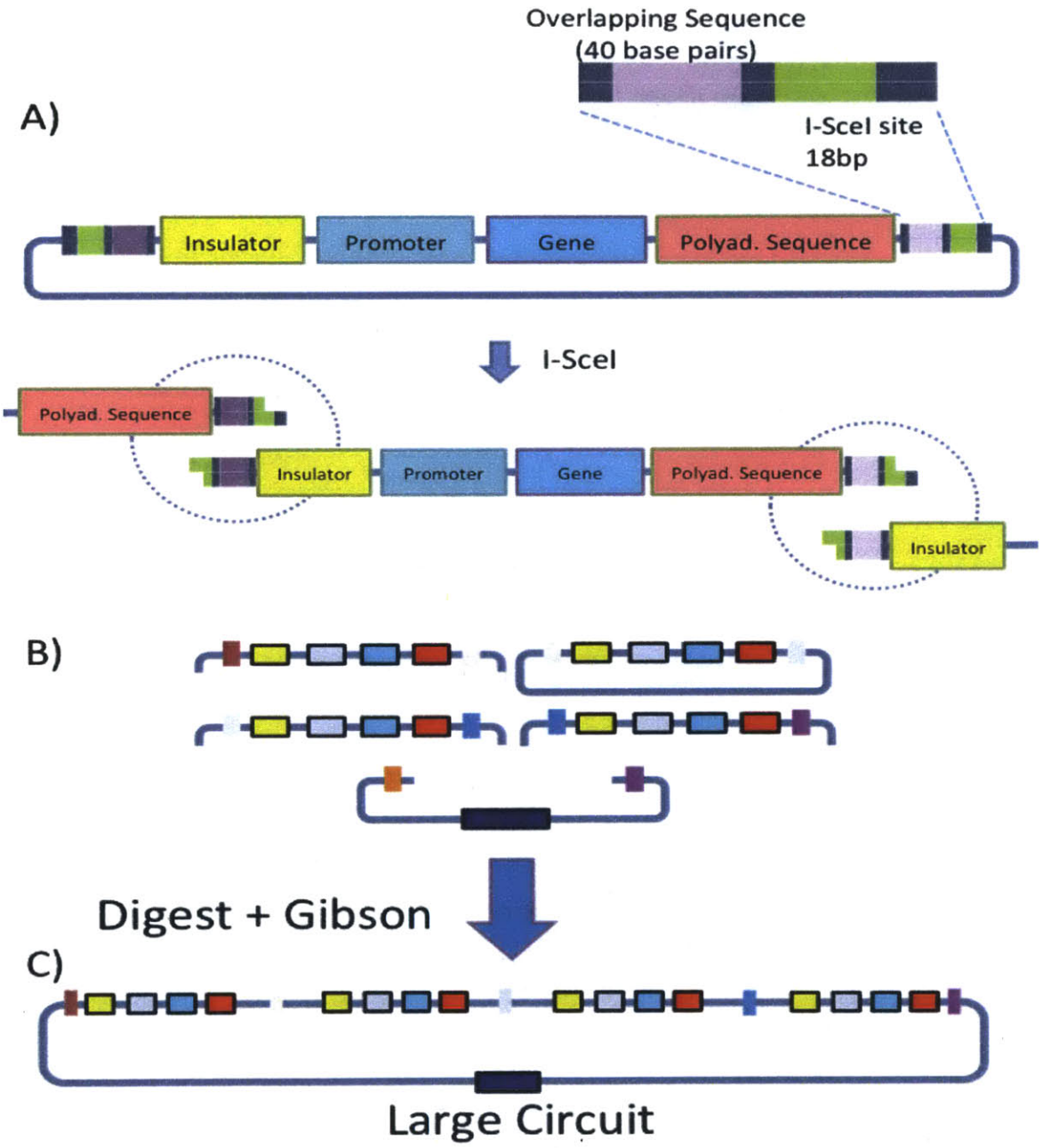


Figure 3 Gibson Assembly Developed in Weiss Lab
A) Depicts a standard expression module containing Gibson specific 40bp overhangs. Step 1 is to digest with I-SceI exposing the overhangs so that congruent overhangs can be sewn together. B) A schematic representation of the one pot reaction where numerous expression cassettes are combined into one carrier vector to create a large circuit. C) The resultant large circuit contained on a single plasmid [Figure courtesy of Patrick Guye and Yinqing Li]

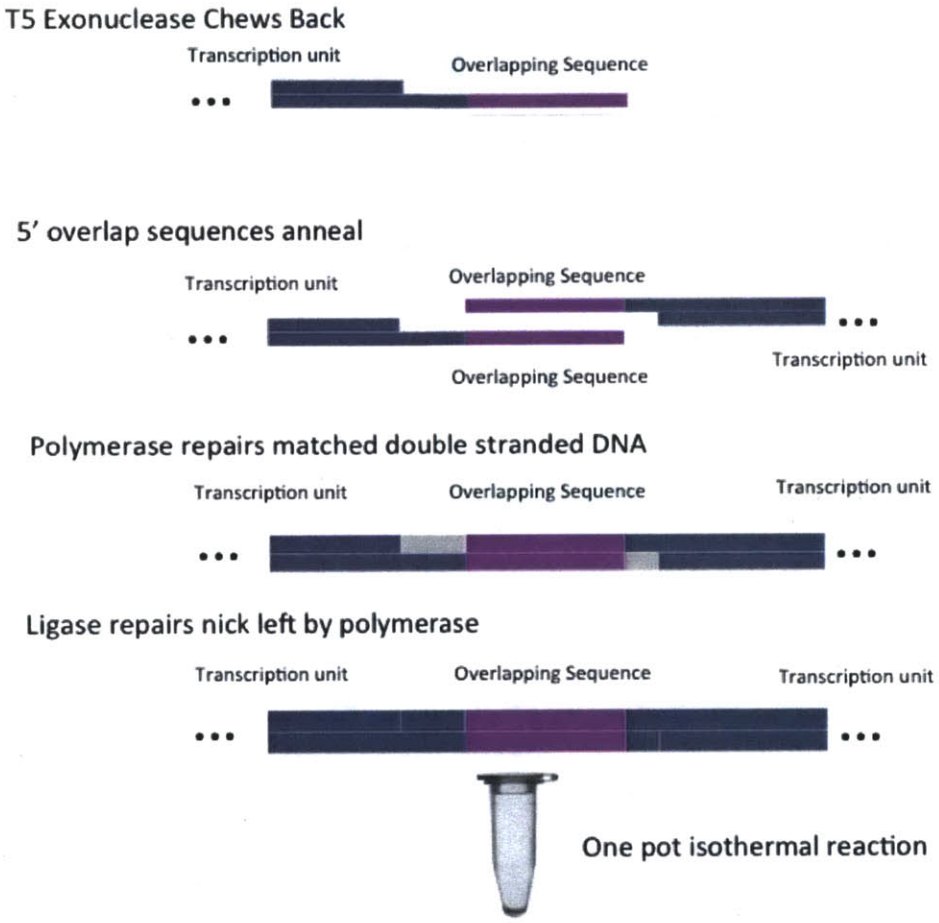


Figure 4 Gibson Assembly Protocol
The first step in the gibson reaction is to expose an end of DNA such that a T5 exonuclease can act. This is done in our lab through restriction digestion, but if one starts with a PCR product the end is already exposed. Once the T5 has exposed the homologous regions to be connected, they anneal to one another and the polymerase and ligase fill in and repair the missing DNA. This occurs in a one pot isothermal reaction Figure adapted from [8].

3.2. Discussion

The new assembly method based on Gateway and Gibson protocols allows for the modular assembly of very large circuits, containing up to 14 promoter gene pairs at the present (>60kbp). A hierarchical method of Gibson assembly was also created and was shown to assemble much larger circuits. To use the hierarchical method of assembly, one must first create a circuit using a new carrier vector that contains additional Gibson sites that can be exposed after its construction. The large (but still incomplete) circuit can then be digested with I-SceI like the other expression units and combined in the same fasion as the normal assembly method. The difference in this case is that one "expression unit" contains several other expression units. A schematic of the hierarchical assembly process can be seen in Figure 5.

In the development of this this assembly technology I have contributed by adding over 70 vectors to the promoter and gene libraries in a very short time frame compared to traditional methods. I have also constructed and tested over 300 hundred expression vectors that can be used for Gibson assembly or directly for transfection experiments. I have also created a third library for 3'UTRs and a new set of destination vectors that can accommodate three-entry-vector gateway assembly. I have helped in the development and testing of this very helpful DNA assembly technology.

These DNA assembly methods have allowed significant progress to be made towards building and testing various circuits in mammalian cells that would not have been possible otherwise. By using this DNA assembly method, the burden has

shifted from the bench top and creating the physical DNA to the researcher, deciding what DNA he/she wishes to make. By helping develop these tools and libraries of parts I was able to create and test numerous large genetic circuits in matter of weeks instead of months and quite possibly years.

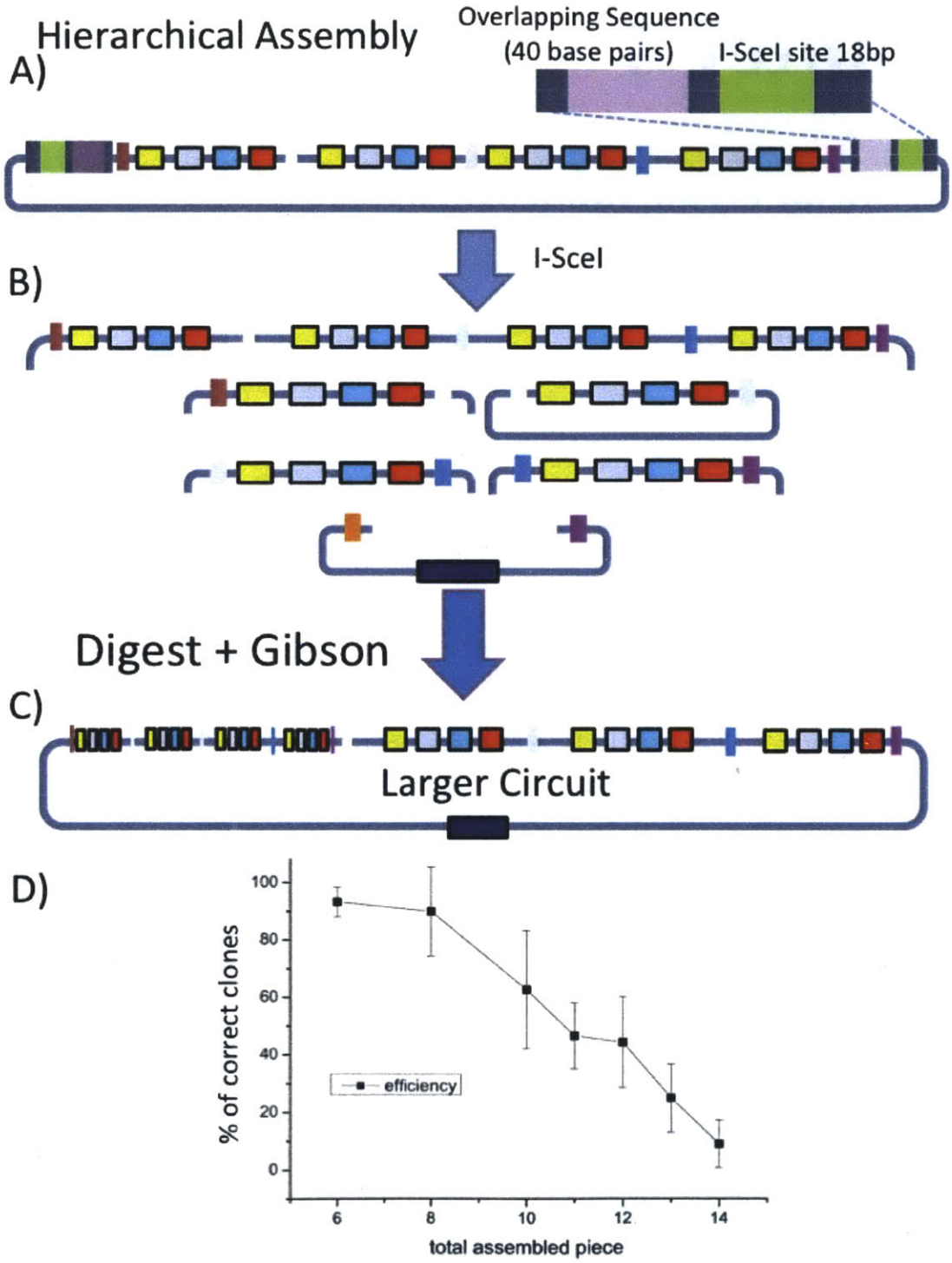


Figure 5 Heirarchical Gibson Assembly
Depicted here is the weiss lab's heirarchical assembly that uses the same Gibson protcol. The only difference is which carrier vector is used. A) This depicts a circuit created using a different carrier vector that contained sites to be used in heirarchical assembly. The Gibson sites surrounding this circuit are used for this purpose B) The large circuit and other smaller expression cassettes can be digested to expose the 40bp overhangs and combined together in a standard Gibson reaction. C) The final product is a very large circuit that has been heirararchical assmebled. D) Thus far 14 pieces have been gibson'd together to form up to ~60kbp circuit. The assembly efficiencies are shown here. [Figure Courtesy of Patrick Guye and Yinqing Li]

Chapter 4

4. Artificial Tissue Homeostasis

The ability to engineer large multicellular systems is an important challenge that face synthetic biologists today. Tissue homeostasis is one example where the ability to engineer such systems would be beneficial. Tissue homeostasis is the balance between growth and death. The growth of new cells can come from cells replicating to form identical copies of itself and from stem cells differentiating into a predefined cell type. Studying the challenges associated with engineering a tissue homeostasis system can give invaluable insight into how endogenous systems perform the same task as well as how to correct natural systems from failure. Mis-regulation of tissue homeostasis plays an important role in Type I Diabetes.

Current standard treatments for Type I Diabetes include the maintenance of insulin levels by blood monitoring, diet, and exogenous insulin injections. More radical treatments include full organ transplants, islet cell transplants or β-cell transplants [29]. Even when patients are lucky enough to be chosen for an allogeneic pancreatic organ transplant, they must take immunosuppressants in order to battle graft vs. host disease [29]. A recent attempt to use islet cell transplant therapy provided short-lived relief in most patients but the transplanted β cells subsequently died or ceased to produce insulin in a majority of the initial successful transplants [30]. Clearly another approach is necessary to alleviate the problems caused by diabetes and address the root causes of the disease.

The best possible treatment would be one in which a person's own adult stem cells are collected, turned into iPS (induce pluripotent Stem) cells, re-programmed, and reintroduced into the body to relieve the disease state. In this situation, the correct "program" needs to be developed in order to cure the patient. Synthetic biology holds the promise in the development of such a program; engineered artificial tissue homeostasis. This project focuses on engineering an isogenic population of human embryonic stem cells (hESC) or adult (e.g. iPS) stem cells that will have the capability to produce a stable population of insulin producing β cells.

As seen in Figure 6a, the proposed system[101] is very complicated and contains several stand-alone modules that can be worked on independently of the others. By replacing different modules of the artificial tissue homeostasis system one can also relieve other disease states. If instead of pancreatic β cells, one wanted neuronal cells, all one would have to do is change the differentiation module that determines cell fate and this system could serve to restore function to damaged neuronal tissue as a potential cure for Alzheimer's.

In Figure 6a the large system is broken up into modules by the light shading and boxed regions. The modules outlined consist of an intercellular communication system that can relay the number of stem cells in the system (cell-cell communication system I). Another cell-cell communication system (in green) is used to relay the number of β-cells in the system. The differentiation control circuit (top right) contains the logic circuit that decides when to have cells grow/differentiate/quiesce. Once a cell has met certain conditions it will

differentiate. The differentiation module is seen in the bottom right of Figure 6A. Also, it might be necessary to include a safety mechanism in the event that a stem cell re-locates to an undesired location (bottom left).

The differentiation control circuit provides numerous challenges that face mammalian synthetic biologists. By focusing on this module I am able to work on several important questions in synthetic biology; Is it possible to engineer a orthogonal extensible intercellular communication system? Is it possible to predictively compose large genetic circuits based on prior characterization? What characterizations are necessary for the predictable composition of circuits? What is the optimal circuit topology for a robust consistent oscillator in mammalian cells? Below I enumerate the function and progress of each component in the system I am working on.

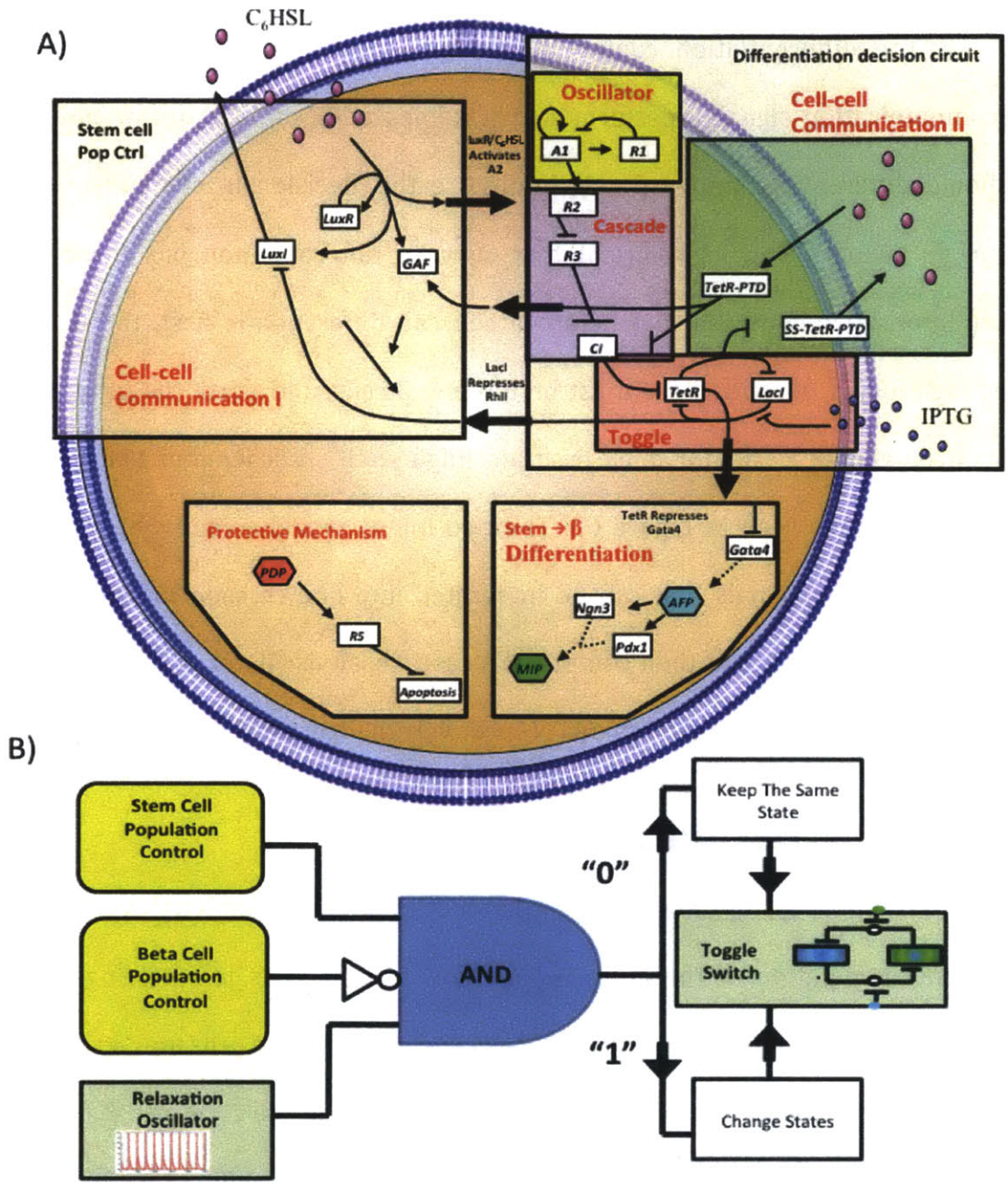


Figure 6 β-Cell Tissue Homeostasis
A) Depicts the entire tissue homeostasis system proposed to cure type I diabetes. There are 4 modules that can be worked on independently and can be interchanged to perform a different task if necessary. The "stem cell population control" is a quorum senseing system to indicate stem cell population size. The differentiation control circuit performs the logic computation used to decide to differentiate or not. The differentiation circuit takes the stem cell to whatever cell type desired with out exogenous cues. The protective mechanism makes sure that the cells do not cause cancer or relocate to another area. B) The logic function being performed in this system takes in three inputs and AND's them together. If all three conditions are met it will flip the toggle switch and cause differentiation to occur. Figure adapted from [101]

4.1. Differentiation control circuit

The differentiation control circuit is designed to take in several inputs (communication signals), perform a logic computation on whether or not the cell should differentiate into a β cell, and store this decision in a memory element (toggle switch). The logic circuit that controls differentiation has three specific conditions that must be met before a cell can differentiate, first, there must be enough hESCs, second, there must be too few β cells, and third, the input from an artificial genetic oscillator must be in its "high state". A schematic example of this decision can be seen in Figure 6b. The two inputs would come from intercellular communication signals that can be integrated into one cascade through an 'AND' gate. The control circuit then comprises a cascade, an oscillator and a toggle switch (yellow, blue and red boxed regions respectively in Figure 6a)

4.1.1. Oscillator

The purpose of the genetic oscillator is to provide symmetry breaking of the isogenic population. If the entire population of stem cells is receiving the same information, i.e. there is not enough β cells and enough stem cells, one would expect that the entire population of stem cells would all decide to turn into β cells at the same time. If this scenario were to occur the system would collapse and be a transient non-permanent solution [101]. Nature solves this problem by creating a niche for stem cells, and cells that are not located in the niche, differentiate. The oscillator creates an artificial niche by introducing heterogeneity in an isogenic population. This asynchrony, which arises due to the stochastic variability between

each cell, creates the heterogeneity. In this situation, the endogenous noise in a biological system is being used to increase stability of the overall system.

The proposed oscillator design consists of an activator that activates itself and another activator, which in turn activates a repressor. The repressor represses the expression of the first activator (Figure 9a). The principles behind the design of this oscillator come from numerous natural oscillators [102]. It has a positive feedback loop combined with delayed negative feedback (which is increased by having a second activator). After modeling the system using a stochastic Gillespie algorithm, certain design features became apparent for stable oscillations. First, repression needed to be dominant over self-activation; otherwise the system will never be in the "low state" (i.e. when activator one is off) [101]. Second, the dynamics of the repressor needed to be much slower than that of the activators (i.e. the half life of the repressor needed to be longer than the activator) [101]. This can be accomplished by fusing decay tags to the ends of the transcription factors to vary their stability. Clonetech sells several decay tags destabilization domains (DD) that can control the degree of decay via a small molecule. This way one can test more thoroughly the effect of differing rates of decay on the oscillations.

A problem arose initially when trying to use the DD tag because the mechanism by which the DD works is to recruit the proteasome to the protein that contains the tag and have it degraded. However, the proteins that were tagged with the DD domain were localized to the nucleus because they are TF with an NLS domain. The proteasome did not have access to these TF's while they were in the nucleus and therefore had no effect on their activity or stability. Figure 7a shows

that shield (the small molecule that reverses the effect of the DD) had no effect on the downstream expression of EYFP caused by DD-rtTa (the corresponding circuit is below the graph). This problem was circumvented by attaching an NES (nuclear export signal) to the C-terminus of the TF's. This provided the protein with the ability to shuttle back and forth between the nucleus and the cytoplasm and allowed access to the proteasome for degradation. In Figure 7b shield has a significant effect on expression once the TF has the NES domain. I have already characterized and tested several components for the oscillator including: TRE-LacO1Oid, DD-rtTa3-NES-4xFF4, DDg-LacI-NES-mkate, DD-VP16Gal4-NES.

After validating that all these parts work, I transfected the two-stage oscillator. The circuit diagram can be seen in Figure 8. Three hours after I transfected the cells with the plasmids, I changed the media to Optimem, (Invitrogen; Carlsbad, CA) which does not contain phenol red, and added the inducers. I then took microscope images for the next 48 hours at 15-minute intervals. Figure 8 shows three time points (0hrs, 24hrs, 48hrs) for each of the reporters used in the circuit. The constitutive EBFP2 seems to continually increase over the 48-hour period. The EYFP and mkate, which report on the activator and repressor respectively, increase for 24 hours and then level off (and some cells even seem to decrease). This is highly suggestive that the circuit might oscillate under the correct conditions. By observing that the EYFP and mkate do not follow the same pattern of the EBFP2 we can infer that these two elements of the circuit are interacting with each other and behaving differently than a constitutive reporter (EBFP2).

The future work for the oscillator is to combine all these parts onto a single plasmid via Gibson assembly and integrate the whole circuit into the genome for testing. Integration into the genome is necessary since the system will need to be monitored over many days (more than 2 days) and transfection does not provide the necessary stability of each DNA construct that would be needed for the time course movie. Single cell tracking software will also need to be developed in order to analyze the microscopy images to determine if oscillations are taking place.

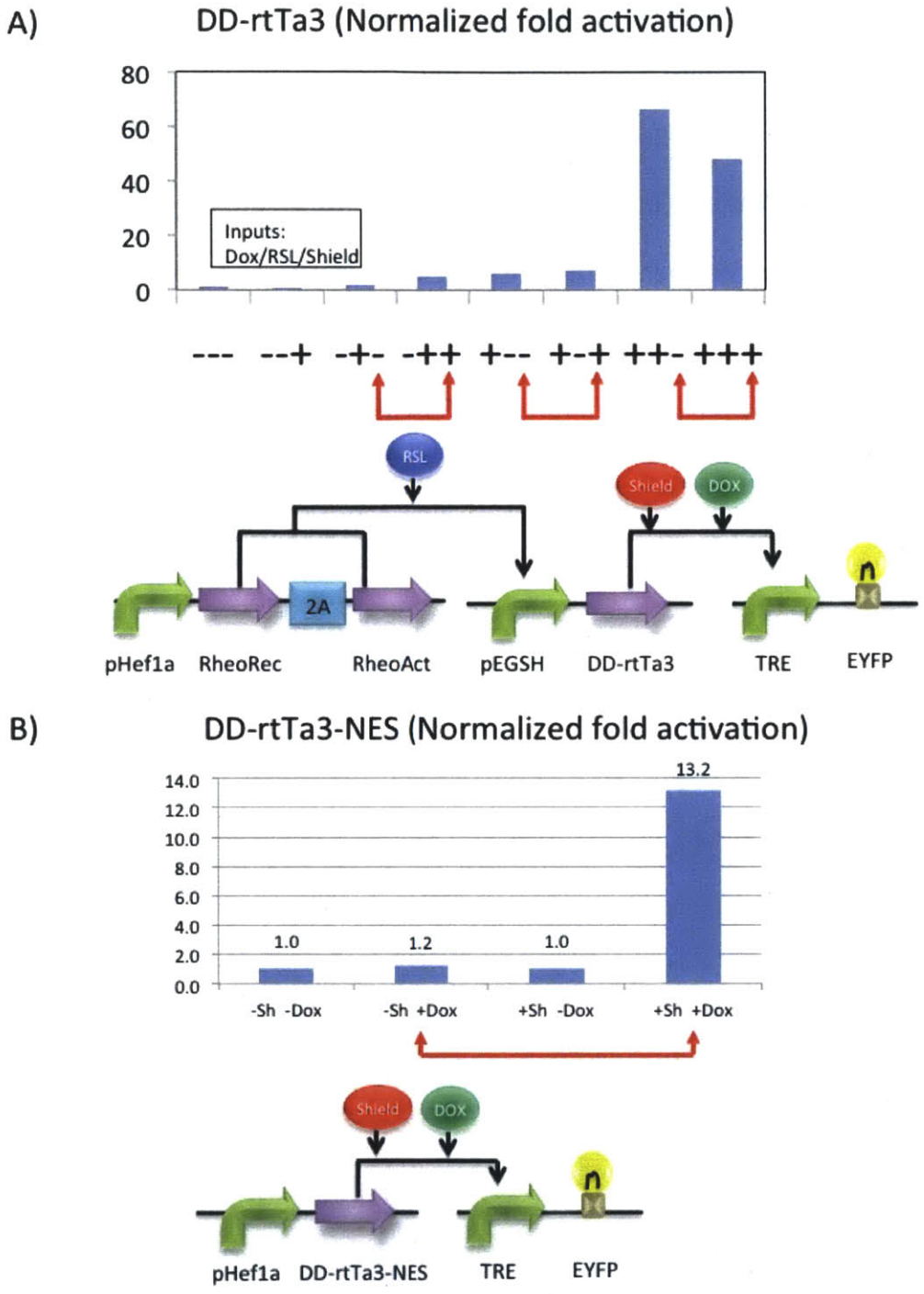


Figure 7 Destabilization Domain and Nuclear Export Signal
A) The graph shows the 8 different combinations of inducers for the system and their outputs. The legend shows the order in which the data is graphed. The red arrows indicate the places in which the only difference is +/- Shield. To determine if the DD domain is working one would compare the pairs indicated by red arrows. Below the graph is a circuit topology for this experiment. B) This graph shows the efficacy of the DD domain once an NES has been fused to a transcription factor. Here there is only two inputs since DD-rtTa3-NES is constitutive. The red arrow again indicates the pair in which only shield is different.

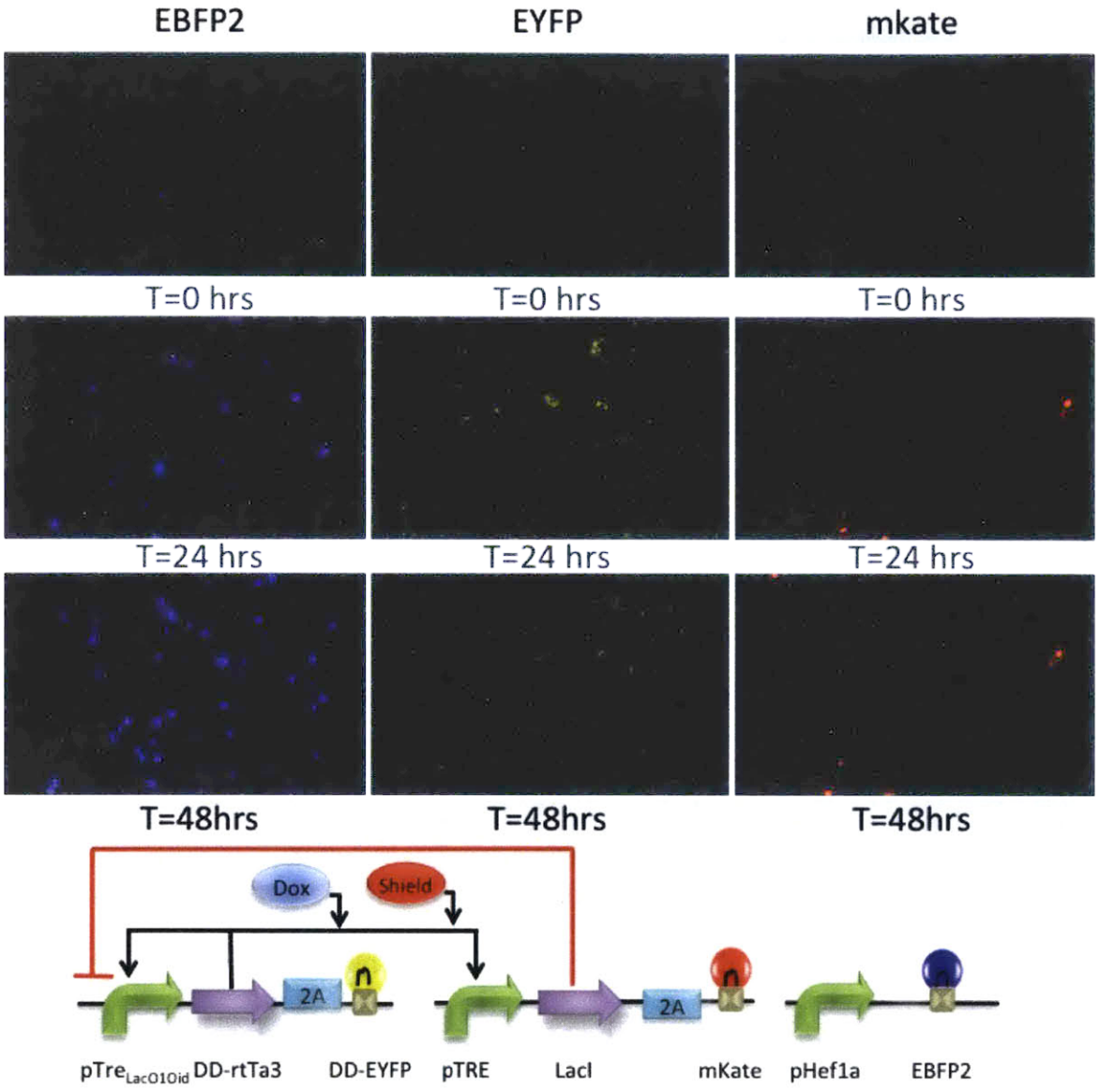


Figure 8 Oscillator Movie
This Figure shows a time course movie of a transfection experiment of the above circuit. The cells were cultured in Dox. at a concentration of 200nM and Shield at a concentration of 0.1nM for 2 days. Images were taken every 15 minutes. The EBFP2 appears to increase in brightness from 24-48 where as the EYFP and mkate appear to level off and decrease a little. This is highly suggestive that the cells might be oscillating.

4.1.2. Toggle Switch

The design of the toggle switch is a simple cross-repressional system in which one repressor inhibits the production of the other, similar to Collins' toggle switch in *E. coli* [3]. Previously, Sairam had constructed and tested a toggle switch [1], which was composed of TetR and LacI both fused to KRAB (kruppel-associated box) repression domains. This system worked in human embryonic kidney cells (HEK293FT) but did not work in murine ES cells (mES) because of endogenous gata factors which bound to a part of the Tet operator site inevitably causing the system to always switch into a state in which TetRKRAB was high [31, 32]. Also the dynamics of the system were not ideal because the switching time needed to change from one state to the other was 6 days. The hypothesis for the slow switching time is that the KRAB domain caused methylation and deacetylation of the histones and the cell is slow to recover from these epigenetic changes.

To improve the shortcomings of the previous toggle switch a new toggle switch based on the same design principles of strong but equal cross-repression is being built (Figure 9b). But, if we want to use any two repressors, which may not have small molecule inducers, then we need a method for switching from one state to the other. The approach used to solve this issue is depicted in Figure 9c. Two inducible activators that each express one of the repressors involved in the memory element are placed before the toggle in order to switch the states. To switch the state of the system one would add the corresponding small molecule to activate the desired arm. Instead of repression domains, many different repressors are being characterized to determine which ones would be best suited to be used together

based on three criteria; high fold repression, increased stability, and fast switching time. Several parts have already been characterized and numerous others are in progress of being characterized. A large library of TALER (TAL effector repressor) proteins are currently being constructed and will soon be characterized. Several TALER TF's have already been characterized and show promising data for their use in this genetic architecture (data shown in Chapter 5). There is also ongoing work to create and characterize a library of TetR homologs found in bacteria to be used in mammalian cells. One of these homologs has already been tested and appears to show a 60-fold repression change (data shown and discussed in chapter 5).

4.1.3. Transcriptional Cascade

The cascade consists of an inducible promoter that activates the expression of a repressor. This repressor in turn shuts off the expression of another repressor (Figure 9c). This describes a two-stage cascade. In the differentiation control circuit, each level of the cascade is connected to one state of the toggle switch. Activation of the cascade chooses which state of the toggle is going to be expressed and therefore which state the cell is in (Figure6a). The cascade is described in more detail in chapter 5 and is the basis for the design and implementation of predictive composition of genetic components.

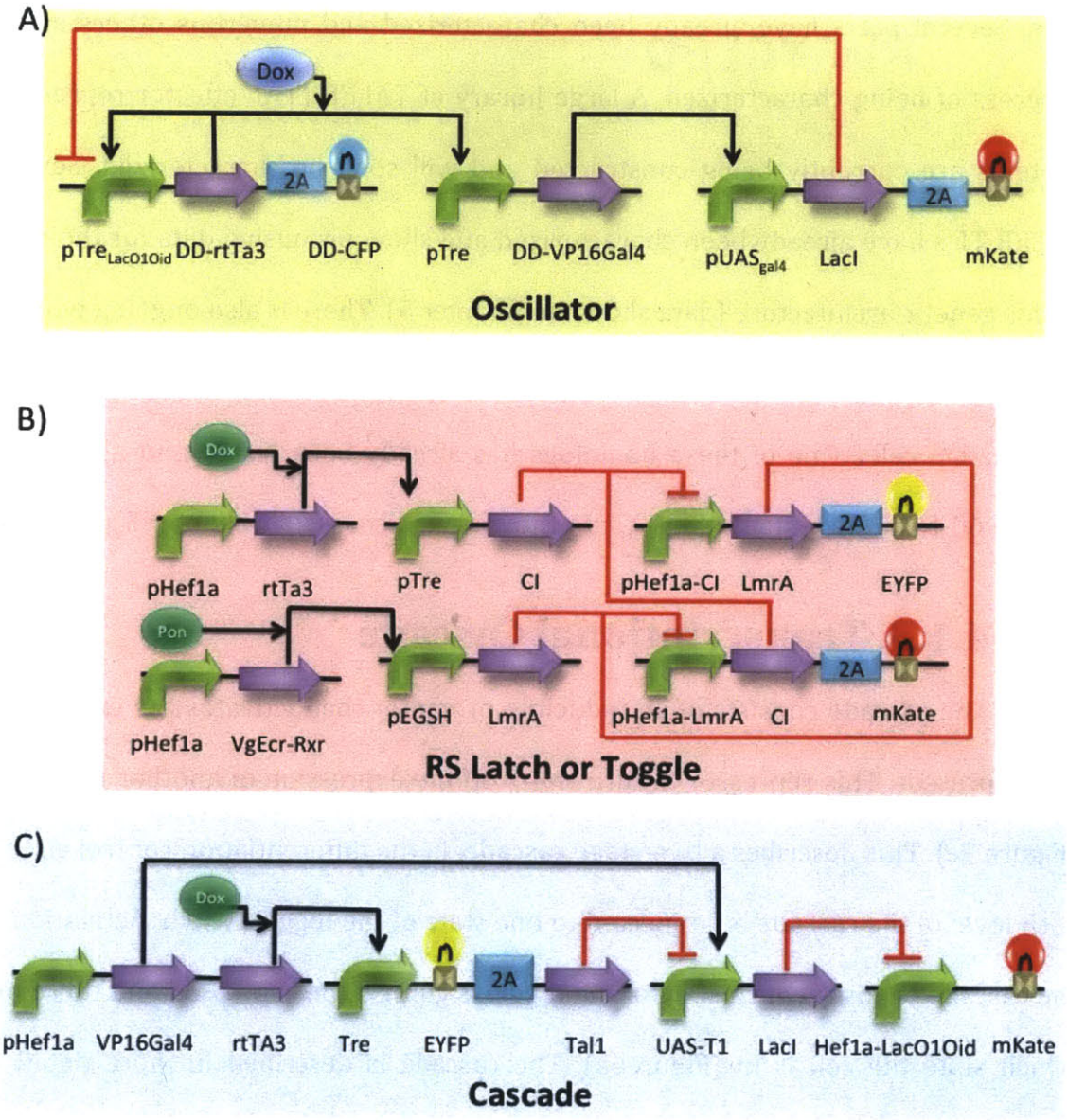


Figure 9 Cascade, Oscillator, Toggle
A) A genetic representation of an oscillator for the β cell project.B) A schematic of a switchable toggle or RS latch memmory element. C) A genetic circuit diagram of a proposed cascade for use in the β cell project.

4.2. Discussion

In this chapter, I have presented the overall system design for artificial tissue homeostasis with regards to replacement of β cells in the treatment of Type I diabetes. I have concentrated my work on the differentiation control circuit and presented several designs for the components of this logic circuit. Ten different cascades have been built and verified (explained in more detail in chapter 6). A new cross-repressional toggle switch was presented for use in this particular context to hold the state of the cell. The oscillator has been stochastically modeled and shown to be functional if certain requirements are met [101]. Several genetic parts have been constructed and tested for use in the oscillator. I have transfected and tested a version of the oscillator (Figure 8) and it is highly suggestive that oscillations could occur under the right conditions. The oscillator is currently being integrated into mammalian cells for microscopy verification over a longer time scale. The integration will also hopefully reduce the cell-to-cell variability.

Chapter 5

5. Engineered Mammalian Cell-Cell Communication

Developing tissues and organs through out development are governed by a complex interplay of intracellular and extracellular signals that are integrated by endogenous genetic circuits to produce a functional organism with all the right parts [103,104]. As scientists elucidate different aspects of these processes they are able to recapitulate them in controlled laboratory environments. Through a combination of artificial scaffolds and the addition of exogenous extracellular signals they have been able to recreate certain tissues and organ systems for use in medical therapies. They have even been able to recapitulate the growth of specific tissues in vivo in specially created microenvironments [33].

The ideal situation would be to engineer a population of stem cells with a genetic program to differentiate and pattern themselves into a tissue without the use of artificial scaffolds and exogenous signals. This chapter of my thesis enumerates the different strategies that I have explored to create artificial inter-cellular signaling systems.

5.1.1. AHL Communication

There are many natural cell-cell communication systems in biological systems. Even the relatively simple organisms such as bacteria have developed cell-

cell communication systems termed "quorum sensing (QS) systems" [105]. They are able to communicate the size of their own population and change their behavior accordingly (i.e. creating biofilms) [106]. A requirement of the components in our system for engineered tissue formation is that they are orthogonal with endogenous processes. For the cell-cell communication system this means that we either develop a completely new mechanism or pathway for communication or try to implement a different system from another organism that will have minimal cross talk with our natural system.

From earlier efforts researchers have had success translocating acyl-homoserine-lactone(AHL) signaling system from *Vibrio fisheri* into *E. coli* [34]. *Vibrio fisheri* are a marine organism that live symbiotically with other marine dwelling organisms and produce light upon certain stimuli. The *lux* operon is responsible for producing and sensing the QS molecule 3-oxo-C6-homoserine-lactone (3OC6HSL) as well as responding to different concentrations of 3OC6HSL by producing light under the proper conditions [35, 36]. The way the system works is there is a gene LuxI that is a synthase that produces 3OC6HSL [37, 38]. Another gene, LuxR, responds to this small molecule through a conformational change that allows it to bind the Lux operator and activate transcription [37-40].

Previous efforts were made to try to implement this system in mammalian cells. The system was designed to be a sender-receiver circuit to recapitulate the behavior seen in bacteria. Initially, only *LuxI* and mammalian optimized *LuxR* fused to an activation domain VP16 combined with a minimal promoter with 7 *Lux* operators was constructed and tested. The mammalian and bacterial type II fatty

acid synthesis (FAS) pathways appear to be similar enough such that the precursor metabolites would be available for *LuxI* to produce enough 3OC6HSL for detection with *LuxR* [34]. The initial system design was unable to produce the desired behavior. The receiver or detection component of the system, *LuxR*, was re-engineered to use a hyper-sensitive version fused to a new activation domain that was mammalian optimized and had the addition of a nuclear localization signal (NLS), p65H4LuxRFmNLS (Figure 10a&b) [1, 41, 42]. The receiver worked upon addition of exogenous AHL and was able to detect 100nM AHL concentrations (Figure 10c). Unfortunately, the senders were not able to produce enough AHL to create a functional sender-receiver system. It was theorized that the problem was in the availability of intermediary metabolites from the fatty acid synthesis (FAS) system in mammalian cells for the production of AHL by *LuxI* [43]. The FAS-II system in bacteria is much more inefficient compared to mammalian cells. Mammalian cells use a single FAS gene that encodes a multi-domain enzyme that catalyzes all steps of fatty acid synthesis, but bacteria have individual enzymes to catalyze each part of the reaction (Figure 11)[34, 43]. In a first attempt, pinocytotic vesicles were used to introduce the metabolites needed by LuxI to produce AHL, namely, s-adenosyl methionine (SAM) [44] and hexanoic acid. This again did not yield any detectable concentrations of AHL. Since these metabolites short half-lives' might have been a reason for the lack of production of AHL, the next step was to endogenously increase production of SAM and acyl carrier protein (ACP). Lenti viral constructs that encode the synthase of SAM and the gene for ACP were used but one still added exogenous hexanoic acid to the system [45]. This was unable to produce

levels of AHL that were detectable by "bacterial sensors" that are able to sense biologically relevant levels of AHL. The "bacterial sensor" was *E. coli* that had been transformed with the LuxR gene and the promoter pLux, which controlled expression of EGFP. The bacteria were able to detect as low as 2nM AHL. (Figure 12b). The next course of action was to reconstitute part of the bacterial FAS-II system in mammalian cells in order to create available intermediate metabolites for LuxI to use to make 3OC6HSL. Since there were so many components it was unlikely through co-transfection and/or co-infection that a large enough percentage of cells would contain the required 9 genes.

In order to solve this issue, new constructs were created using small 2A domains from different viruses that enable multi cistronic expression of genes from a single promoter [46]. This would allow for the use of 3 constructs instead of 9. However, this also did not yield any positive results. The literature had one possible explanation; lactone rings are degraded quite rapidly in mammalian cells and cell culture media [47]. The half-life of AHL in the normal media was only 2.5 hours (Figure 13). Without knowing the reason for the lack of production of AHL as well as AHL's short half-life in a mammalian context future work would be needed to employ this system. The numerous open questions led to the exploration of different types of communication systems that might be better suited in mammalian cells. The first system re-engineered was one that had already been successful in transmitting messages from one mammalian cell to another, viruses!

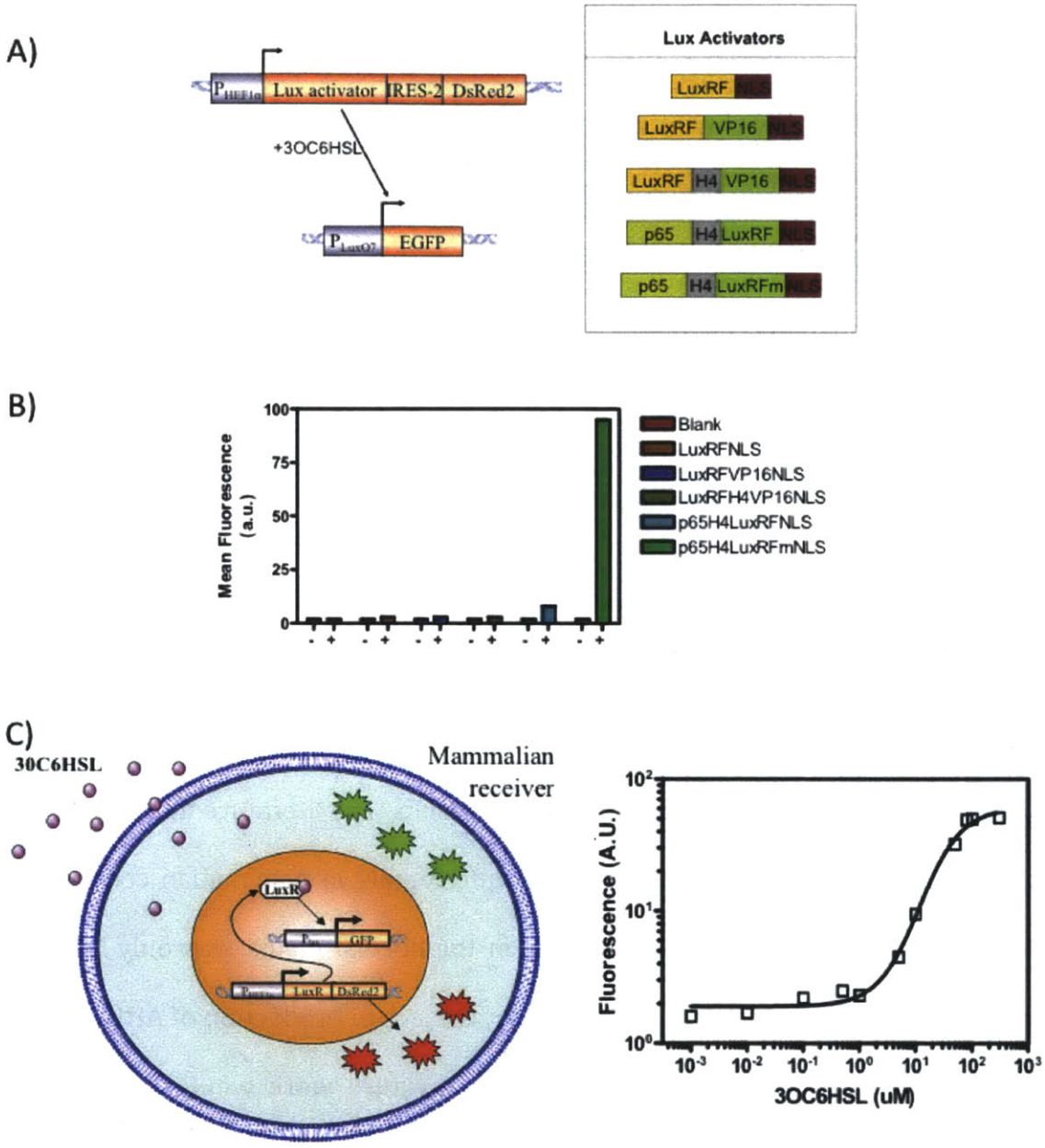


Figure 10 Mammalian LuxR
A)The genetic representation of the constructs tested [1]. B) the results of testing the response of each new LuxR variant with exogenous AHL of 100uM [1]. C) The mammalian receiver desing of the entire system tested. And the dosage response of the best LuxR construct from B. [1] (This figure is adapted from [1] for explanatory purposes)

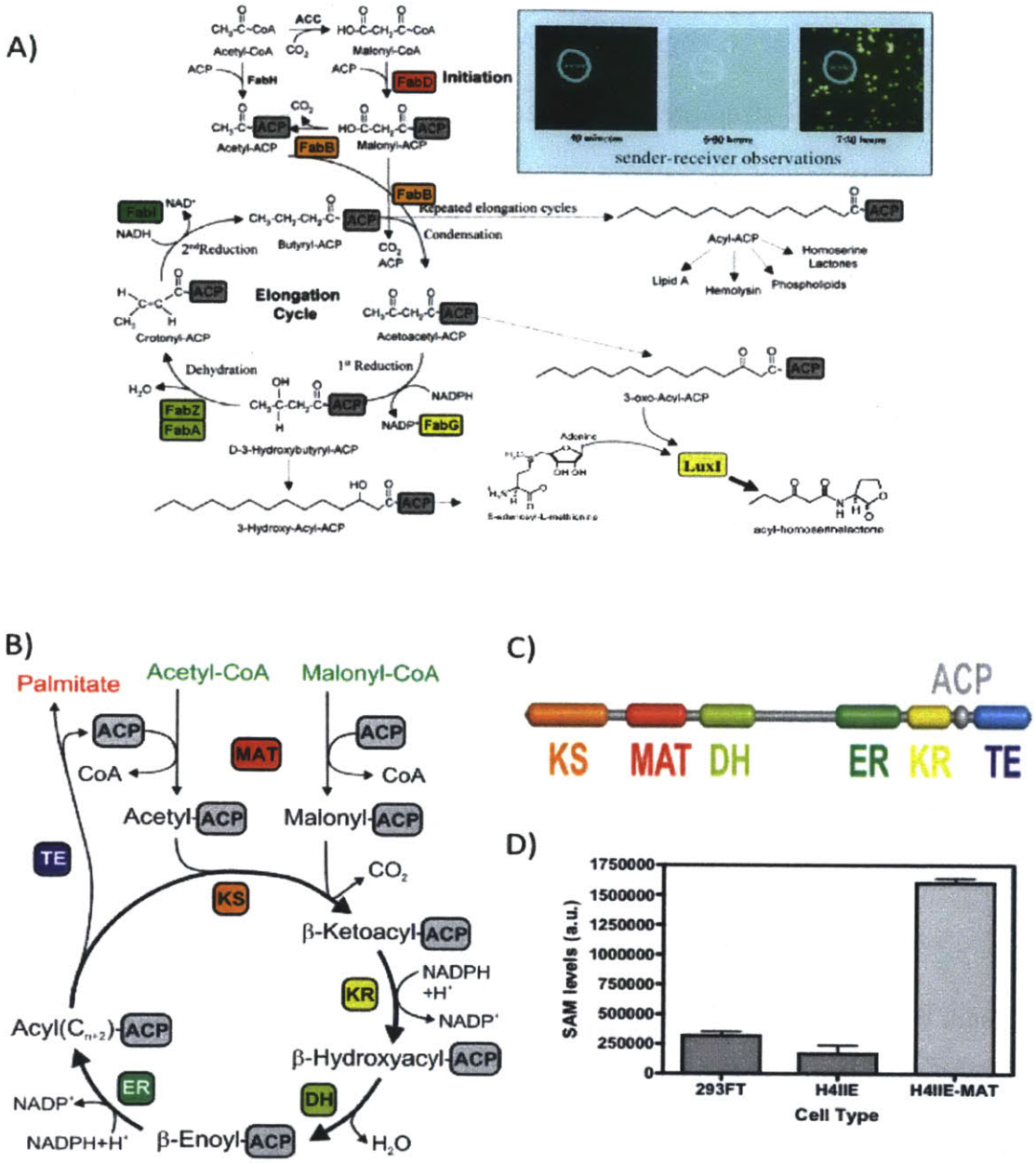


Figure 11 Bacterial and mammalian FAS II pathways
A) Bacterial FAS II system. The inset shows a time course for the bacterial sender receiver system for AHL communication [1, 34]. B) Mammalian FAS II system [43] C) The multi domain mammalian protein that carrys out the function of the mammalian FAS system. D) Mass spectrometry levels of SAM produced upon addition of MAT genes to increase pre cursor metabolites for production of AHL [1]. (This figure is adapted from [1] for explanatory purposes)

5.1.2. Virus Based Communication

After attempting to engineer the AHL communication system from bacteria in mammalian cells, I attempted to implement a system based on virus production and infection. Viruses have already developed a very efficient way of delivering information from one cell to another. Lenti virus particle production is a well understood method having been used for numerous years [107]. This seemed like an ideal system for "proof-of-concept" for this new communication system.

A TF was fused to an endogenous protein of HIV that gets incorporated into the inactive HIV particle. A protein involved in infectivity of HIV-1 called *Nef* is incorporated into the HIV viral particle on the order of 10 nef molecules per virion. Recently a lab discovered a mutant of nef called nef7. A variant was found to have over a 100-fold increase of incorporation into the virus and had fused EGFP and thymidine kinase (TK) to nef7 and shown transport from sender cells to receiver cells. Cell death was shown to be induced by the delivery of nef7-thymidine-kinase viral constructs in the presence of ganciclovir [48]. This system seemed attractive to be used to transport a transcription factor, such as reverse tetracycline trans activator (rtTa), instead of EGFP or TK.

The first step was to recreate the experiment done by Peretti et al. A Nef7-EGFP fusion was created and transfected into mammalian cells. Images were taken with the Leica confocal microscope and they had the ability to visualize the individual viral particles containing EGFP adsorbed on the surface of cells (Figure 14a).

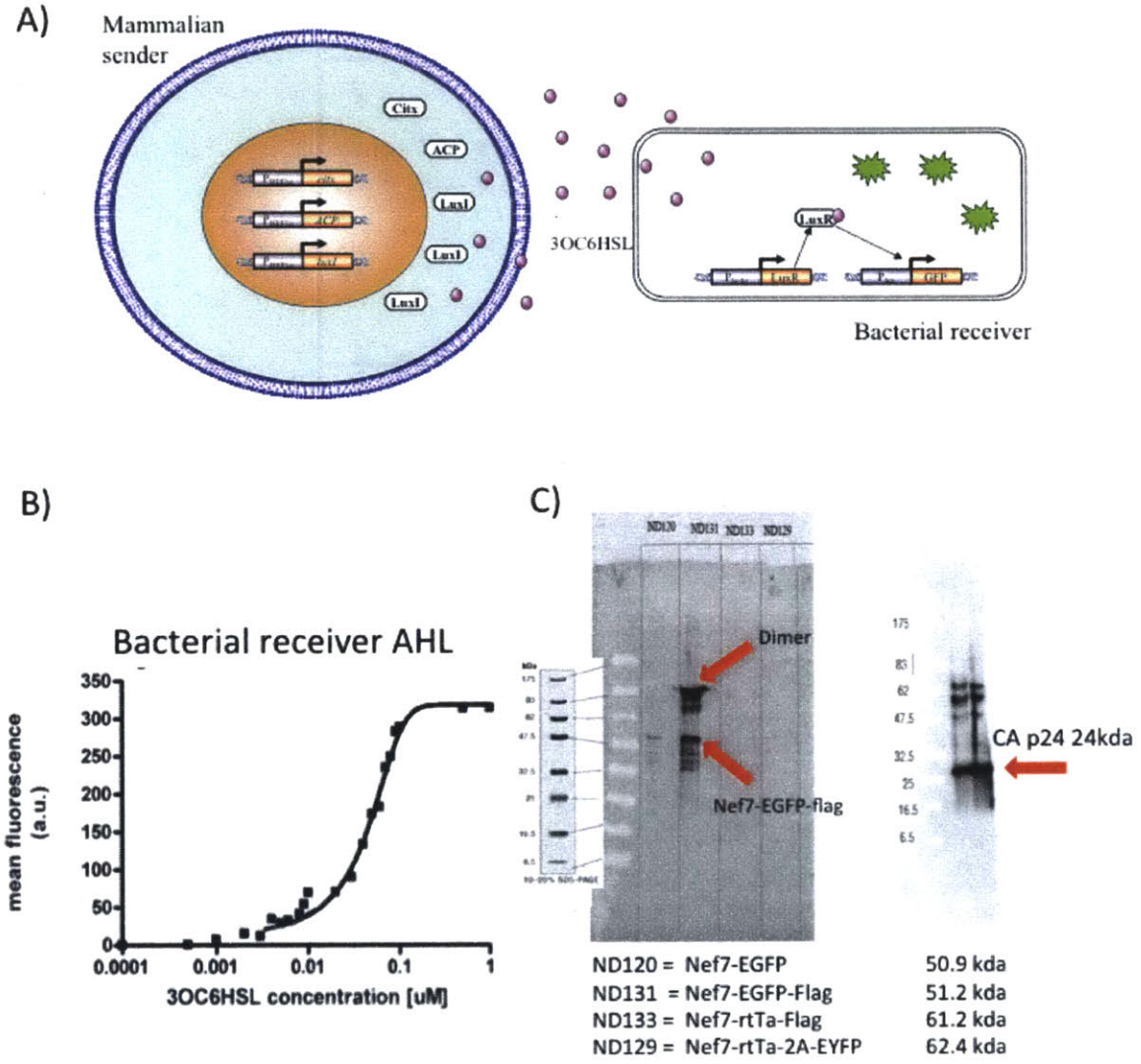
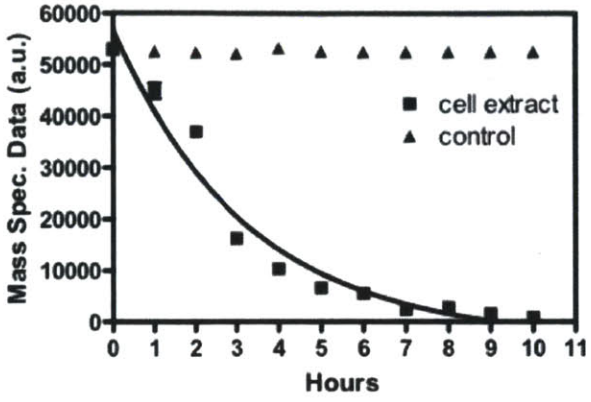


Figure 12 Mammalian AHL Senders
A) Schematic design of the mammalian AHL sender. It has multiple genes responsible for producing AHL. To test the sender separate from the mammalian receiver to debug the system a bacterial receiver was used and grown in mammalian media. B) The dosage response of the bacterial receiver from exogenous AHL. C) A western Blot for Nef7-XX constucts where XX represents Nef7's cargo. Only Nef7-EGFP was seen in the wester blot. Also a control for the Capsid protein of HIV on the right.

A)



B)

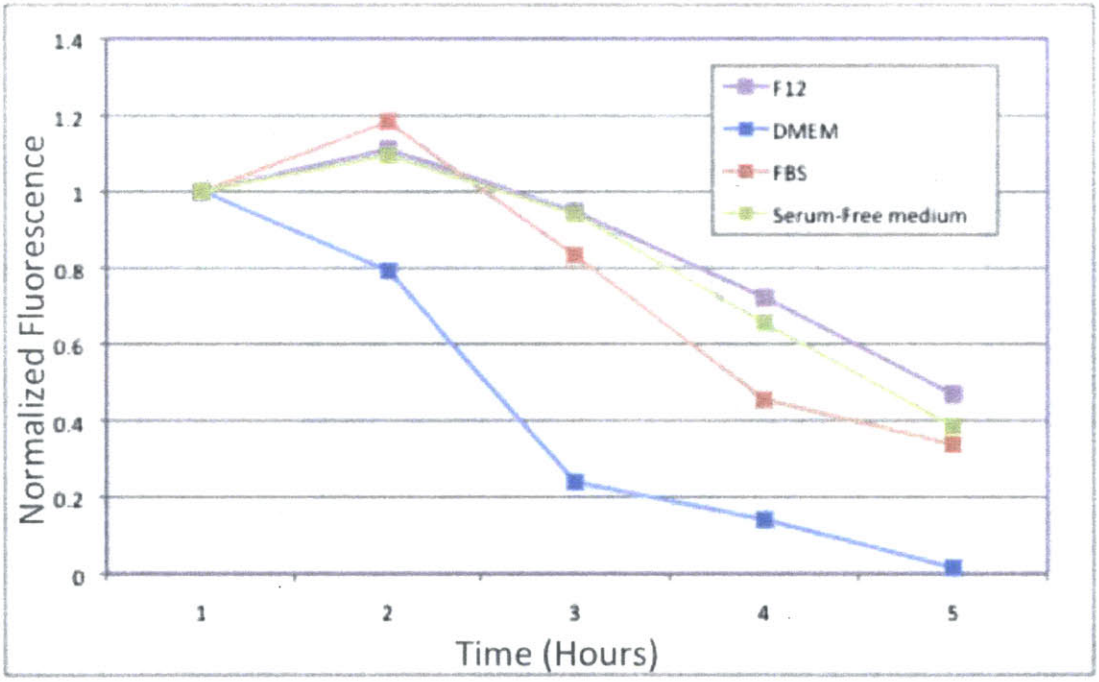


Figure 13 Decay of AHL in Mammalian Cells
A) Mass spec results every hour of AHL in mammalian cells [1]. B) Bacterial response in mammalian media incubated prior to culturing the receivers at room temperature for the number of hours indicated.

After several attempts to create viral particles that showed an effect on receiver cells (HEK 293 cells that were infected with TRE-EGFP), it seemed rtTa was unable to make it into the nucleus of receiver cells to activate transcription of TRE-EGFP. There were several possible points of failure: 1) Nef7-rtTa fusion did not yield a protein that was successfully incorporated into viral particles. 2) If Nef7-rtTa was being incorporated then it was not making it to the nucleus, possibly because of the myristoylation sequence attached to the N terminus of Nef7 (the purpose of this myristoylation tag is for incorporation into the virion). 3) There is not enough rtTa reaching the cell to cause a measurable effect. Verification that fusing Nef7 and rtTa did not destroy the function of rtTa was not possible because nef7 excluded rtTa from the nucleus.

The third hypothesis was easiest to test and was therefore approached first. In a dose dependent manner, more and more viral particles were added to the same number of cells increasing the effective multiplicity of infection (MOI). At an MOI of 10 the cells showed significant signs of cytotoxicity but still no signs of detecting rtTa (data not shown). However, this experiment did provide useful information pertaining to the maximum concentration of virus that can be used without significant cytotoxicity (which is an MOI of 5).

The other two hypotheses were approached simultaneously because both required modification to the Nef7-rtTa construct and were orthogonal problems. To visualize incorporation of Nef7-rtTa into the virion a FLAG-tag peptide was attached to the C-terminus of the construct [49]. To address the issue that rtTa is prevented from entering the nucleus because it is still tethered to the membrane (or vesicle), a

native protease domain [50], used by HIV to cleave the capsid protein from the matrix protein, was inserted between the Nef7-rtTa fusion along with flexible GGGGS linkers [51]. This should allow for cleavage of Nef7 from rtTa once packaged into the virion since the viral protease is inactive until having budded from the host cell [52, 53].

I also constructed a version of rtTa containing N-terminally the amino acid residues that would be present post cleavage and C-terminally a FLAG-tag, in order to check functionality of rtTa with these additions and to use as a positive control for future western blot experiments (PR-rtTa-FLAG) (Figure14c). The newly constructed protein was still functional with these additions as seen through co-transfection with TRE-EGFP into HEK 293 cells (data not shown).

Having these new constructs at my disposal, I began testing each of the aforementioned hypotheses. In order to test if the construct is being included in the lenti viral particles, a western blot was performed on purified virus particles and whole cell lysates that were expressing PR-rtTa-FLAG from a strong constitutive promoter (Hef1a). Both the PR-rtTa-FLAG from the whole cell lysates and the Nef7-PR-rtTa-FLAG constructs were undetectable in the numerous western blots performed (Figure 12c). I turned to the literature to see if anyone else had had success in seeing rtTa in a western blot. The answer was no [54, 55]. I knew rtTa was present because of my functional assay that used TRE-EGFP even though I was unable to see it in a western blot. I had reached an impasse and chose to pursue another system that had fewer complications. There are also numerous other

concerns of *in vivo* use of lenti viral particles as a communication system in mammalian cells, dealing with immune response and prolonged use. [56-58]

5.1.3. TEV Communication

The TEV (tobacco etch virus) system is a receptor based communication system that depends on the dimerization of two membrane bound constructs to induce an intracellular proteolytic event to release a tethered transcription factor, which can then enter the nucleus and mediate expression [9].

The initial design of our system included a fusion protein containing a membrane localization signal and transmembrane domain from the fibroblast growth factor receptor 2 (FGFR2) fused to a TEV protease cleavage site (TCS) and VP16Gal4 (which remained intra-cellular along the cell membrane) as seen in Figure 15a. The extracellular portion of the FGFR2 receptor was replaced by a rigid helical linker sequence that contained the cMyc peptide recognition sequence at its amino terminus. The helical linker was to serve as a spacer to provide distance between the cMyc peptide and the surface of the cell membrane. A similarly constructed part contained the TEV protease instead of the TCS-VP16Gal4 as the intracellular domain. The purpose of the cMyc peptide is to cause dimerization by addition of cMyc antibody between the two adjacent constructs. Upon dimerization a proteolytic event would lead to VP16Gal4 being released into the nucleus. cMyc is easily produced in HEK 293 cells and was re-constituted in the Weiss lab by Adrian Slusarczyk and confirmed by western blot [59-61].

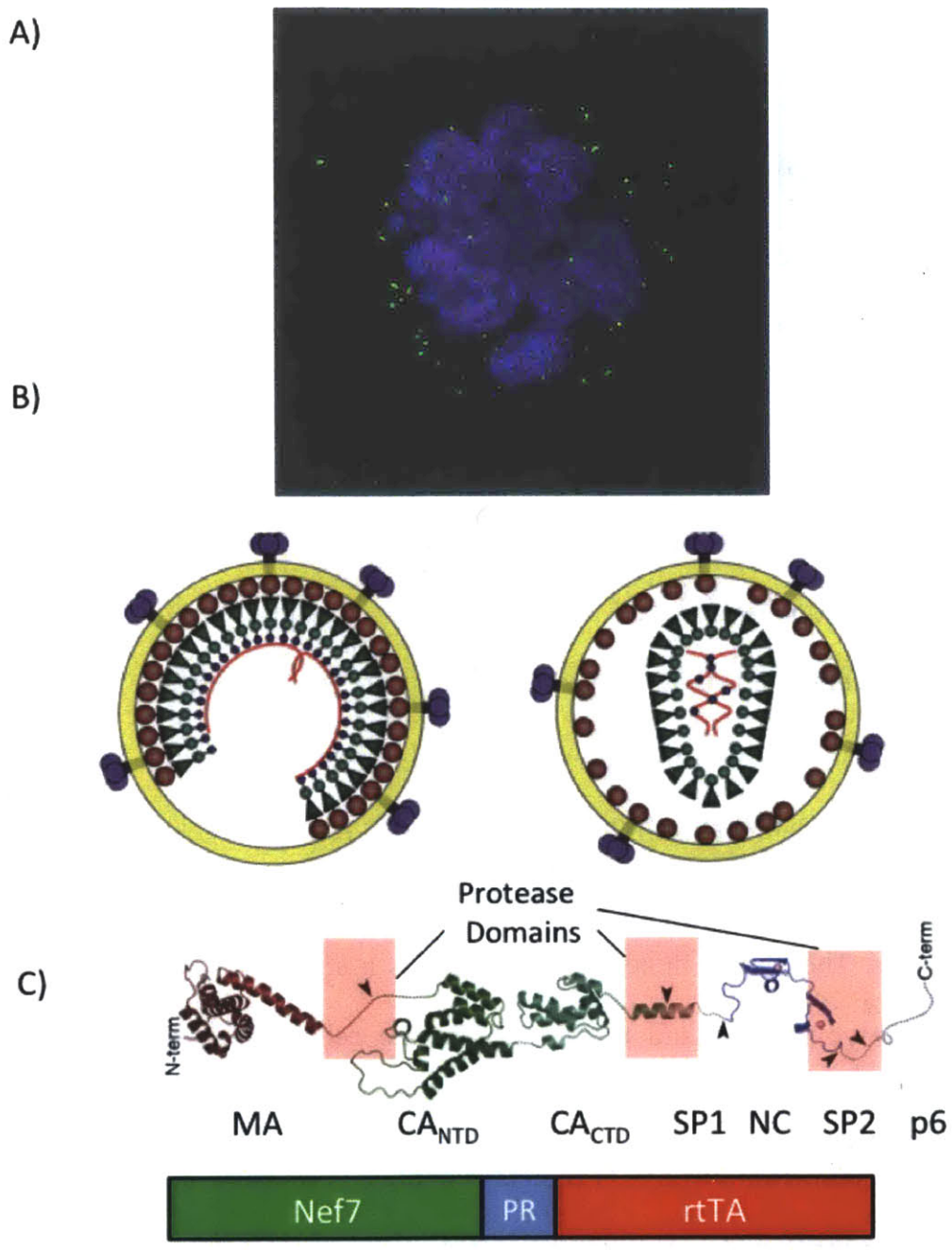


Figure 14 Virus Protein Design
A) A confocal microscopy image of HEK 293 cells stained with DAPI. The Green dots are viral particles containing NEF7-EGFP adhered to the membrane. B) Viral budding and proteolytic cleavage of matrix, capsid and other proteins into a mature virion. C) Structural diagram of the multi protein made by the viral genome with protease domains. Also the design of my NEF7 carrier that would include a protease domain and a transcription factor. (Figure C adapted from [50])

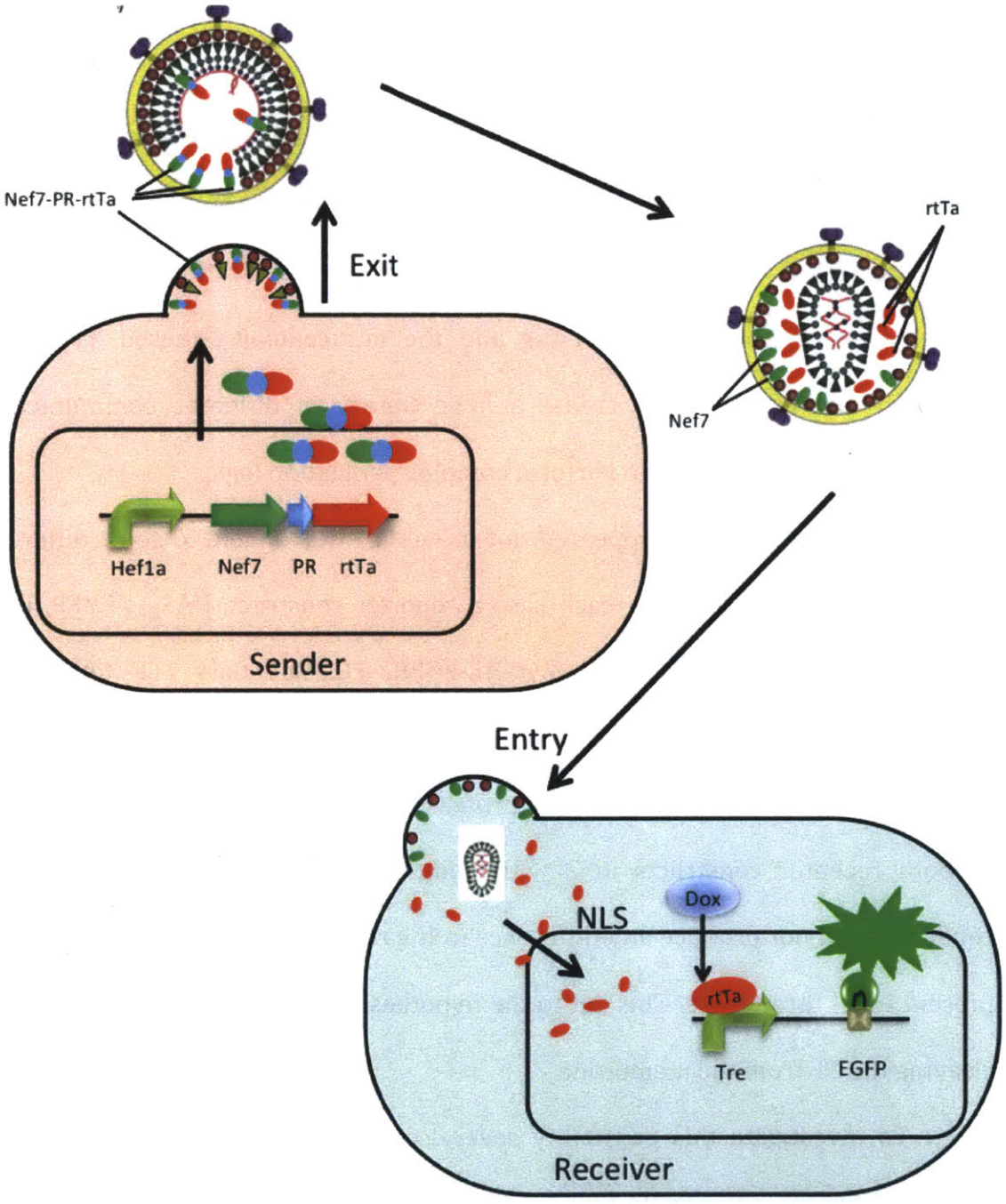


Figure 15 VLP System Design
A schematic diagram of the viral sender-receiver design. The pink cell is the sender that produces the viral particle. After the virion buds off the protease becomes active and produces a mature virion. This process also separates the Nef7 from the transcription factor. The transcription factor is then available to enter the nucleus after infection occurs.

The system was designed to anchor any transcription factor to the membrane and upon dimerization, induced by adding an antibody, the TEV protease would cleave off the TF that would then localize to the nucleus and modulate transcription of the gene of interest (Figure 16). The design's modularity allows for quick expansion of the number of orthogonal communication channels. The extracellular epitope recognition tag and the intracellular tethered TF can be expanded and combined to create a huge library of different communication channels that would be able to perform complex population logic.

Initially, the system appeared not to work. I was unable to see a difference between three conditions (in each case a reporter construct, UAS_{gal4}-EYFP, is co-transfected): 1) the construct, cMyc-HL-FGFR2-TCS-Vp16Gal4 (TF tethered to membrane), alone 2) both receptor constructs, cMyc-HL-FGFR2-TCS-Vp16Gal4 and cMyc-HL-FGFR2-TEVpr (contains TEV protease), without addition of cMyc antibody 3) both receptor constructs in 2 with addition of cMyc antibody. These three conditions did not produce any difference in the reporter (UASgal4-EGFP) that was present in all three cases. This led to the hypothesis that the TEV protease was not cleaving the TF from the membrane.

To investigate this possibility several constructs were built containing an additional fluorescent protein fused to the C-terminus (EBFP2 was added to the TF and mkate was added to the TEV protease) in order to visualize the cleavage and change in localization by microscopy. We were unable to see any change in localization, which was highly suggestive that the TEV protease was not cleaving the VP16Gal4. Another possibility was that not enough of the surface receptors were

being cleaved to visualize the change in localization. We researched the problem and discovered that the TCS being used was in fact the recognition sequence for the TEV protease; however, the recognition sequence was more context dependent than previously thought. Several more surrounding amino acids needed to be purported to our fusion protein in order for the TEV protease to recognize and efficiently cleave our construct.

After adding in the new residues to our fusion protein we were able to see a difference with and without the TEV protease construct (Figure 16d). This shows that the TEV protease construct is cleaving and allowing translocation of the TF from the membrane to the nucleus to activate UASgal4-EYFP. This shows promising results that the TEV system is working within a single cell. The next step would be to see if it is possible to relay a signal from outside the cell inside. Addition of exogenous cMyc antibody should tell us if our system is working and then we can combine the senders and receivers.

5.1.4. PhoB Two Component Signaling

Cells have developed the ability to sense different signals in their environment through a variety of mechanisms. One method involves proteins inside the cell binding specific ligands that affect its ability to regulate transcription and ultimately how the cell responds to particular stimuli. One example is how bacteria respond to different concentrations of glucose in their environment using the lac operon [62-64]. In this particular situation glucose diffuses into the cell and directly affects transcription. Cells have also developed the ability to sense changes in their

environment through receptor-mediated pathways [65-68]. Specialized membrane proteins that usually consist of an extracellular, transmembrane, and intracellular domain are able to sense things in the environment and transmit that information inside the cell [69-73].

This receptor mediated signaling is a major method for transmitting information into the cell. The extracellular domain usually consists of a receptor that binds a specific ligand and either causes a conformational change to the intracellular domain or induces dimerization. These two events usually proceed to cause a phosphorylation cascade that ends in transcriptional regulation. The phosphorylation cascade in bacteria is dominated primarily by histidine and aspartate residues that transfer the phosphate group [74-76]. In mammalian cells, the dominant residues that transfer the phosphate groups are serine, threonine, and tyrosine [77-79]. The histidine-aspartate relays have been seen to be more promiscuous than the serine-threonine-tyrosine ones which may be a reason mammalian cells evolved away from using the histidine-aspartate ones [80-82]. The promiscuity of the His-Asp relays has afforded us the ability to "mix and match" different intracellular components with different receptors to create novel responses from existing pathways.

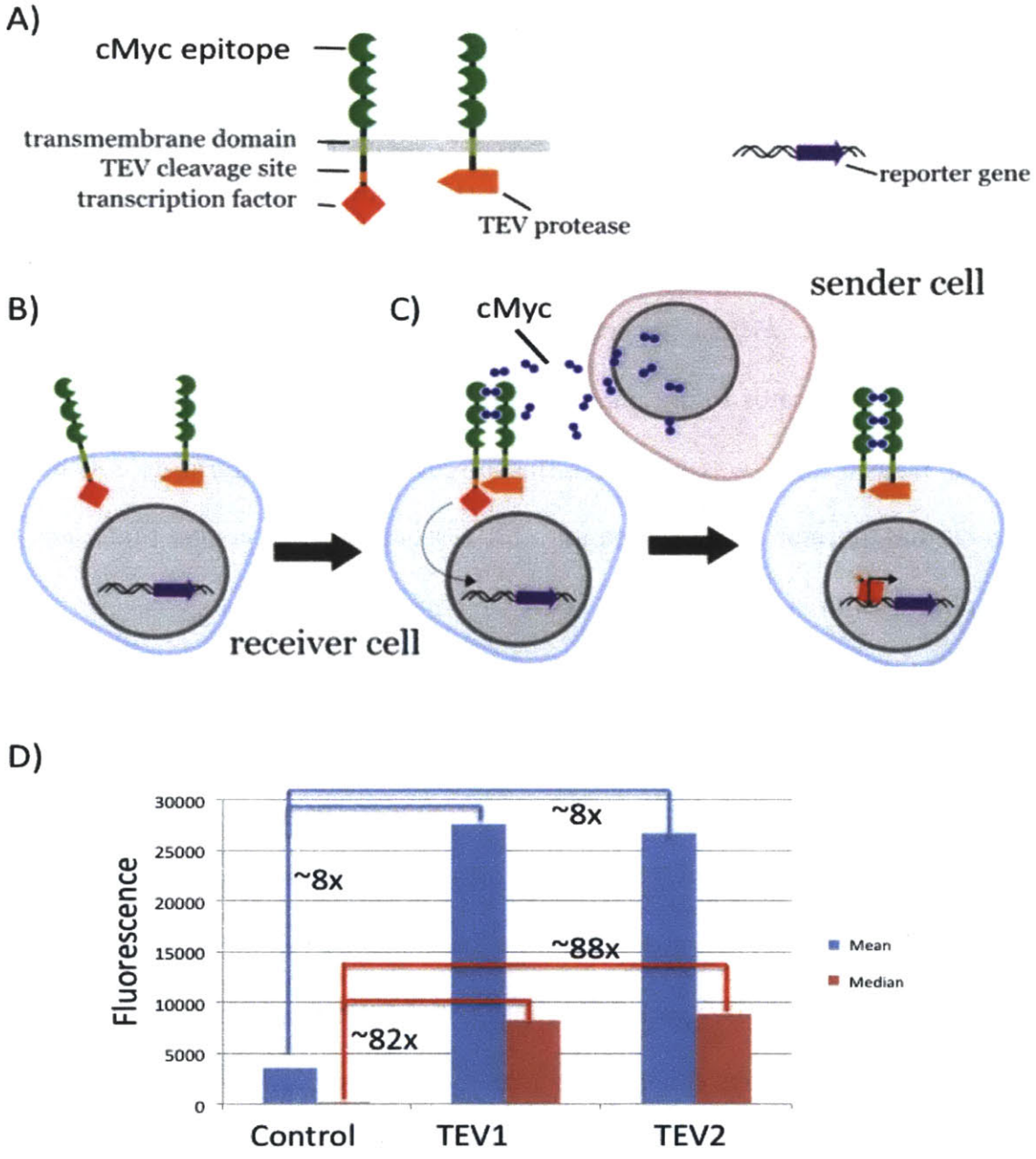


Figure 16 TEV Communication Design
A) Shows the receptor design. Extracellularly there is a cMyc epitope that can be bound by cMyc antibody. Intracellularly there is a TF bound to one receptor with a cleavage site in between and a TEV protease bound to the other receptor. The receptors are designed to hybridize and cause cleavage. B) The receiver cell for the system. C) the sender and receiver system working to express a protein. The TEV protease cleaves off a TF to activate transcription. D) Transfection results in HEK cells with and without the TEV protease construct. There are two versions of the TEV protease 1 and 2.

Recently, a group was able to take a bacterial TF, PhoB [83, 84], which is part of a bacterial two-component system and use it in plant cells [10]. Here they used the AHK4 (Atcre1a) receptor in plants, that responds to the ligand trans-zeatin and has a lower affinity to isopentenyladenine (IP); both are plant hormones. In this system AHK4 upon binding trans-zeatin phosphorylates AHP5 (a phospho-transferase) and AHP5 phosphorylates PhoBVP64. Upon Phosphorylation PhoBVP64 translocates to the nucleus and binds its respective operator sites. The nuclear translocation was an interesting discovery since bacteria do not have a nucleus and therefor have no need for an NLS sequence. It is thought that when the phosphate group is transferred to PhoB the conformational change reveals a DNA binding domain that contains several positive residues that are recognized by nuclear transport proteins and translocated to the nucleus.

This system seemed like an attractive option to try to purport from plants to mammalian cells for several reasons. A previous graduate student Sairam had tried to implement a two-component system in mammalian cells from yeast and was unsuccessful. The most probable cause of failure was that the TF when phosphorylated in yeast changes the activity of its activation domain. Since the activation domain from yeast does not work in mammalian cells and was replaced by a mammalian activation domain the phosphorylation event would not affect transcription of the mammalian version. The PhoB system does not suffer from this problem. The yeast two-component system used the same receptor, AHK4 [25]. Also, we have already shown that mammalian cells have the ability to create IP (Figure 17) [1].

The first step was to take all the components from the paper and get them synthesized for use in mammalian cells. The codons were optimized for human cells, PhoBVP64 was changed to PhoBVP16, and I created a mammalian minimal promoter with PhoB operator sites (minCMV-PhoBx6). The proposed system would function as seen in Figure 17a.

Once I had all the components, I transfected them into HEK 293 cells with the addition of different concentrations of exogenous IP. As seen in Figure 18b, the cells are able to respond to different amounts of IP. However, you will notice that as IP increases the output from our promoter decreases and this system, as designed, should respond in a positive manner to IP. Also, you will note, that my constitutive fluorescent protein is also decreasing slightly with increasing IP. My initial hypothesis was that IP was causing some sort of toxic effect on the HEK cells causing protein production to decrease. To determine if the decrease in EYFP was in fact some sort of non-specific toxic side effect, I created an inverter to show that the response of my system is specific to IP and increases as IP increases. I was indeed able to show that IP is causing the change in expression and not some overall toxic effect that decreases protein production (Figure 19a). Also note that the response from the PhoB promoter is greater than the change in the constitutive promoter. To further validate that PhoB is causing the expression from the designed promoter, I put PhoBVP16 under the control of an inducible gene, rtTa3, varied Dox. and recorded the changed in expression. As you can see in Figure 19b, there is a positive correlation between the amount of PhoBVP16 and the activation from the PhoB promoter.

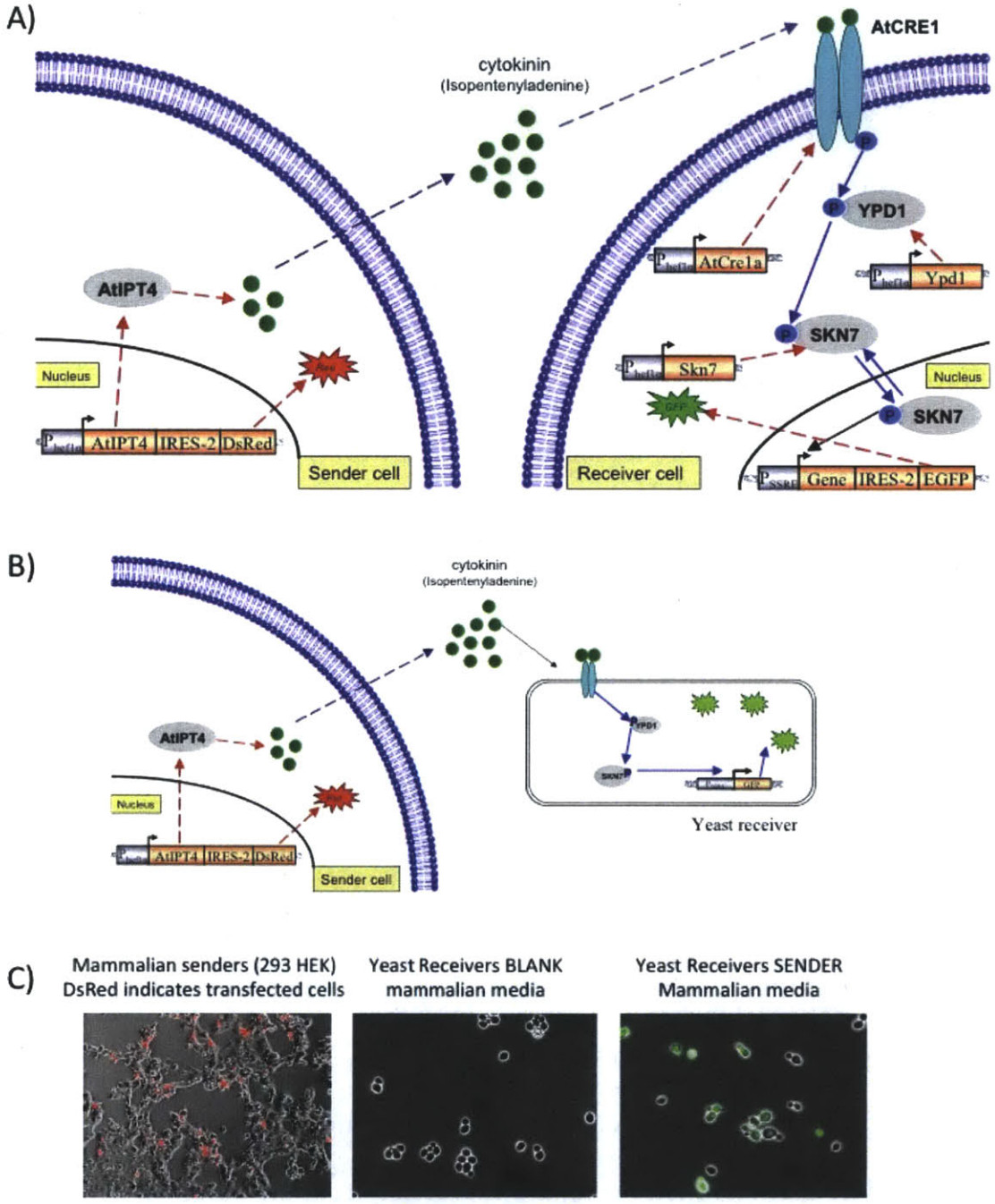


Figure 17 SKN7 Two-Component Signaling
A) Schematic diagram depicting a sender receiver system for Two-component signaling based on the yeast system that uses SKN7. The cell on the left produces IP and the cell on the right would detect it through a phosphoyrlation cascade. B) Depicts a system to test the mammalian sender with a yeast receiver. C) the results from testing a mammalian sender with a yeast receiver. Mammalian senders are on the left. Yeast receivers with blank mammalian media are in the center. Yeast receivers with Sender mammalian media are on the right. Figure adapted from [1] for explanatory purposes

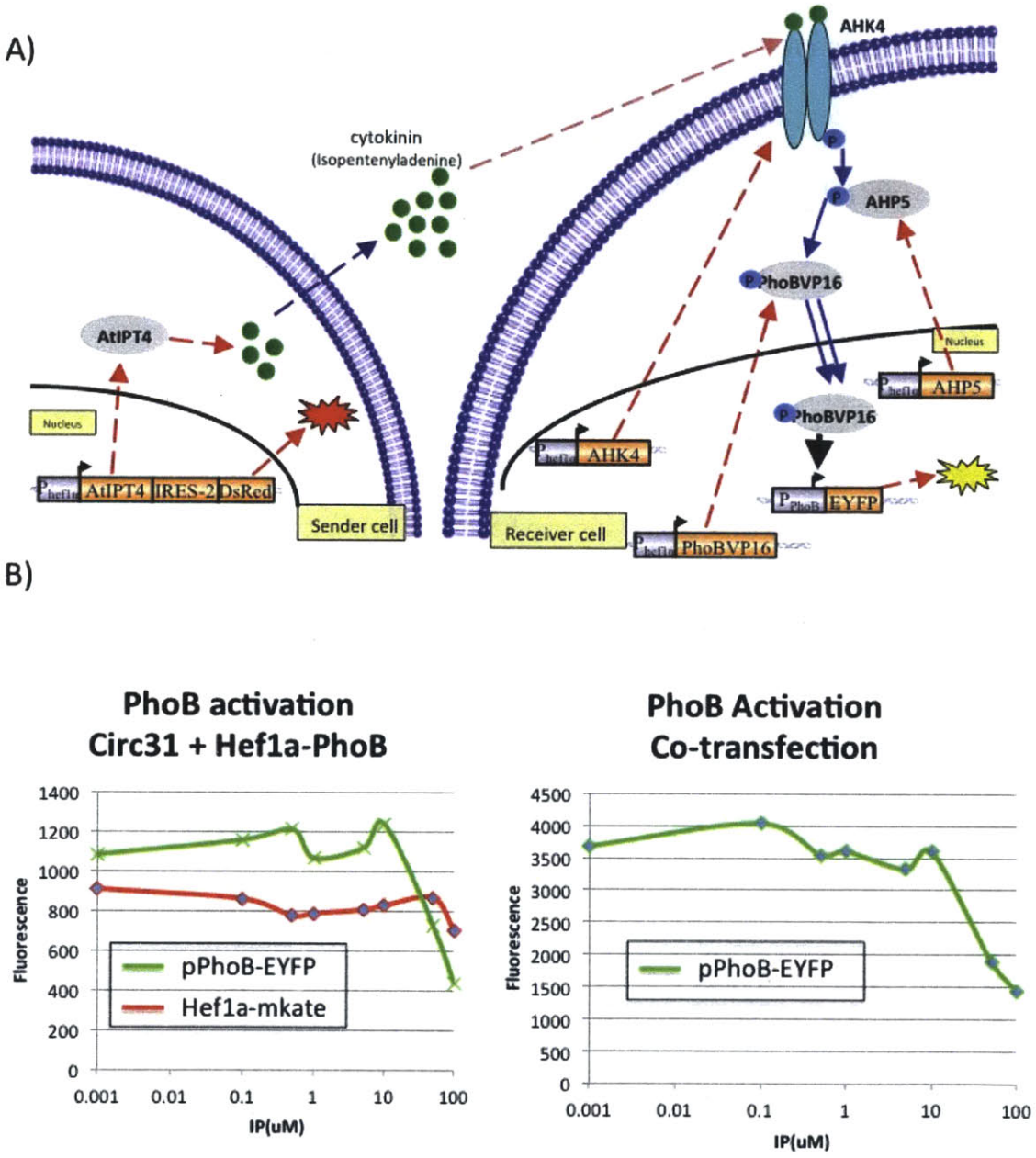


Figure 18 PhoB Two-Component Signaling
A) Schematic diagram depicting a sender receiver system for Two-component signaling based on the plant system that uses PhoB. The cell on the left produces IP and the cell on the right would detect it through a phosphoyrlation cascade. B) Results from mammalian receivers with exogenous amounts of IP added. The left graph represents a contransfection with all the components expect Hef1a-PhoB on a single plasmid cotransfected with Hef1a-PhoB. The right is a co-transfection with everything on separate plasmids.

The question still remained, how come the system is behaving contrary to the design? To elucidate this mystery, a red fluorescent protein, mkate, was fused to PhoBVP16 in order to visualize the translocation from the cytoplasm to the nucleus upon induction with IP. There did not appear to be a difference in mkate localization with IP induction. It appeared to be mostly cytoplasmically localized (data not shown). This appeared to suggest that small amounts of PhoB were needed to fully activate the system and the response from IP is not due to localization changes.

Future work would include determining if PhoBVP16 is being phosphorylated. Since His-Asp relays are hard to work with because the transfer of the phosphate is fast, I sought to find mutants of PhoB that are seen to be constitutively active or inactive regardless of phosphorylation state. A comparison could then be made between the activity of these mutants and the non-mutated version of PhoBVP16 currently in use. This might shed some light on what is taking place in this system.

5.2. Discussion

This chapter enumerates the multifarious methods I have employed as orthogonal mammalian cell-cell communication systems. Employing the bacterial quorum sensing system had difficulties with the sender cells whereby they were unable to produce sufficient quantities of AHL that were detectable by mammalian receiver cells. The VLP communication system was unable to show versatility in the "message" one would send inside the particle. There were also numerous immunological concerns with *in vivo* use. The PhoB receiver system appears to

detect exogenous IP but the exact mechanism of signal transduction in mammalian cells needs further elucidation. The TEV system has been shown to work in individual cells. The protease is able to cleave the TF from the membrane to activate transcription. This system appears promising and requires further experiments to demonstrate its efficacy. The design and response to antibodies will also have to be verified.

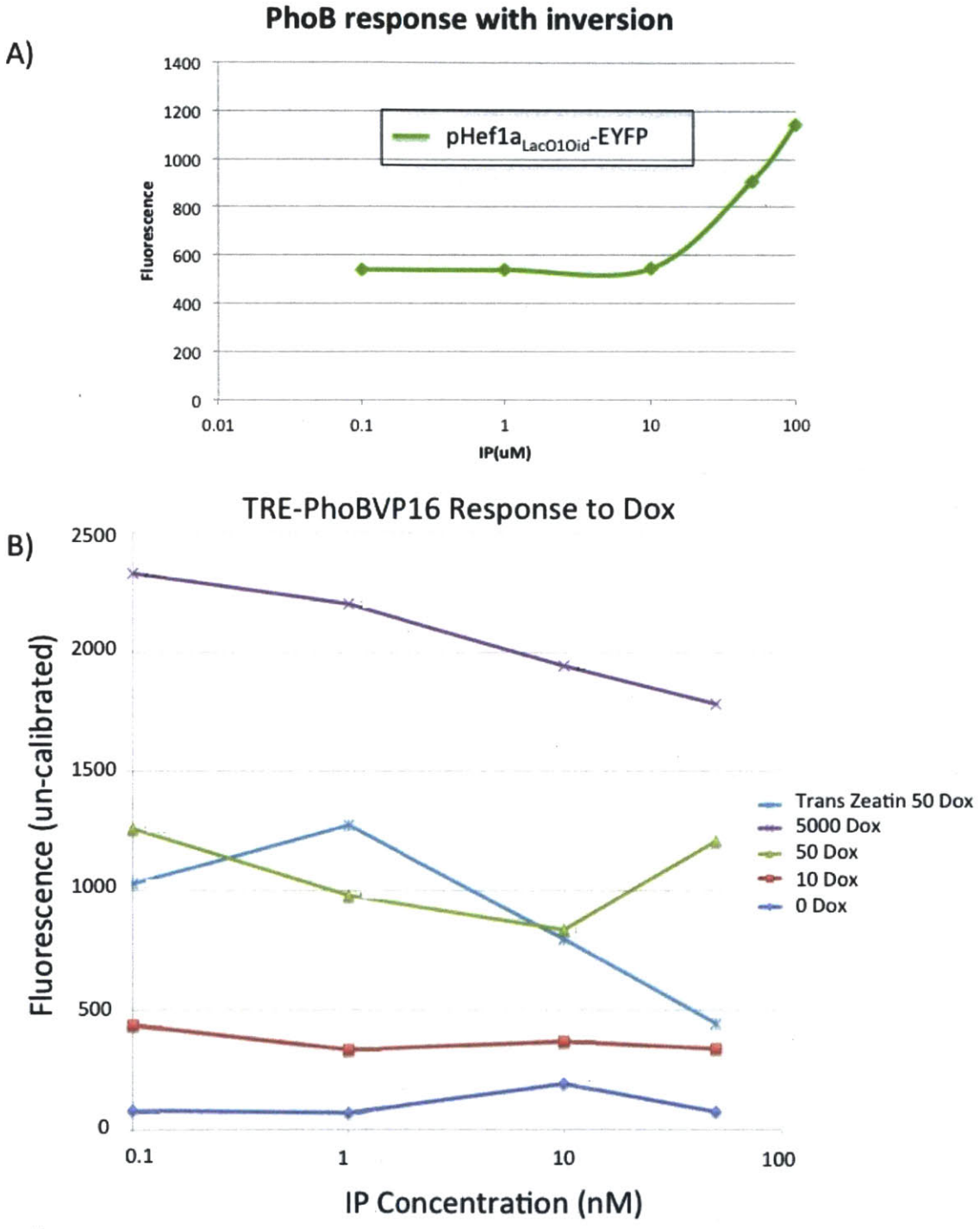


Figure 19 PhoB Characterization
A) A co-transfection of the PhoB receiver circuit, except the output from the PhoB promoter is LacI-mirFF4 which repressed Hef1a-LacO1Oid-EYFP. This made the response positive and increasing with increasing amounts of IP. B) An experiment with an inducible PhoB activator. This would allow us to determine if PhoB is responsible for the response we are seeing. Also would allow us to see if there is an optimal amount of PhoB in the system. Each line reresents a different concentration of Dox. while varying the amount of IP.

Chapter 6

6. Predictive Composition of Genetic Circuits

The TASBE (Tool-chain to Accelerate Synthetic Biology Engineering) project is a collaborative effort between several groups with the aim to make synthetic biology more automated [110]. A large thrust of this project is to be able to characterize a genetic part sufficiently such that a computer would be able to choose which parts to use in a new circuit (based on its functional specifications) and a robot would construct the necessary DNA to be tested. In collaboration with people from Raytheon BBN Technologies (Jacob Beal PhD, Fusun Yaman PhD, and Aaron Adler PhD) a new characterization methods was developed to help realize the end goal of this project; automatable design and construction of genetic circuits.

Synthetic biology aims to forward engineer genetic circuits from individual characterized parts. Many efforts have been made toward uncovering the necessary requirements of a characterization method that is capable of producing quantitative predictions that can be used for composition of novel genetic circuits. However, previous efforts were unable to garner enough information about individual parts to be able to engineer new genetic architectures in a quantitatively predictive manner. Here we show a new method for characterizing individual genetic parts, that is context independent and which allows predictable construction of novel genetic circuits in mammalian cells. Characterization of individual genetic parts was carried out through transient transfection, which allowed the in-depth study of the circuit's

temporal dynamics. Through direct use of each part's individual transfer function combined with the temporal expression dynamics we were able to successfully predict quantitatively a novel combination of two characterized parts. Six genetic cascades were created to validate the predictive power of the new characterization and prediction methodology. We were able to achieve a quantitative prediction within 1.25 fold average error for a cascade with greater than 15-fold gain. Our results demonstrate the validity and power of the characterization and prediction method developed. We anticipate this method to be used in creating libraries of well-characterized parts in mammalian cells that will be used to predictively engineer complex genetic circuits that were unable to have been created previously. This modular predictable composition method is a foundational tool of synthetic biology that will lead the way for better design principles and allow more complex circuits to be created.

6.1. Cascade

The cascade module is the component that was chosen as the 'proof of principle system' to show that with proper characterization and computational methods quantitative predictions could be made about the behavior of previously characterized parts in a novel architecture.

The reason the cascade was chosen was because the other two modules have added complexities. The toggle is a closed loop system and as such, in order to properly test its functionality it needs to be stably integrated it into the genome for long term steady state analysis. From the expression data in mammalian cells, the

switching dynamics appear to be longer than the time of a single transfection experiment due to the speed of transcription/translation in mammalian cells. The oscillator would also need to be integrated into the genome for long-term analysis. The cells will not be synchronized and as such flow cytometry measurements cannot be used to analyze the circuit behavior. Long-term microscopy experiments will be needed to characterize the oscillatory behavior and custom software will need to be developed for cell tracking and analysis.

The cascade (described previously in Figure 9c) functions as a double inversion gate. As Dox is added rtTa3 activates TRE and increases the amount of the first repressor. This repressor binds and deactivates the transcription of the second repressor. The second repressor levels drop and relieve repression of the output. The behavior of the cascade can then be summarized such that as Dox increases, the output increases. The timing dynamics and open loop system allowed for experiments to be done with transient transfection. Flow cytometry data could also be used to analyze individual cells behavior. By using flow cytometry we could obtain data on large numbers of cells, between $1x10^4$ and $2x10^5$ cells per induction point.

6.2. Characterization Circuit

The main thrust of Synthetic biology has been to develop a set of standardized parts that are characterized, modular, and can be assembled in any order. Most of the focus has been on developing a standard for the construction of these genetic parts (i.e Bio Bricks, Bgl Bricks, goldengate, etc...) [7, 27, 87] and little

has been done to develop a standard for the characterization of these parts. A characterization circuit that allows us to gather high quality data on a specific promoter-gene pairs in mammalian cells has been developed. An example of this circuit can be seen in Figure 20ab where the promoter-gene pair that is being characterized is Hef1a-LacO1oid and LacI-mirFF4 respectively.

The characterization of a genetic part involves obtaining the input-ouput relationship between the amount of TF (input) needed to produce a certain level of output response from its cognate promoter. This input-ouput relationship is called a transfer function. The characterization circuit seen in Figure 20 has been designed such that you can obtain this input-output relationship. This relationship is garnered by obtaining flow cytometry for the input fluorescent protein (IFP), which represents the input transcription factor, and the output fluorescent proteins (OFP), which directly measures the response of the promoter.

The TF is controlled through an inducible gene, (i.e rtTa3_Dox, RheoSwitch_RSL, VgEcR-Rxr_PonA, etc...)[88-90] rtTa3, and its inducer doxycycline (Dox) in a positive manner (i.e. the more Dox one adds, the more TF one gets). The inducible gene also controls a fluorescent protein (from an identical promoter on a different plasmid) that represents the amount of the TF being characterized. In some cases the cognate promoter for the repressors being characterized is a "hybrid" promoter in that it needs an activator to be in the "on" state but also contains operator sites for the repressor around the minimal promoter. In our design the activator used is VP16Gal4, which is produced from a strong constitutive promoter (CAG). (Figure 20)

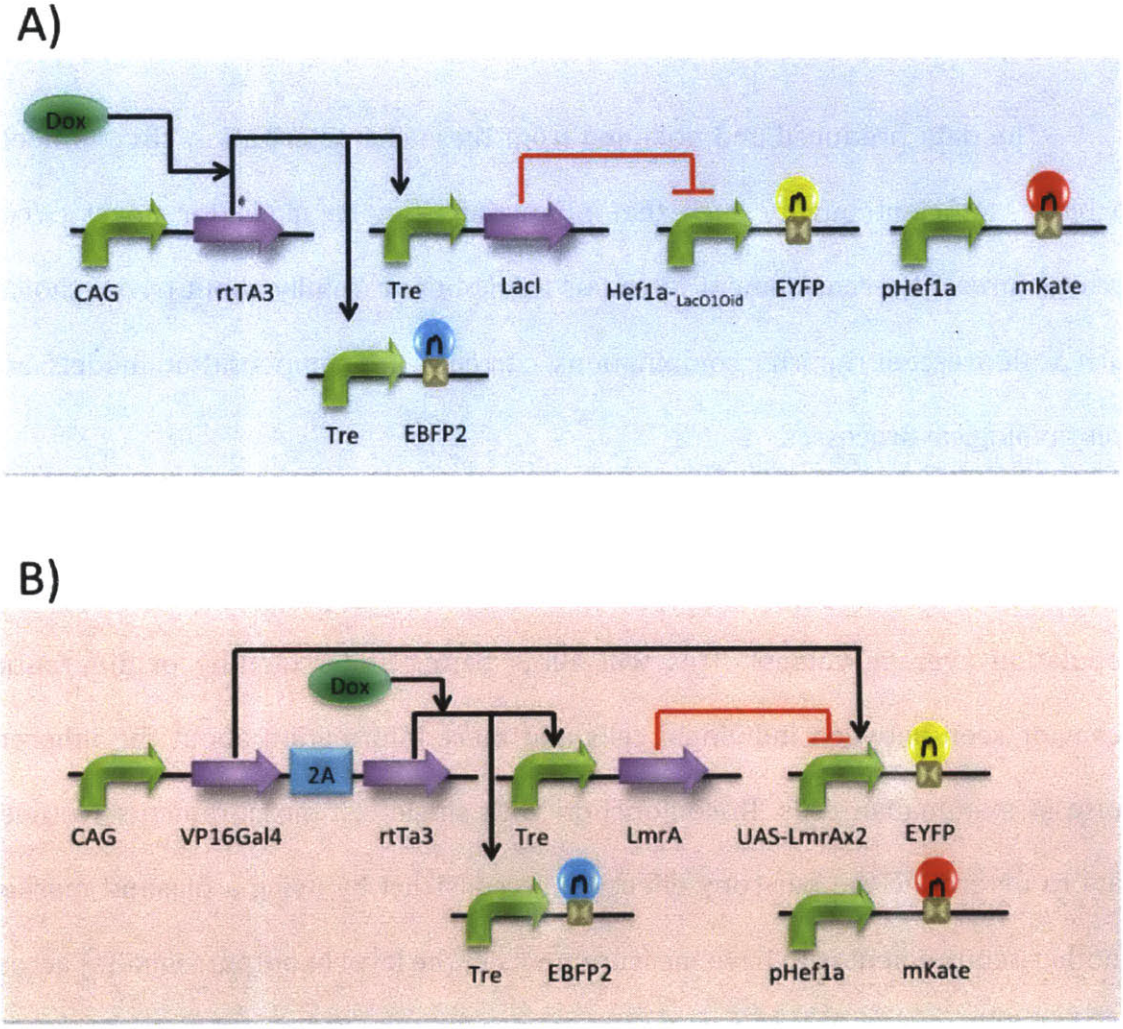


Figure 20 Characterization Circuits
A) A characterization circuit of a simple gene-promoter pair. In this circuit the transfer function obtained is of LacI-mirFF4 and Hef1a-LacO1oid pair. B) Depicted here is another characteriation circuit of a "hybrid" promoter. "Hybrid" because it requires an activator to be on, but can also be repressed by an orthogonal transcription factor. If the activator is always present then it can be considered to be a standard repressible promoter.

6.3. Data Analysis

6.3.1. Obtaining quality data

The data produced and obtained from these characterization circuits needs to be of sufficient quality such that predictions can be made for novel circuit architectures. This requirement forced us think more carefully about trivial choices such as fluorescent reporter combinations, current color compensation models, and noisy biological processes.

There are five main requirements on the data collected from each part. The first criterion is to have large numbers of single-cell measurements (as opposed to population average values). This will allow better understanding of differential behavior seen between individual cells and more information about the inherent noise in mammalian cells. In conjunction with single cell measurements we need data to determine the per-copy effect of the construct by using a plasmid marker. Another requirement is to have measurements of the level of output signal(s) across the full dynamic range of levels of part input signal(s). Without the full range of information about each part we would not be able to compose new circuits regardless of context. In order to estimate the noise in the system we need the statistical distribution of single-cell output levels for each input level. And finally we need information about the expression of the different proteins over time in order to predict larger composite circuits in which the course of expression over time is significant.

None of the previous efforts mentioned in Chapter 2 satisfy all five requirements above, and thus cannot produce the kind of characterization data that is necessary for predictable part composition. Understanding the minimal set of necessary and sufficient information on each part allowed for the development of a method to characterize parts for the predictive composition of genetic circuits.

The current characterization system comprises three fluorescent reporters, IFP (input fluorescent protein), OFP (output fluorescent protein), and CFP (constitutive fluorescent protein). Fluorescent proteins have wide emission spectra with a long "tail" that is likely to overlap with another fluorescent protein emission spectra. Initially the IFP, OFP, and CFP chosen were AmCyan, EYFP, and mKate respectively. A complication immediately arose from this combination of fluorescent reporters. The contribution from AmCyan to EYFP was approximately 10%. That meant the OFP was incorrect by 10% and the transfer curve is not accurately representing the input-output relationship. Greater than a 1% bleed through (contribution from one fluorescent protein to the channel of another fluorescent protein) from one channel to another was unacceptable because of the noise created and confounding of contribution of signals. This limited the quality of data and therefore the accuracy to which our predictions could be made.

There were several avenues for improvement that were explored after this issue arose, a biological, a hardware, and an analytical one. In the end, a combination of all three was employed. The biological avenue consisted of finding a new fluorescent protein or new set of fluorescent markers that could be used together to produce IFP, OFP and CFP with less than 1% bleed through (less than 1% of the

signal being contributed from a neighboring channel). EBFP2 is one such protein that fluoresces farther away from the EYFP spectrum [91]. Several commercial products (HaloTag Promega; Madison, WI) that produced narrow spectral fluorescence were considered. However, to use these tags one had to add a specific amino acid sequence in which the fluorescent probe could bind. These were non-anti body based and could be used for live cell analysis but required more work than using fluorescent proteins and were a secondary measure if the fluorescent proteins did not work.

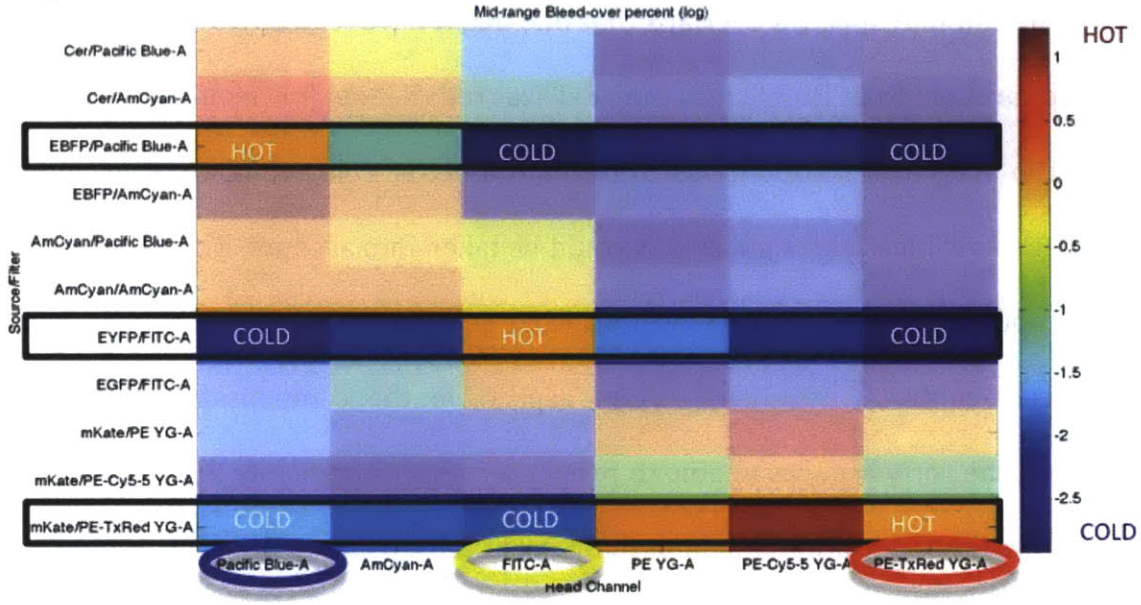
The flow cytometer hardware used is the LSR Fortessa from BD Biosciences and can house up to 4 lasers with numerous filters for each laser. The correct laser/filter combination is selected for each channel in order to maximize signal and minimize crosstalk. There is some flexibility in choosing which laser and filter to use in order to find the optimal combination.

To determine the optimal choice of fluorescent protein and laser/filter combination, a heat map was created testing all three of these parameters (Figure 21a). In this Figure, on the left is a combination of a fluorescent protein and a filter and this is compared to the filter on the bottom. For instance, if one look at EYFP/FITC-A on the left hand side compared to FITC-A on the bottom there is a color value of 0 (the normalized value for this protein filter pair). For that particular row, (EYFP/FITC-A) we can compare how well other filters detect EYFP compared to FITC-A. In this case no other channel detects EYFP signal better than the FITC-A channel. The optimal combination of fluorescent proteins and filters can therefore be determined by selecting fluorescent proteins/filter pairs that are uniquely hot on

the map when compared to each other. An interesting consequence of this heat map is that one can find spectral overlaps that one might have logically overlooked such as 'bleed through' from EBFP2 into the PE-Texas red channel (or mkate channel).

The analytical solution was to develop a better color compensation model such that small 'bleed through' effects could be taken into account. The BD software has the option to perform compensation from one channel to another when using the correct controls. However, certain aspects of the compensated data were troubling. The software was unable to properly compensate highly fluorescent cells at the same time as low and medium fluorescent cells. Using their software one had to choose which data to compensate. This was unacceptable for our data requirements. The issue that existed in the BD software was that it took into account the autofluorescent population and tried to compensate everything equally. By ignoring the autofluorescent population a simple linear relationship can be found that compensates properly for each pair of channels (Figure 22b). If a new fluorescent protein and/or filter was to be used, one could just use the same method to develop a new compensation curve for the new combinations of colors.

A)



B)

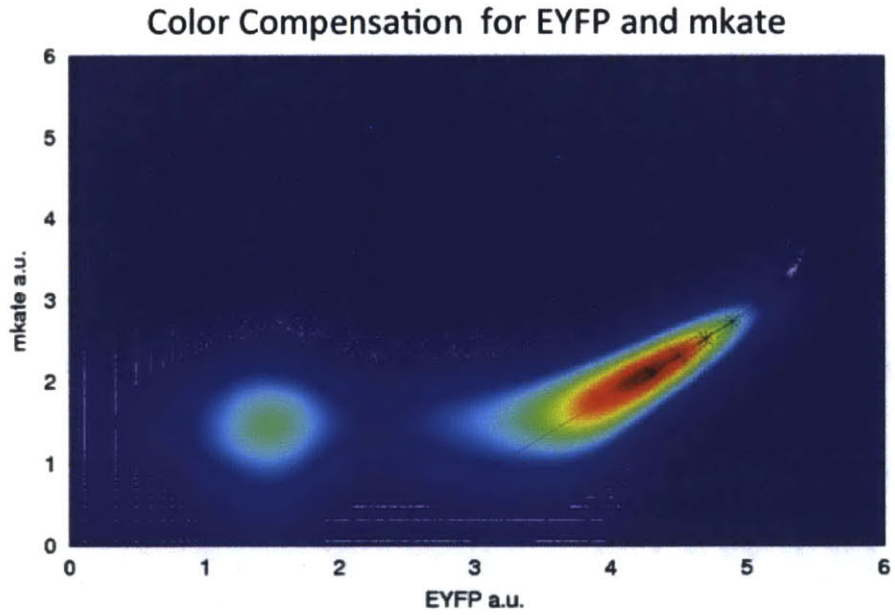


Figure 21 Spectral Overlap And Compensation
A) A heat map that describes the bleed through of proteins into other channels. The left hand side has protein/filter pairs and the x-axis has the read-filter. The proetin/filter pair on the left gets compared to the signal read from the filter on the bottom and a heat value is assigned compared against a self filter normalization. EYFP/FitC-A when compared to FitC-A gives a heat value of 0 and then EYFP/FitCA when compared to Pacific blue-A is cold. B) A graph representing the color compensation fit used to correct for bleed through effects. The light blue dots are data from a single constitutive color and the x-axis is the fluorescence in the Fitc channel, the y-axis is the fluorescence in the mkate channel. Here you see bleed through into the mkate channel from high EYFP values that is non- linear. A linear relationship is represented by the red line which does not fit the data properly. Figure adapted from [85]

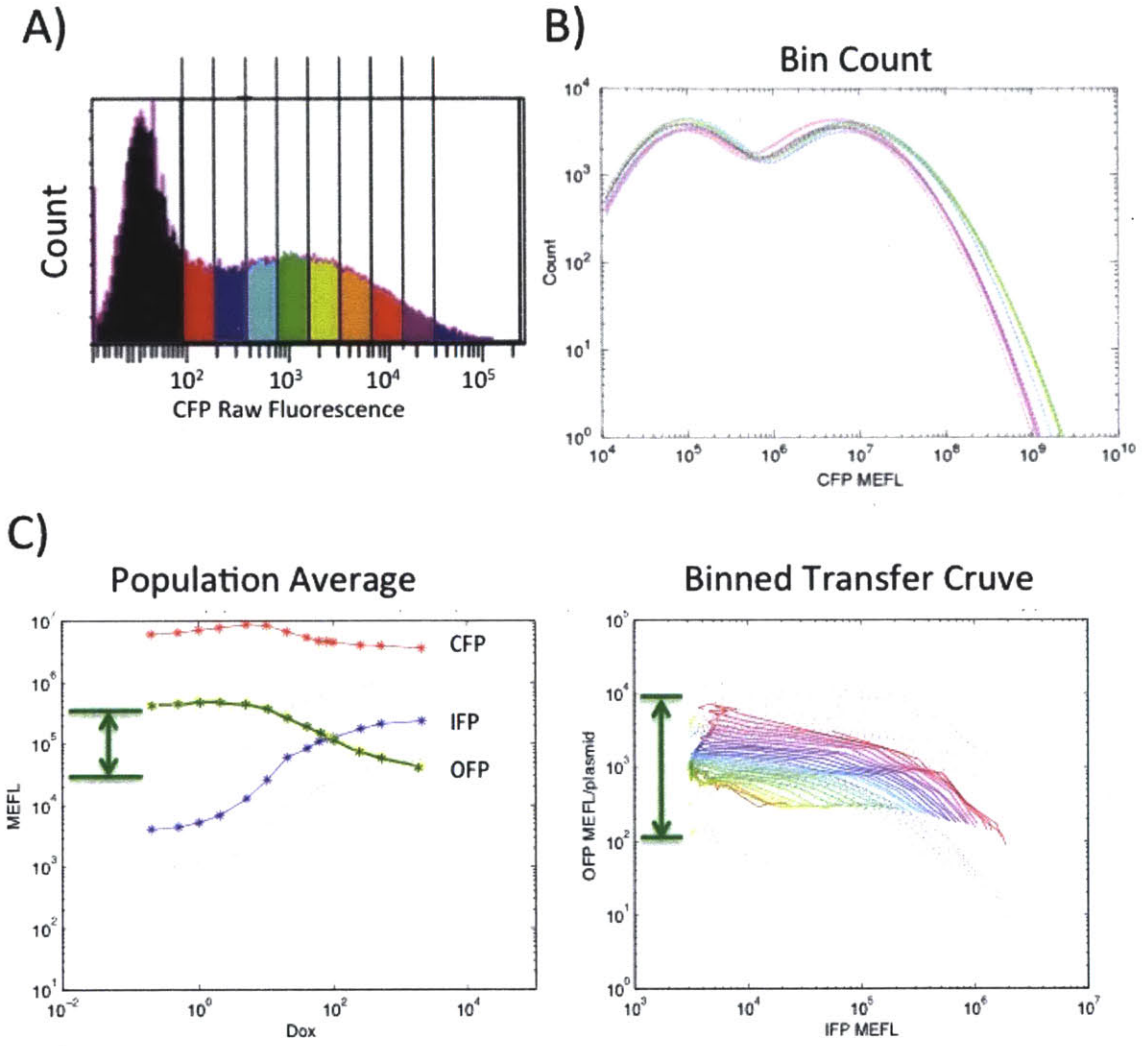


Figure 22 Binning
A) A representation of binning on the constitutive fluorescence. The blue lines represent the division of the cells into bins. In practice there are bins every 1/5 of a decade. The colors of each bin correspond to the color coding of the transfer functions. This particular binning profile is only a schematic representation of the actual binning B) A representation of the number of data points in each bin and the computer fit to each of these. There are two gaussian curves modeling the CFP to get a better description of cells to be included in analysis. C) The difference between analyzing population averages versus binning the cells into small bins that better represent individual cell behavior. Figure adapted from [85]

6.3.2. Analyzing data

Following the practices laid out in the previous section to obtain data of sufficient quality, one must now analyze the data to understand the behavior of the system. Most data analysis used for transfer functions has looked at population behavior [92]. The expectation for mammalian cells was such that we believed population behavior would be insufficient to fully understand the system. One reason for this hypothesis was due to the method of delivery of the genetic circuits to the cells. Lipofection was used to introduce 6 separate plasmids into HEK 293 FT cells (Invitrogen) using Metafectene Pro (Biontex). This method of DNA delivery produces a wide variance in the number of circuits each cell receives. It is also generally accepted that this method of DNA delivery allows each cell to get approximately equal numbers of each plasmid for the circuit due to the "packet delivery" method seen in lipofection [93].

Under these assumptions we might see differing circuit behavior in cells containing one copy of the circuit compared to cells containing hundreds to thousands of copies. Using population level analysis could obfuscate the underlying difference in cells with more or less copies of the circuit. To investigate this possibility, a two-dimensional binning analysis was developed to separate out the cells by induction level and plasmid copy number [85]. As mentioned before, to obtain a marker for copy number a plasmid producing a constitutive fluorescent protein was included in the circuit that was transfected. The more plasmids a cell receives, the more fluorescent protein produced [94].

Using this constitutive copy number the cells were divided into bins based on the strength of their constitutive fluorescence (Figure 22a) Initially, there were not enough data points obtained from the flow cytometer to split the cells into too many bins. This becomes important for the bins for very low and very high constitutive fluorescence, since one expects a somewhat Gaussian distribution for copy number. If the bins do not contain enough cells the data will be too noisy and unrepresentative of the proper behavior of the cells. There should always be at least several hundred cells in the smallest bins and thousands of cells in the center bins. A representative cell-bin count is seen in Figure 22b. The CFP distributions are approximated by two overlapping Gaussian distributions, the autofluorescent population and the population that received our circuit. By using this double Gaussian one can separate out much better the cells that should be considered in the analysis and the ones that should not.

Once the cells are split into separate bins, a transfer function can be garnered for each bin that describes the behavior for that copy count. From this we can see whether or not different copy counts give different types of behavior for the same circuit. There is a differential behavior in the cells with lower number of circuits as can be seen in (Figure 27a middle), and without binning we lose 10-fold sensitivity (Fig 22c). In order to compare the circuits from one cell to another with differing copy number we needed to normalize the data in each bin. For our transfer functions we normalized only the output. This gives the response per promoter for a given input. Since we are able to get the full range of input we do not care if the

input came from one promoter or many, a single copy of the output promoter still "sees" that particular level of input (Figure 23a).

As one can see from Figure 23b, there is a spreading of the data from one bin to another. There is also an "inversion" from the raw data to the normalized data in that the high copy number bin (in red) is on top in the raw data and is now on the bottom in the normalized data (Figure 23b). These phenomena caused quite a bit of concern for a while. There were two possible explanations for the spreading of the data, one was that as copy count increases the output from an activatable promoter (such as TRE) levels off and at some point does not increases linearly like the constitutive promoter does. The other possibility is that by binning the cells one creates a sampling bias. As one goes farther away from the mean number of plasmids per cell the effect becomes stronger (and a sigmoidal like appearance is seen in the expression from TRE as compared to constitutive).

We did not think that the activatable promoter would behave differently from a constitutive promoter and thus sought an analytical solution. The sampling bias is explained by the unequal contribution from neighboring bins. This occurs because expression noise is greater than the bin size and because of the underlying plasmid copy-count distribution. This contribution becomes more pronounced the farther away from the mean copy count one goes.

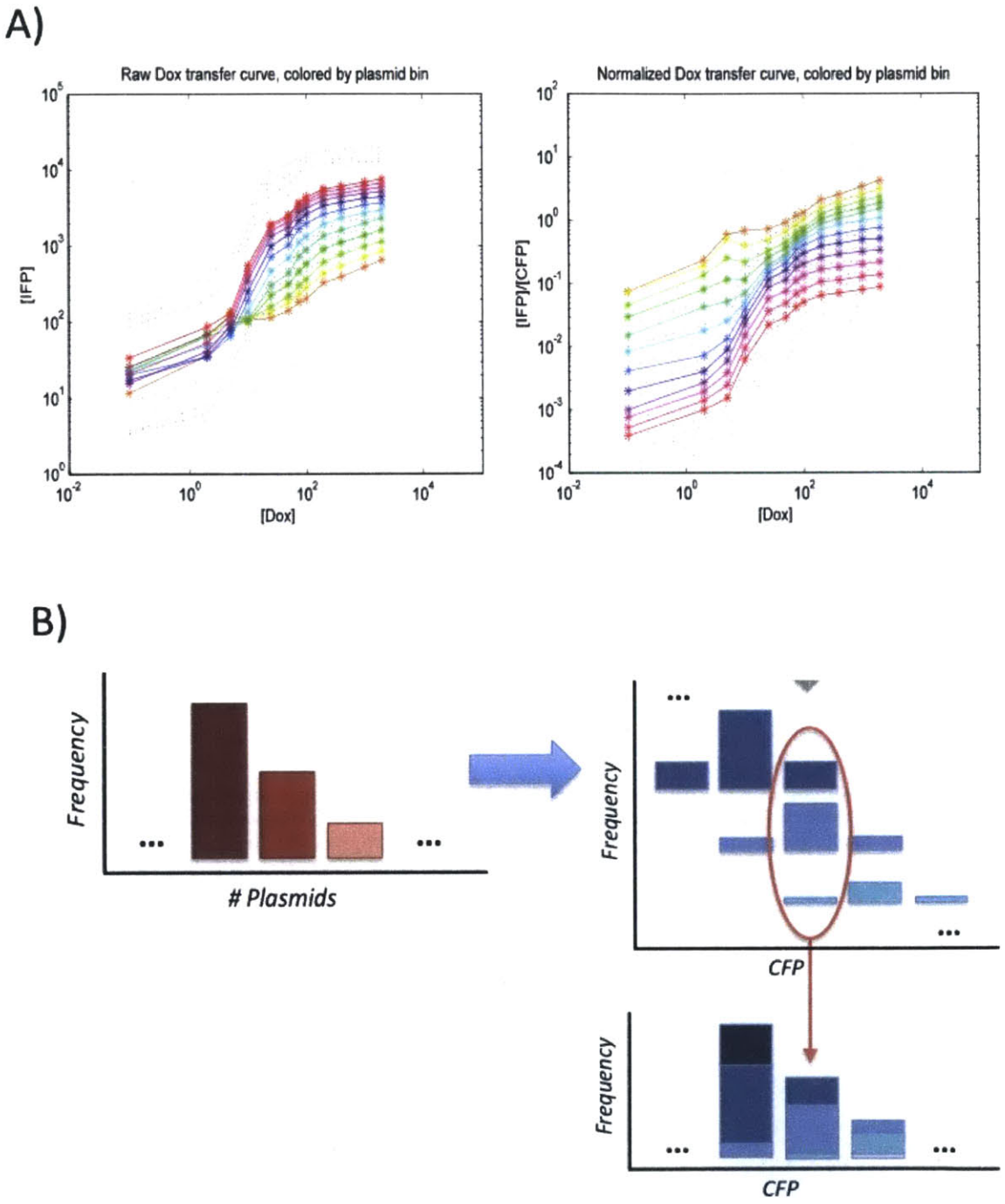


Figure 23 Normalization Inversion and Sampling Bias
A) There are two Dox induction curves. The one on the left is not normalized according to plasmid count. The one on the right is. There is an inversion in colors from the graph on the left to the graph on the right meaning there is over normalization for high plasmid count, and under normalization for low plasmid count. B) This describes the sampling bias. Our actual constitutive distribution is made up of cells from neighboring bins because of expression noise. The ratio of contribution from neighboring bins changes as you move farther away from the plasmid copy number mean. We are able to deconvolve the contributions from each bin and correct for it. Figure adapted from [85]

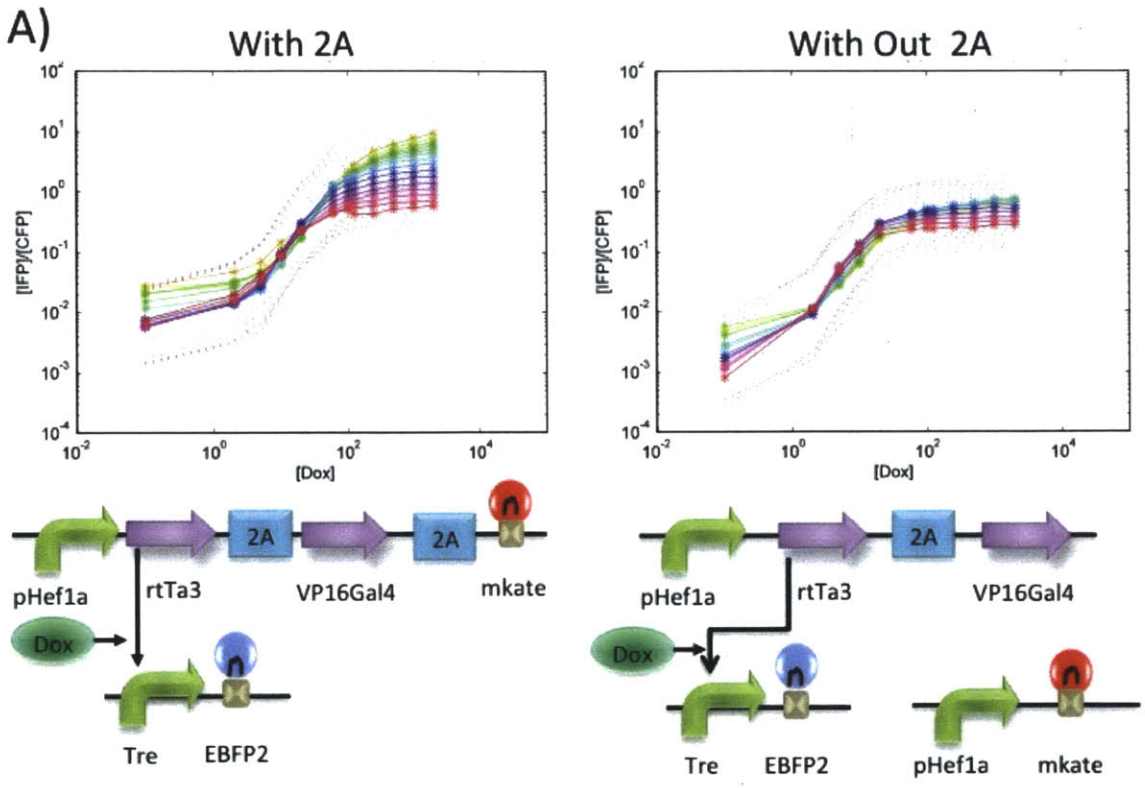


Figure 24 2A Effects
A) A comparison of noise with and without a 2A tag used in the expression of the consitutive fluorescence. On the left the construct includes a 2A tag and as such has a wider spread than the graph on the right. The two network topologies are underneath each graph respectively.

The solution to this problem is found by using Bayes' law, where the convolution of two Gaussians is a Gaussian where the geometric mean is multiplicative and the geometric variances are summed. In this system there are two types of distributions that have been convolved together to create the constitutive fluorescence distribution (which is obtained directly through flow cytometry measurements). The two distributions that are convolved together to create the observed distribution are expression noise and copy number distribution.

The expression noise is measured by taking the variance in the distribution in the two other channels (i.e. IFP and OFP). By measuring expression noise distribution and constitutive expression distribution, and assuming plasmid distribution is a Gaussian (since it is a random process), Bayes law allows us to deconvolve expression noise from our constitutive expression distribution to obtain our plasmid copy number distribution (Figure 23b). Once we are able to obtain the plasmid copy number distribution we can then un-bias our sampling bias as seen in Figure 24a [85].

Another consequence of stringent requirements on the quality of our data was that slightly more noisy processes, which normally were overlooked from a biologists' perspective, could not be ignored from an engineering standpoint. A problem arose when trying to express multiple proteins from one promoter using certain biological "tricks" found in nature (namely the 2A tag). Normally in mammalian cells only one protein can be expressed from one promoter, called mono-cistronic expression.

Certain viruses have developed a trick to express multiple proteins from one promoter (most likely due to their inherent size limitations). The 'trick' employed in these constructs was the use of an amino acid sequence known as a 2A sequence (discussed previously). This sequence causes the ribosome to 'skip' and translationally creates two separate proteins from one single transcript. However, the ribosome does not skip 100% of the time. Only a certain percentage of the transcripts are separated, another portion of the transcripts only have the first protein translated because the ribosome presumably 'falls off' the transcript. And yet another portion of the transcripts could create a giant non-functional protein where both the first and second interfere with each other [46].

This creates noise in the expression of this construct that is not taken into account in our current analysis and models. Therefore, fluorescent proteins and input proteins to be characterized cannot use this 2A tag. We found a significant increase in the accuracy of our analytical methods when the expression of the constitutive marker, mkate (which accounted for plasmid copy count), was expressed from its own promoter (Figure 24a). This is seen by the reduced spread from the normalization technique when comparing with and without the 2A tag expressing the constitutive fluorescent protein.

6.3.3. Timing analysis

The dynamic behavior of gene expression led us to study more closely the effect this had on our characterization circuits and genetic cascades. Because we use relatively stable proteins and transient transfection, our circuits will not generally be in equilibrium; we gather time series data to characterize how the observed expression level changes over time which is particularly important for transient transfections, but also relevant for any circuit with non trivial dynamics.

We measured expression levels for constitutive expression and for rtTA3 and VP16Gal4 driven expression at 16 time points ranging from 12 to 72 hours post-transfection (Figure 25). We found that the fraction of cells with observable (i.e. above auto-fluorescence) constitutive expression of a fluorescent protein increases linearly following a short delay, saturating at approximately 70% of cells at 42 hours (Figure 25). Given an unsynchronized population of cells, this observation is consistent with typical lipofection efficiency and the standard lipofection hypothesis of plasmids entering the nucleus during mitosis. There is no significant difference between the rate at which the mean fluorescence of the expressing population rises to its saturated level for constitutive expression in comparison to activation by rtTA3 and VP16Gal4 (Figure 25). This implies that transcriptional activation delays in protein production are not significant factors in the behavior of the circuit architecture under consideration and are further obfuscated by the wide range of expression variation observed in the transfer curves.

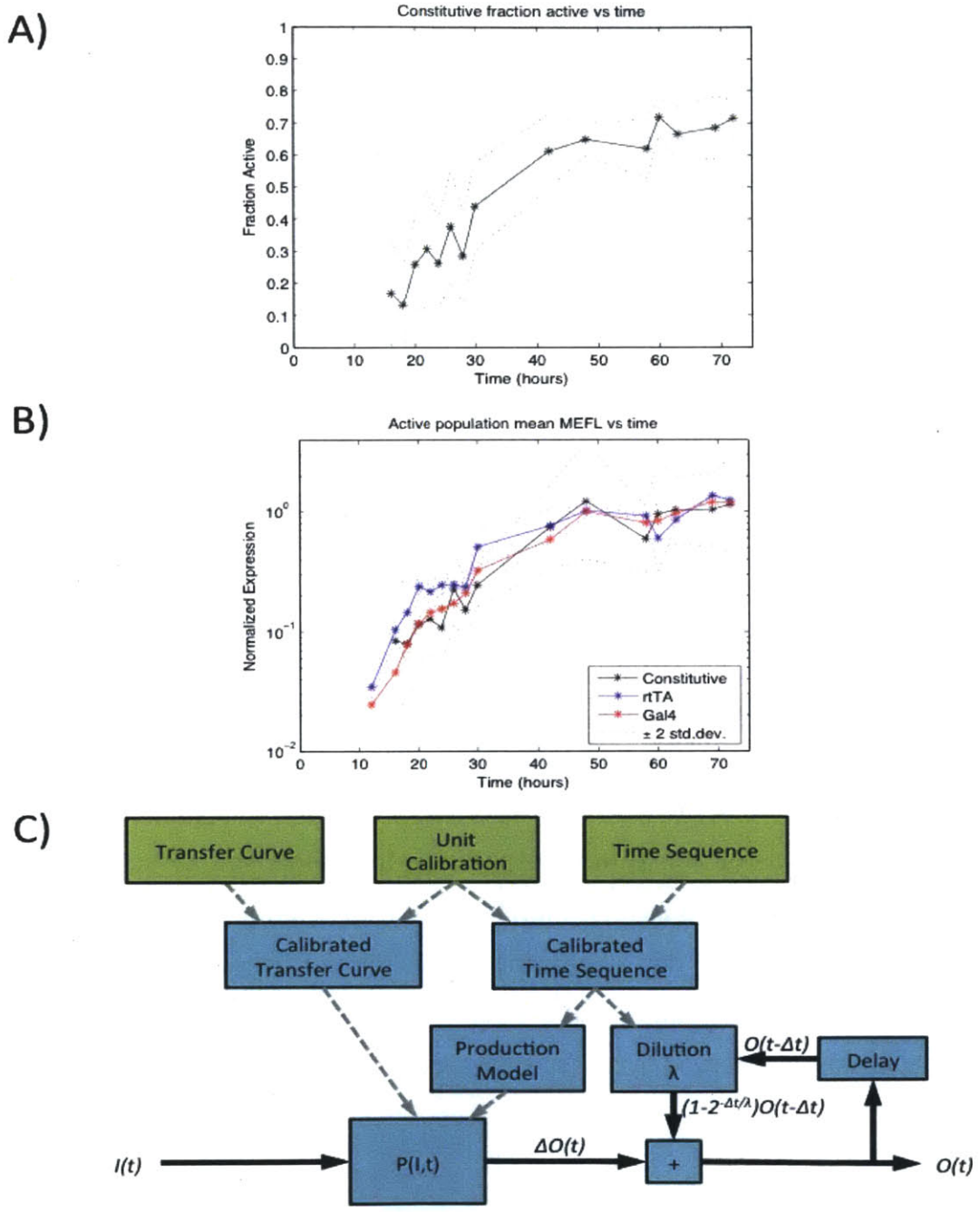


Figure 25 Time Course Information and Prediction Methdology
A) Following a short delay, time sequence characterization shows a linear increase in the fraction of cells constitutively expressing a fluorescent reporter (a), until reaching a saturated level of approximately 70% transfection efficiency. Progression of mean fluorescence level (b) is similar for both constitutive and activator-driven fluorescence, implying little impact from time delay or unsaturated activation. Normalized expression for each sequence is computed by dividing by mean MEFL for t = 48 to 72. (C) EQuIP predicts the output expression level $O(t)$ of a biological device at time t from its input $I(t)$ using an incremental model of the production and loss of output. This model is taken from empirical characterization of transfer curve, time sequence, and unit calibration. Figure Courtesy of Jacob Beal.

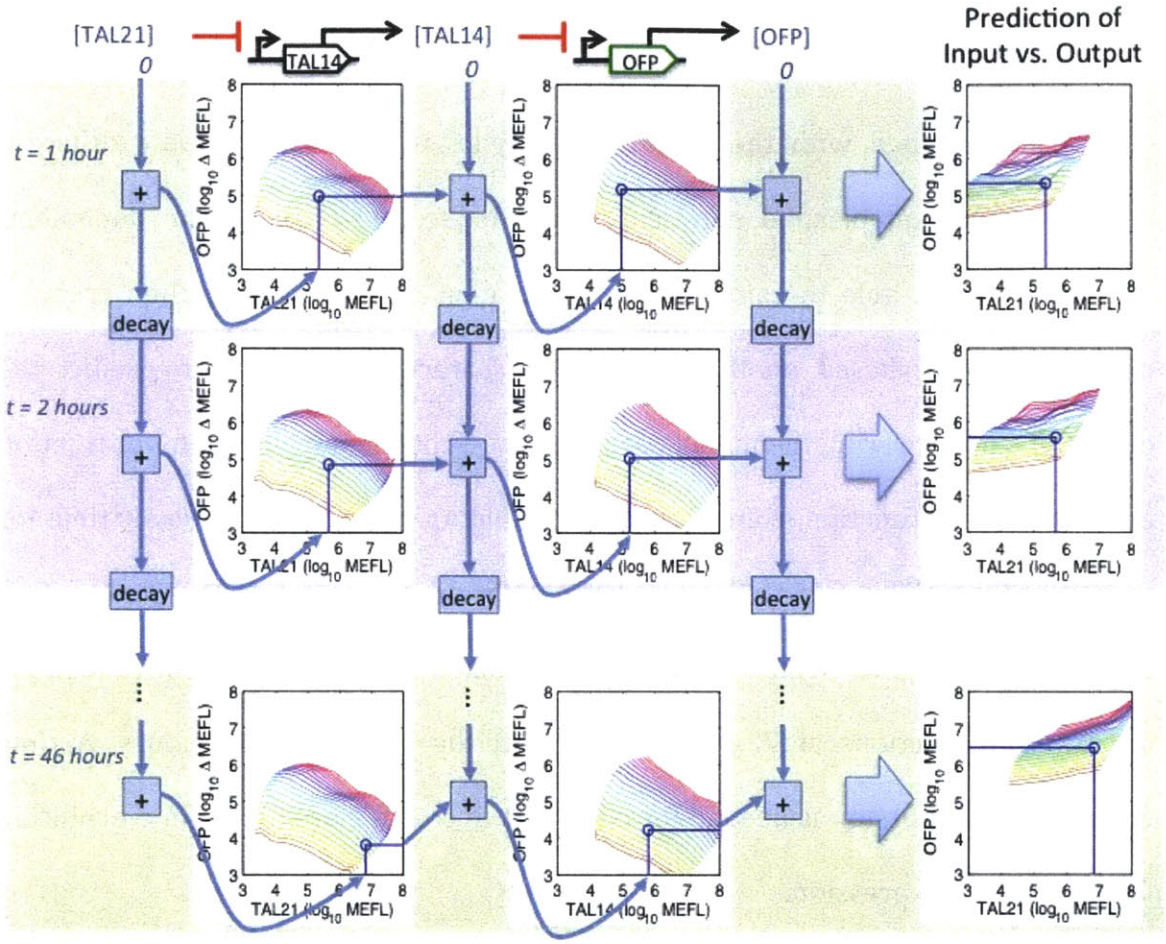


Figure 26 Prediction Method
A) Example of prediction for TAL14-TAL21 circuit, showing prediction of the 7th CFP bin under induction with 50 nM dox. Information flows through the model down over time and right from device to device. Concentration of the input repressor, TAL21, rises and dilutes with each time step. A lookup on the rescaled transfer curve finds the amount of TAL14 produced, which accumulates and is looked up on its own transfer curve to find the amount of OFP produced. The predicted transfer curve (right) thus shifts incrementally over time, predicting the relationship at different instants of time. Figure Courtesy of Jacob Beal.

6.3.4. Predictive model

With the ability to obtain high quality, single cell, measurements for each of our parts combined with the advanced analytics used to differentiate different behavior based on plasmid copy count and inducer level in a time dependent manner we were able to gain enough information on each part to then create a predictive model based on the individually characterized parts. To predict the output of a novel circuit, the model uses the time-dependent expression information and the transfer function from each part to build up a transfer curve over time for the new circuit (Figure 25c). Each part's transfer function is taken at 72 hours (the same time the cascade transfer function is obtained). It is not sufficient to just use the transfer functions at 72 hours to predict the cascade at 72 hours. A time evolution of expression is needed for each part to create a predicted time evolution for the cascade expression.

After obtaining time-dependent information about the dynamics of the repressors and activators used in the circuits, a model was built to simulate the expression over time for the cascade that led to the transfer functions obtained for each part at 72 hours. The predicted output for the cascade is taken directly from the transfer functions of the individual parts over time. The output from repressor 1 for a given induction level, plasmid number, and time point is taken as the input for repressor 2 at the same plasmid number, induction level, and time point. This is then traced through the transfer curve for repressor 2 to predict the output of the new cascade circuit at a given time point (Figure 26). By generating the time evolution of the cascade expression from the individual transfer functions based on

their expression dynamics, we are able to produce quantitatively accurate predictions for dynamic circuits without ever knowing specific rate constants or having detailed information about the operation of each part. We use the transfer function obtained for each part combined with time dependent expression information to predict novel genetic architectures.

6.4. Characterization and Prediction results

The characterization data obtained consists of a doxycycline induction curve that provides exact input levels and a corresponding transfer function relating the input level and the normalized output. A sample of this type of data can be seen in Figure 27. I have characterized 3 parts, LmrA, TAL14, and TAL21.

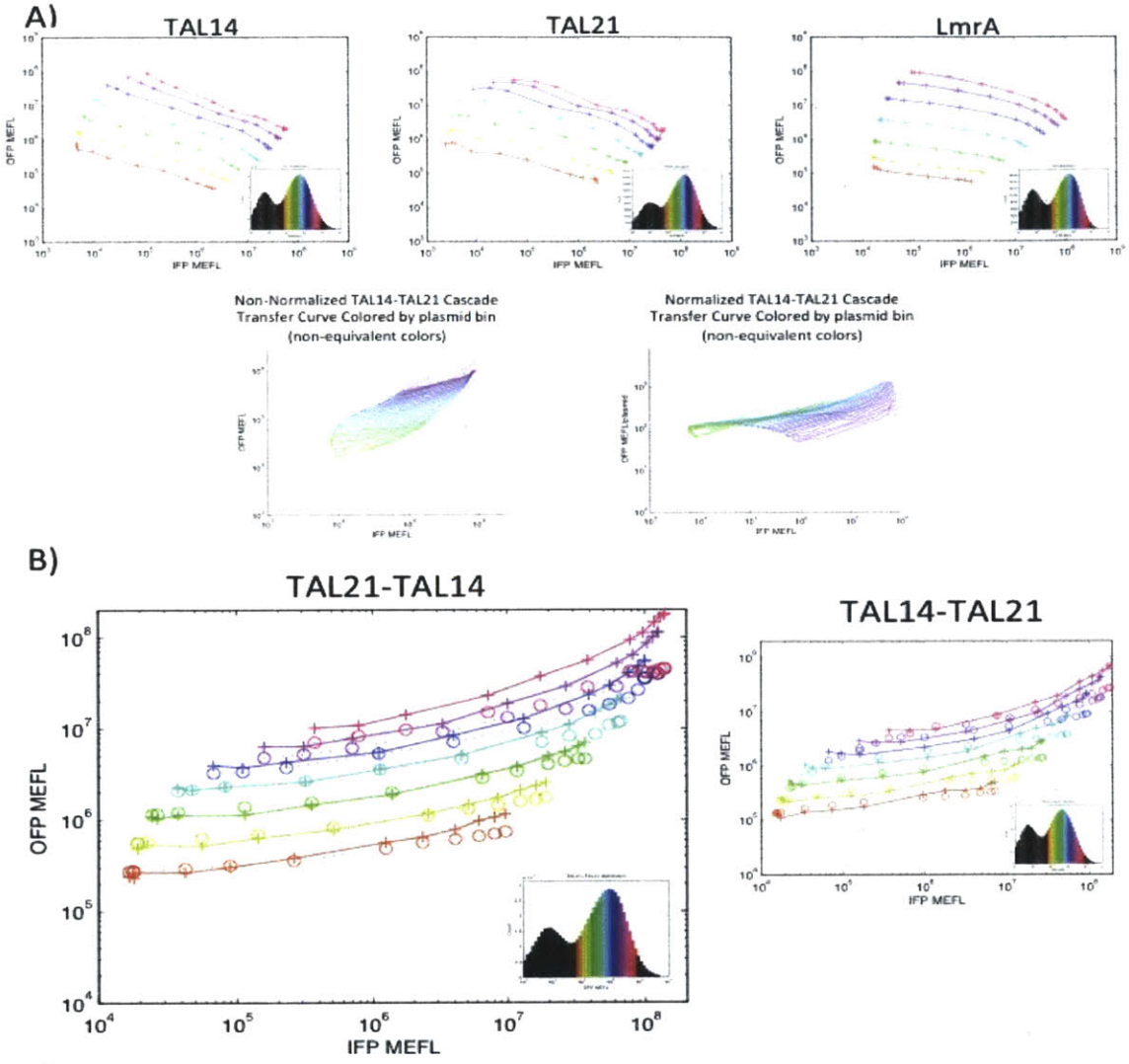
As mentioned previously, when collecting data to fully characterize a part, it is necessary to properly cover the range of input. It was determined that a logarithmic progression ranging from 0.0nM to 2000nM was optimal (i.e. 0.0, 0.1, 0.2, 0.5, 1, 2, 5, 10... 2000). 2uM was the maximum induction concentration because toxicity was observed above 2uM without an increase in output.

Figure 27 shows predictions for two sample genetic cascades (TAL21-TAL14 and TAL14-TAL21) (Figure 27b) and sample individual transfer functions of some parts characterized (Figure 27a). The circles in Figure 27b represent the predicted result for each cascade and the stars are the data obtained through flow cytometry of each cascade. The lines connect different induction level data belonging to the same bin (based on CFP levels, like the inset at the bottom shows). The non-normalized view allows a clearer picture as to how well the predictions line up with

the actual data. The predictions for the cascades are within a 1.6 mean squared error.

6.5. Discussion

The characterization framework, analysis tools, and prediction models developed make a significant contribution to mammalian synthetic biology. Our ability to make predictions without specific rate constants or operational details about the parts makes this a powerful abstraction. Using this method we merely require the transfer function of the individual part and the time-dependent expression information for these parts in mammalian cells. Synthetic biology, like electrical engineering before it, needs abstractions like this in order to build more and more complex circuits. By having standard predictable modular units that can be composed into larger more complex architectures it shifts the focus from "what **can** we create?" to "what **should** we create?". This new powerful tool will help to foster the creation of huge libraries of characterized parts that can be used to usher in anew wave of increasingly more complex rationally designed mammalian circuits.



C)

Cascade	Mean Prediction Error	
	Output Level	+/- Induction
TAL14-TAL21	1.49x	1.61x
TAL14-LmrA	1.72x	1.34x
TAL21-TAL14	1.46x	1.87x
TAL21-LmrA	1.52x	1.66x
LmrA-TAL14	1.72x	2.27x
LmrA-TAL21	1.84x	1.93x

Figure 27 Characterization and Prediction Results
A) Characterization transfer curves for each of three parts, LmrA, TAL14, and TAL21. Inset represents actual bin colors in the transfer curve. BOTTOM LEFT: A non normalized TAL14-TAL21 cascade transfer curve. BOTTOM RIGHT: A normalized TAL14-TAL21 transfer curve B) LEFT: Prediction results overlaid on non-normalized data for easier viewability for TAL21-TAL14 cascade. Inset represents actual bin colors for transfer curves RIGHT: Prediction results for TAL14-TAL21 cascade. Inset represents actual bin colors for transfer curves. C) Table showing the mean prediction error for all cascades created. The cascades curves not shown in (B) can be found in the appendix. Figure Courtesy of Jacob Beal.

Chapter 7

7. Conclusion and Future work

7.1. Conclusions

This thesis establishes a foundation for engineering genetic regulatory networks and intercellular pathways of communication in mammalian cells. A library of well-characterized parts and modules is of paramount importance when complex systems are to be engineered in multi-cellular organisms. The DNA assembly method presented in Chapter 3 allows for unprecedented speed and versatility when constructing mammalian genetic components. The advance in DNA construction technology permits us to concentrate on design and testing. This creates the opportunity for future researchers to pursue more complicated and less well understood systems, such as the one proposed in Chapter 4,. This system outlines a basis for artificial tissue homeostasis and defines the necessary components that need to be created in order to realize one possible method to create this system. The modules described here attempt to provide answers to some fundamental questions of synthetic biology (i.e. can we predictively compose genetic circuits? Can we create an orthogonal extensible intercellular communication system in mammalian cells?).

The control of cells in a population requires that information be exchanged across cellular boundaries. Hence, a cell-cell communication system is required. Since pattern formation through programmed differentiation of embryonic stem

cells is one application of this research, the communication signal should be capable of establishing concentration gradients and the cells should be able to respond to these gradients in a pre-determined fashion. In Chapter 5, I introduced several novel cell-cell communication system designs adapted for mammalian cell use from systems in organisms ranging from the marine bacterium *Vibrio fischeri* to the plant Arabidopsis thaliana. I ported the bacterial Type II FAS synthesis pathway into mammalian cells by codon optimization and expression of the relevant enzymes in a compact manner using 2A tags. I constructed a virus like particle (VLP) that can carry a load from a sender cell to a receiver cell. In Chapter 5, I also present a design for a two-component system adapted from a hybrid mix of plant and bacteria. I show initial promising results of the response to exogenous IP and processing of that response in a cascade format to invert the signal. I also propose a versatile and promising system using the TEV protease. I have shown that TEV protease cleaves TF's from the cell membrane and activates transcription with in a single cell.

In Chapter 6, I address a fundamental question of synthetic biology; is predictive composition of genetic circuits possible in mammalian cells? Yes! We are able to collect high quality characterization data on numerous genetic parts, analyze and process the data to extract the relevant information (transfer functions), and feed this information into computer models that are able to predict the behavior of novel genetic architectures. I validated these predictions by constructing and testing several genetic cascades in mammalian cells. It establishes a general method for characterizing TF's for use in novel circuits. This research creates a foundation to engineer large complex genetic circuits from well-characterized parts.

7.2. Future Work

The engineering of synthetic inter-cellular communication modules coupled to exogenous intra-cellular transcription networks offers tremendous possibilities to program cells at the population level. The work presented in my thesis presents numerous avenues that can be expanded upon to push the boundaries of synthetic biology. Below I present some areas that have promise for future research.

The PhoB system presented in Chapter 5 appears to work. In order to show that the PhoBVP16 is actually responsible for the response to IP one just needs to tease out the mechanism of action in the cells. Several mutants have already been created based on constitutively active and inactive versions of PhoB [95, 96].

The TEV system is the most promising cell-cell communication system seen thus far. One reason for its attractiveness is because of its expandability to many different channels. The proof of concept of this system has been demonstrated and with a little more characterization I believe that the system will be able to function as designed. The TEV system also could be used in a different manner.

The most attractive cell-cell communication system would be one where the receiver cells did not have to be genetically modified. The sender cells would send messages that could influence the behavior of unmodified cells by actuating on endogenous pathways. The TEV system, with a couple small modifications could become such a communication system. Recently, it has been demonstrated that acylated tagged proteins accumulate at the membrane and get incorporated into exosome vesicles, that are taken up by neighboring cells[97]. If we anchor the TEV constructs to the membrane with an acyl tag instead of the FGFR2 transmembrane

domain the TEV constructs could get incorporated into these vesicles and transmit TF or other proteins to any unmodified receiver cells one desired. This would be a very attractive solution for creating IPS cells where the cells are genetically unmodified.

The future work for the characterization project would be to create huge libraries of characterized parts that could be used in mammalian cells. Further validation of the predictive software we have developed would be needed for more complicated systems. Also, new analytical tools will need to be developed for closed loop systems such as the toggle, where there is feedback, as well as the oscillator. The parts can still be characterized in the same method that has been developed in this thesis. But the analytical and predictive tools will need more work to better understand these more complex systems.

In order to truly have an expansive library of parts, we will need to employ automation. The tedious, time consuming, and repetitive nature of characterization lends itself immediately to full automation. Some places have already begun automating their own characterization processes (i.e. BIOFAB). Currently, in our lab, Jonathan Babb is working on automating every step of the DNA assembly to FACS preparation of the mammalian cells. If he were successful, one would only need to perform the transfections and cell culture work. The BD LSR Fortessa already has a high throughput device for 96 well plates. I began to use this device to collect data towards the end of the characterization work when the number of samples I had for the flow cytometer reached unmanageable levels.

The future of synthetic biology is wide open and with the new tools developed in this thesis, mammalian synthetic biology will be able to grow much faster than before. We will be able to construct large complex genetic circuits that work based on the initial design. The only limit on what we create in the future will be our own imagination.

Materials and Methods

A. Strains and Culture Conditions

A.1. Bacterial Strains and Culture Conditions

E. coli DH10b, F- endA1 recA1 galE15 galK16 nupG rpsL ΔlacX74 Φ80lacZΔM15 araD139 Δ(ara,leu)7697 mcrA Δ(mrr-hsdRMS-mcrBC) λ-, [98] was used to clone and propagate plasmids all plasmids. LB broth (Difco, Detroit, MI) with the appropriate antibiotic(s) was used as a growth medium in all experiments. For the construction of non lentiviral plasmids, 50 μg/mL kanamycin (Shelton Scientific, Shelton, CT) and 100 μg/mL ampicillin (Sigma, St. Louis, MO) were used. For the construction of lentiviral plasmids 100μg/mL ampicillin (Sigma, St. Louis, MO) was used. AHL was added at the specified concentration. The AHL 3-oxohexanoyl-homoserine lactone (3OC6HSL) was acquired from Sigma-Aldrich. For all growth experiments, cultures were incubated at 37 ◦C in a shaker at 250 rpm. All minipreps of plasmid DNA were done using the Qiagen Miniprep kit (Qiagen, MD) and midipreps and maxipreps of DNA were done using the Qiagen midiprep/maxiprep kits (Qiagen,MD).

A.2. Mammalian Culture Conditions

293FT (Invitrogen) human embryonic kidney fibroblasts (stably expressing the SV40 large T antigen) were used for transfection experiments that tested non-lentiviral plasmids and for virus harvest. Polybrene (Sigma) was used at a

concentration of 10 μg for infecting cells. All the cells were grown at 37◦C and 5% CO_2 in a sterile tissue culture incubator. Cell culture media for culturing 293FT cells is composed of 87.9% DMEM (Hyclone), 10% Tet-approved Fetal Bovine Serum (Clontech), 1% Penicillin-Streptomycin (Hyclone), 0.1% Fungin (Invivogen) filtered through a 0.45 mu filter (Nalgene).

B. Selected Protocols

B.1. Lentivirus Production and Infection

Adapted from Coleman et al. Physiol Genomics 2003 12:221-228 by Dr. Christoph Schaniel.

Day 0: 293FT (Invitrogen R70007) cells are plated at a density of $7.8 * 10_6$ per 10cm dish or $1.75 * 10^7$ 293FT cells per 15cm dish. Generally 3 15cm dishes are used. The cells are incubated in culture medium (88.9% DMEM, 10% FBS, 1% Penicillin-Streptomycin, 0.1% Fungin) at 37◦C, 5% $C0_2$.

Day 1: To produce the virus, 293FT cells are first transfected with two packaging plasmids [99] and the lentivirus vector containing the circuit of interest using Superfect transfection reagent (Qiagen). The supernatant media containing the virus is harvested 30h post transfection and an equal amount of growth media is added to the dish. The harvested supernatant is filtered through a 0.22μm low protein binding filter (Nalgene) to remove cellular debris.

Day 2: Supernatant is again collected 45-72h post infection, filtered as before and combined with the supernatant from Day 1. The virus is concentrated by either by ultrafiltration using Centricon Plus-70 100 kDa spin filters (Millipore) or by

ultracentrifugation at 50000 g for 2.5 hours. Following concentration, the virus can be stored at −80◦C without significant loss of viability for several months. If the viral particles are pelleted by ultracentrifugation, the pellet is carefully re-suspended in PBS by shaking overnight at 4◦C and the viral particles are aliquoted (20µl each) and stored at −80◦C.

To determine virus titer 293 HEK FT cells are seeded at a density of $5 * 10^4$ per well of a 24 well plate. Polybrene (Sigma Aldrich 10768-9) is added to culture media so that the final concentration is 10µg. 24 hours after seeding, the culture media of the cells is replaced with media containing Polybrene. Three different serial dilutions of 20µl of virus are made and added to individual wells. The media is replaced with normal cell culture media 24h post infection with virus. The cells are assayed using the FACS 48h post infection and the viral titer is determined. Once the titer is determined, target cells are infected at the desired MOI by using virus and media supplemented with 10µg Polybrene.

B.2. Transfection

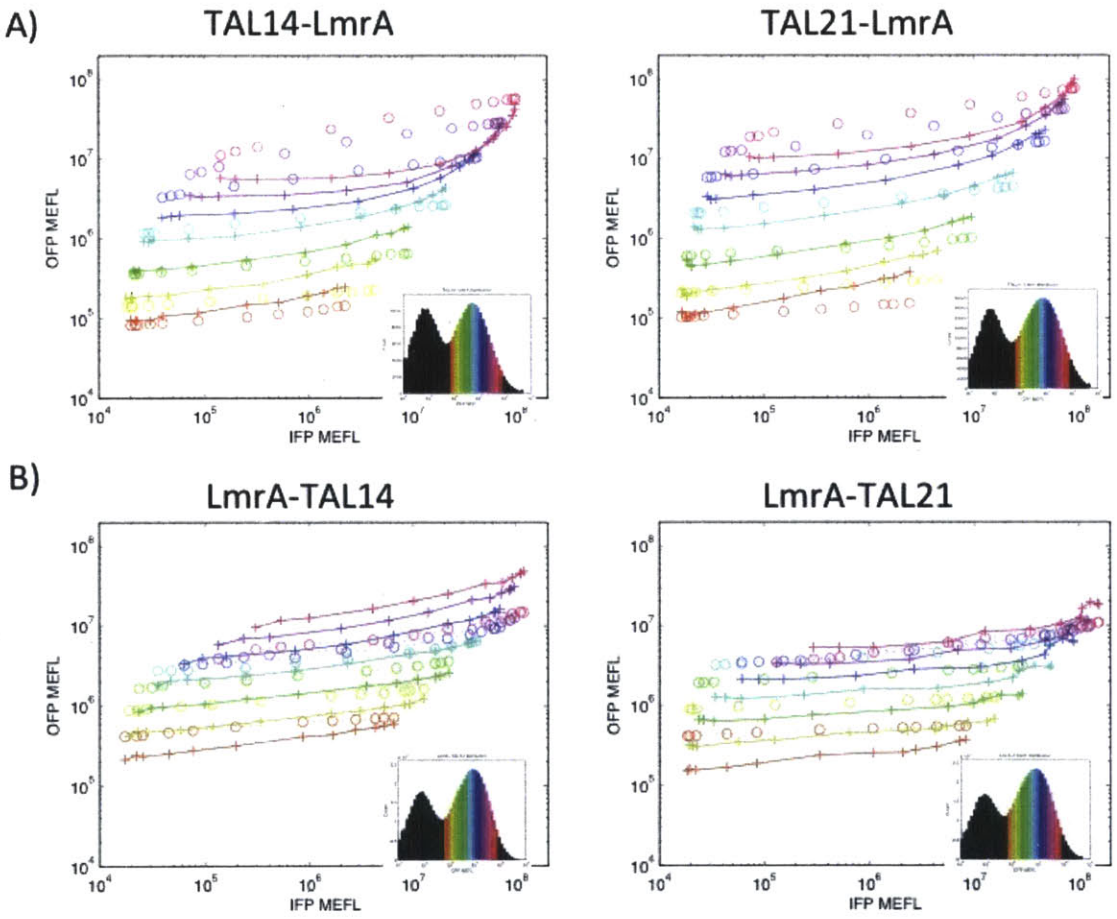
Transfections for characterization or testing purposes were carried out with Metafectene Pro (Biontex). For one well of a 24 well-plate I used 500ng of DNA that is first mixed into 60ul of DMEM (no antibiotics or serum) and then I add 1.5ul of Metafectene Pro and wait for 15 min at room temperature. During this time I coat the well with 0.1% gelatin and split the cells. I seed 0.5 ml of cells at a density of $6 * 10^5$ cells per 1ml media. I then add the transfection mixture to the cells. The media is changed 24 hours post transfection and the cells are assay 72 hours post transfection.

B.3. Western Blot Analysis

Western blot analysis is performed using two techniques. Expression of proteins is assessed by lysing the cells and running the lysate on a 4-20% acrylamide gradient gel. Proteins are transferred using a standard Western blot protocol. The desired proteins are detected with HRP-conjugated antibodies raised against the T7, His, HA, or FLAG tag expressed with the given protein. Detection is performed via a colorimetric assay using TMB stabilized substrate for HRP, or via a luminescence assay using ECL Western blotting detection reagents.

C. Prediction Transfer Curves

The figures below are prediction transfer curves for the 4 remaining cascades that are not shown in the main text. The table described the mean squared error for each of these in the main text in Figure 27c. (Figure courtesy of Jacob Beal)



D. Plasmids

The following table provides an overview of the plasmids constructed for the various systems described in this thesis. Plasmids beginning with pENTR are entry vectors. Plasmid starting with pZDONR are expression vectores. Plasmids beginning

with the initials 'pLV' denote lentiviral vectors. The earlier ones were constructed using basic molecular cloning techniques described in standard molecular biology laboratory manual [100]. The plasmids were constructed by the use of custom designed oligonucleotides and PCR Accuprime Pfx Supermix (Invitrogen)that PCR amplify parent plasmids with overhangs containing the appropriate restriction sites. Restriction enzymes were purchased from New England Biolabs. Synthetic oligonucleotides were ordered from Integrated DNA Technologies. All the cloning was performed with heat shock transformation using DH10b chemical competent cells made in the laboratory. Sequencing reactions were performed at Genewiz Inc. (Cambridge, USA). Synthesis of DNA was done by Geneart (Germany).

1	pENTR_L4_CAGop_R1_ND
2	pENTR_L4_CMV5-CUO_R1-12-20-10
3	pENTR_L4_Hef1a_R2
4	pENTR_L4_Hef1a-5xGal4-5xCuO_R2
5	pENTR_L4_Hef1a-5xCuO_R2
6	pENTR_L4_Hef1a-6xC7_R2
7	pENTR_L4_Hef1a-LacO1Oid-12-11-09_R2
8	pENTR_L4_Hef1a-TAL1x2_R2
9	pENTR_L4_Hef1a-TAL4x2_R2
10	pENTR_L4_TRE_R1
11	pENTR_L4_TRE-LacO1Oid-59.5_R1
12	pENTR_L4_TRE-LacO1Oid-59.5_R1
13	pENTR_L4_TRE-LacO1Oid-81.5_R1
14	pENTR_L4_TRE-LacO1Oid-81.5_R1
15	pENTR_L4_Gal4UAS_R2
16	pENTR_L4_Hef1a-CI434x2_R1
17	pENTR_L4_minCMV-1xCI434_R1
18	pENTR_L4_minCMV-4xCI434_R1
19	pENTR_L4_Hef1a-LexAx3_R1
20	pENTR_L4_minCMV-1xLexA_R1
21	pENTR_L4_minCMV-4xLexA_R1

22	pENTR_L4_Hef1a-Mnt1x2_R1
23	pENTR_L4_minCMV-1xMntI_R1
24	pENTR_L4_minCMV-4xMntI_R1
25	pENTR_L4_minCMV-7xMntI_R1
26	pENTR_L4_minCMV-1xPho_R1
27	pENTR_L4_minCMV-4xPho_R1
28	pENTR_L4_minCMV5xPho_R1
29	pENTR_L4_minCMV6xPho_R1
30	pENTR_L4_minCMV-7xPho_R1
31	pENTR_ L1_AHK4-2A-AHP5_L2
32	pENTR_ L1_AHP2_L2
33	pENTR_L1_AmCyan-mirFF4-2A-LacI_L2
34	pENTR_L1_AtiPt4_L2
35	pENTR_L1_C7C7NLSKRAB-2A-mKate_L2
36	pENTR_L1_Ca-rtta-Flag_L2
37	pENTR_L1_CI434_L2
38	pENTR_L1_CI434VP16_L2
39	pENTR_L1_CymR_L2
40	pENTR_L1_DD-EBFP2_L2
41	pENTR_L1_DD-EYFP-4xFF4_L2
42	pENTR_L1_DD-mkate_L2
43	pENTR_L1_DD-rtTa_L2
44	pENTR_L1_DD-rtTa3-4xFF4_L2
45	pENTR_L1_DD-rtTa-4xFF4_L2
46	pENTR_L1_DD-VP16Gal4_L2
47	pENTR_L1_FGFR2-TCS-VP16Gal4_L2
48	pENTR_L1_FGFR2-TCS-VP16Gal4-mkate_L2
49	pENTR_L1_FGFR2-Tev_L2
50	pENTR_L1_FGFR2-Tev-EBFP2_L2
51	pENTR_L1_FRB-GS-Tev-L2
52	pENTR_L1_HsPum1-NPM1_L2
53	pENTR_L1_NPM1-HsPum1_L2
54	pENTR_L1_LacI_L2
55	pENTR_L1_LacIKRAB_L2
56	pENTR_L1_LacIKRAB_L2
57	pENTR_L1_LacIKRAB-2A-mkate_L2
58	pENTR_L1_LacI-mkate-mirFF4_L2
59	pENTR_L1_LexA-mirFF6_L2
60	pENTR_L1_LexAVP16_L2
61	pENTR_L1_mKate-PEST-D1_L2
62	pENTR_L1_Mnt1_L2

63	pENTR_L1_Mnt1-mirFF6_L2
64	pENTR_L1_Mnt1VP16_L2
65	pENTR_L1_Myr-mkate-TCS-VP16Gal4_L2
66	pENTR_L1_Nef7-MACA-rtTa-Flag_L2
67	pENTR_L1_Nef7-MACA-VP16Gal4-Flag_L2
68	pENTR_L1_Nef7-VP16Gal4-Flag_L2
69	pENTR_L1_PhoB-GS-VP16-_L2
70	pENTR_L1_PhoB-GS-VP16-mkate_L2
71	pENTR_L1_rtTa-D1_L2
72	pENTR_L1_rtTa-D2_L2
73	pENTR_L1_rtTa-D4_L2
74	pENTR_L1_rtta-Flag_L2
75	pENTR_L1_ss-Myc-H2-TM-Tev_L2 (SbfI,NotI) (new new tev)
76	pENTR_L1_ss-Myc-NZIP-H2-TM-rtTa3_L2 6-10-10(SbfI but no NotI)
77	pENTR_L1_ss-Myc-NZIP-H2-TM-Tev_L2 6-10-10(SbfI but no NotI)
78	pENTR_L1_ss-Myc-H2-TM-Tev_L2 (New Tev)
79	pENTR_L1_TEV_L2 (no NotI)
80	pENTR_L1_TEV_L2
81	pENTR_L1_CymR-2A-puro_L2
82	pENTR_L1_EYFP-4xFF4_L2
83	pENTR_L1_EYFP-2A-Hygro-4xFF5_L2
84	pENTR_L1_EGFP-D1_L2
85	pENTR_L1_EGFP-D1bis_L2
86	pENTR_L1_EGFP-D2_L2
87	pENTR_L1_EGFP-D4_L2
88	pEXPR_1-GTW-2_pENTR_L4_CAG_R1|pENTR_L1_rtTA3-2A-Hygro_L2
89	pEXPR_1-GTW-2_pENTR_L4_EGSH_R1|pENTR_L1_AmCyan-mirFF4-2A-LacI_L2
90	pEXPR_1-GTW-2_pENTR_L4_EGSH_R1|pENTR_L1_C7C7NLSKRAB-2A-mKate_L2
91	pEXPR_1-GTW-2_pENTR_L4_EGSH_R1|pENTR_L1_LacI-FF5-miRFF4_L2
92	pEXPR_1-GTW-2_pENTR_L4_EGSH_R1|pENTR_L1_rtTa-D1_L2
93	pEXPR_1-GTW-2_pENTR_L4_EGSH_R1|pENTR_L1_rtTa-D2_L2
94	pEXPR_1-GTW-2_pENTR_L4_EGSH_R1|pENTR_L1_rtTa-D4_L2
95	pZDonor 1-GTW-2|pENTR_L4_EGSH_R1|pENTR_L1_rtTA3-2A-Hygro_L2
96	pZDonor 1-GTW-2|pENTR_L4_EGSH_R1|pENTR_L1_DD_rtTa_L2
97	pZDonor 1-GTW-2|pENTR_L4_UbC_R1|pENTR_L1_FGFR2-TCS-VP16Gal4_L2
98	pZDonor 1-GTW-2|pENTR_L4_UbC_R1|pENTR_L1_Myr-mkate-TCS-VP16Gal4_L2
99	pZDonor 2-GTW-3|pENTR_L4_EGSH_R1|pENTR_L1_AmCyan-L2
100	pZDonor 2-GTW-3|pENTR_L4_EGSH_R1|pENTR_L1_Cer_L2
101	pZDonor 2-GTW-3|pENTR_L4_EGSH_R1|pENTR_L1_CymR_L2
102	pZDonor 2-GTW-3|pENTR_L4_EGSH_R1|pENTR_L1_mKate_L2
103	pZDonor 2-GTW-3|pENTR_L4_Gal4UAS-11-16-09_R2|pENTR_L1_EGFP_L2

104	pZDonor 2-GTW-3|pENTR_L4_TRE-tight_R1|pENTR_ L1_AHK4-2A-AHP5_L2
105	pZDonor 2-GTW-3|pENTR_L4_TRE-tight_R1|pENTR_L1_CymR_L2
106	pZDonor 2-GTW-3|pENTR_L4_TRE-tight_R1|pENTR_L1_EGFP-PestD1_L2
107	pZDonor 2-GTW-3|pENTR_L4_TRE-tight_R1|pENTR_L1_EGFP-PestD1bis_L2
108	pZDonor 2-GTW-3|pENTR_L4_TRE-tight_R1|pENTR_L1_EGFP-PestD2_L2
109	pZDonor 2-GTW-3|pENTR_L4_TRE-tight_R1|pENTR_L1_EGFP-PestD4_L2
110	pZDonor 2-GTW-3|pENTR_L4_TRE-tight_R1|pENTR_L1_LacI-FF5-miRFF4_L2
111	pZDonor 2-GTW-3|pENTR_L4_TRE-tight_R1|pENTR_L1_mKate_L2
112	pZDonor 2-GTW-3|pENTR_L4_Tre-tight_R1|pENTR_L1_VP16Gal4_L2
113	pZDonor 2-GTW-3|pENTR_L4_UbC_R1|pENTR_L1_Myr-FKBPx2-TCS-mkate-VP16gal4_L2
114	pZDonor 3-GTW-4|pENTR_L4_CMV5-CUO_R1-12-20-10|pENTR_L1_AmCyan-mirFF4-2A-LacI_L2
115	pZDonor 3-GTW-4|pENTR_L4_CMV5-CUO_R1-12-20-10|pENTR_L1_EYFP-FF4x4_L2
116	pZDonor 3-GTW-4|pENTR_L4_CMV5-CUO_R1-12-20-10|pENTR_L1_LacI-FF5-miRFF4_L2
117	pZDonor 3-GTW-4|pENTR_L4_EGSH_R1|pENTR_L1_AmCyan-L2
118	pZDonor 3-GTW-4|pENTR_L4_Hef1a-6xC7_R2|pENTR_L1_EYFP_L2
119	pZDonor 3-GTW-4|pENTR_L4_Hef1a-LacO1Oid-12-11-09_R2|pENTR_L1_EYFP-FF4x4-L2
120	pZDonor 3-GTW-4|pENTR_L4_Hef1a_R1|pENTR_L1_FGFR2-Tev_L2
121	pZDonor 3-GTW-4|pENTR_L4_Hef1a_R1|pENTR_L1_mKate_L2
122	pZDonor 3-GTW-4|pENTR_L4_Hef1a_R1|pENTR_L1_Puro_L2
123	pZDonor 3-GTW-4|pENTR_L4_minCMV-4xPho_R1|pENTR_L1_EYFP-FF4x4-L2
124	pZDonor 3-GTW-4|pENTR_L4_minCMV5xPho_R1|pENTR_L1_EYFP-FF4x4-L2
125	pZDonor 3-GTW-4|pENTR_L4_minCMV6xPho_R1|pENTR_L1_EYFP-FF4x4-L2
126	pZDonor 3-GTW-4|pENTR_L4_TRE-tight_R1|pENTR_L1_EYFP-FF4x4_L2
127	pZDonor 3-GTW-4|pENTR_L4_TRE-LacO1Oid-59.5_R1|pENTR_L1_EYFP-FF4x4-L2
128	pZDonor 3-GTW-4|pENTR_L4_TRE-LacO1Oid-81.5_R1|pENTR_L1_EYFP-FF4x4-L2
129	pZDonor 3-GTW-4|pENTR_L4_UASgal4_R1|pENTR_L1_EYFP-FF4x4_L2
130	pZDonor 3-GTW-4|pENTR_L4_UbC_R1|pENTR_L1_TEV_L2
131	pZDonor 4-GTW-5|pENTR_L4_CAG_R1|pENTR_L1_VgEcR-2A-RXR_L2
132	pZDonor 4-GTW-5|pENTR_L4_Hef1a_R1|pENTR_ L1_PhoB-GS-VP16_L2
133	pZDonor 4-GTW-5|pENTR_L4_Hef1a_R1|pENTR_L1_mKate_L2
134	pZDonor 4-GTW-5|pENTR_L4_Hef1a_R1|pENTR_L1_Puro_L2
135	pZDonor 4-GTW-5|pENTR_L4_TRE-tight_R1|pENTR_ L1_PhoB-GS-VP16_L2
136	pZDonor 4-GTW-5|pENTR_L4_TRE-tight_R1|pENTR_L1_AmCyan-L2
137	pZDonor 4-GTW-5|pENTR_L4_TRE-tight_R1|pENTR_L1_Cerulean_L2
138	pZDonor 4-GTW-5|pENTR_L4_UbC_R1|pENTR_L1_VgEcR-2A-RXR_L2
139	pZDonor 5-GTW-6|pENTR_L4_CAG_R1|pENTR_L1_AmCyan-L2
140	pZDonor 5-GTW-6|pENTR_L4_CAG_R1|pENTR_L1_Cerulean_L2
141	pZDonor 5-GTW-6|pENTR_L4_CAG_R1|pENTR_L1_mKate_L2
142	pZDonor 5-GTW-6|pENTR_L4_Hef1a_R1|pENTR_L1_Puro_L2
143	pZDonor 6-GTW-7|pENTR_L4_Hef1a_R1|pENTR_L1_LacI-FF5-miRFF4_L2

144	pZDonor 6-GTW-7|pENTR_L4_Hef1a_R1|pENTR_L1_Puro_L2
145	pZDonor 2-GTW-3|pENTR_L4_Rheo-5xUAS_R1|pENTR-L1-TetR-mirff4-L2
146	pZDonor 1-GTW-2|pENTR_L4_UbC_R1|pENTR_L1_Rheoact-2a-Rheoswitch-L2
147	pZDonor 3-GTW-4|pENTR-L4_Hef1a-tetO2-R1|pENTR_L1_EYFP-FF4x4-L2
148	pZDonor 4-GTW-5|pENTR_L4_Rheo-5xUAS_R1|pENTR_L1_AmCyan-L2
149	pZDonor 2-GTW-3|pENTR_L4_UASrheo5x_R1|pENTR_L1_rtTa-D1_L2
150	pZDonor 2-GTW-3|pENTR_L4_UASrheo5x_R1|pENTR_L1_rtTa-D2_L2
151	pZDonor 2-GTW-3|pENTR_L4_UASrheo5x_R1|pENTR_L1_rtTa-D4_L2
152	pZDonor 2-GTW-3|pENTR_L4_UASrheo5x_R1|pENTR_L1_rtTA3-2A-Hygro_L2
153	pZDonor 2-GTW-3|pENTR_L4_UASrheo5x_R1|pENTR_L1_DD_rtTa_L2
154	pZDonor 2-GTW-3|pENTR_L4_UbC_R1|pENTR_L1_FRB-GS-Tev-L2
155	pZDonor 2-GTW-3r|pEntr_L4_5xUAS_Rheo_R1|TetR-mirff6
156	pZDonor 3-GTW-4r|pENTR_L4_Hef1a-LacO1Oid-12-11-09_R2|pENTR_L1_EYFP-FF6x4-L2
157	pZDonor 2-GTW-3|pENTR_L4_TRE-tight_R1|pENTR_L1_CI434_L2
158	pZDonor 2-GTW-3|pENTR_L4_TRE-tight_R1|pENTR_L1_LexAVP16_L2-version2
159	pZDonor 2-GTW-3|pENTR_L4_TRE-tight_R1|pENTR_L1_Mnt1_L2
160	pZDonor 3-GTW-4|pENTR_L4_Hef1a-CI434x2_R1|pENTR_L1_EYFP-2A-Hygro-4xFF5_L2
161	pZDonor 3-GTW-4|pENTR_L4_Hef1a-LexAx3_R1|pENTR_L1_EYFP-2A-Hygro-4xFF5_L2
162	pZDonor 3-GTW-4|pENTR_L4_Hef1a-Mnt1x2_R1|pENTR_L1_EYFP-2A-Hygro-4xFF5_L2
163	pZDonor 4-GTW-5|pENTR_L4_Hef1a_R1|pENTR_L1_EBFP2_L2
164	pZDonor 4-GTW-5|pENTR_L4_TRE-tight_R1|pENTR_L1_EBFP2_L2
165	pZDonor 4-GTW-5|pENTR_L4_TRE-tight_R1|pENTR_L1_mKate_L2
166	pZDonor 4-GTW-5|pENTR_L4_UASrheo5x_R1|pENTR_L1_EBFP2_L2
167	pZDonor 5-GTW-6-Hef1a-EBFP2
168	pZDonor 6-GTW-7|pENTR_L4_Hef1a_R1|pENTR_L1_EBFP2_L2
169	pZDonor 1-GTW-2|pENTR_L4_Hef1a_R1|pENTR_L1_rtTA3-L2
170	pZDonor 2-GTW-3|pENTR_L4_minCMV6xPho_R1|pENTR_L1_EBFP2_L2
171	pZDonor 2-GTW-3|pENTR_L4_UASrheo5x_R1|pENTR_ L1_AHK4-2A-AHP5_L2
172	pZDonor 3-GTW-4|pENTR_L4_minCMV6xPho_R1|pENTR_L1_LacI-FF5-miRFF4_L2
173	pZDonor 3-GTW-4|pENTR_L4_minCMV6xPho_R1|pENTR_L1_rtTA3-L2
174	pZDonor 3-GTW-4|pENTR_L4_TRE-tight_R1|pENTR_L1_EYFP_L2
175	pZDonor 4-GTW-5|pENTR_L4_attB[BxB1]_R1|pENTR_L1_Bleo_L2
176	pZDonor 5-GTW-6|pENTR_L4_Hef1a-LacO1Oid-12-11-09_R2|pENTR_L1_rtTA3-FF4x4-L2
177	pZDonor 5-GTW-6|pENTR_L4_TRE-tight_R1|pENTR_L1_LacI-FF5-miRFF4_L2
178	pZDonor 6-GTW-7|pENTR_L4_minCMV6xPho_R1|pENTR_L1_EBFP2_L2
179	pZDonor 6-GTW-7|pENTR_L4_minCMV6xPho_R1|pENTR_L1_LacI-FF5-miRFF4_L2
180	pZDonor 7-GTW-8|pENTR_L4_minCMV6xPho_R1|pENTR_L1_EBFP2_L2
181	pZDonor 7-GTW-8|pENTR_L4_minCMV6xPho_R1|pENTR_L1_LacI-FF5-miRFF4_L2
182	pZDonor 7-GTW-8|pENTR_L4_minCMV6xPho_R1|pENTR_L1_rtTA3-L2
183	pZDonor 1-GTW-2|pENTR_L4_Hef1a_R1|pENTR_L1_AtiPt4_L2
184	pZDonor 3-GTW-4|pENTR_L4_CAGop_R1_ND|pENTR_L1_EYFP-FF4x4_L2

185	pZDonor 3-GTW-4|pENTR_L4_Hef1a-TAL1x2_R1|pENTR_L1_EYFP-2A-Hygro-4xFF5_L2
186	pZDonor 3-GTW-4|pENTR_L4_Hef1a-TAL4x2_R1|pENTR_L1_EYFP-2A-Hygro-4xFF5_L2
187	pZDonor 3-GTW-4|pENTR_L4_Hef1a_R1|pENTR_L1_HsPum1-NPM1_L2
188	pZDonor 3-GTW-4|pENTR_L4_Hef1a_R1|pENTR_L1_NPM1-HsPum1_L2
189	pZDonor 4-GTW-5|pENTR_L4_TRE-tight_R1|pENTR_L1_PhoB-GS-VP16-mkate_L2
190	pZDonor 5-GTW-6|pENTR_L4_UASgal4_R1|pENTR_L1_LacI-FF5-miRFF4_L2
191	pZDonor 1-GTW-2|pENTR_L4_UAS-T1T1-72bp_R1|pENTR_L1_DD-VP16Gal4_L2
192	pZDonor 1-GTW-2|pENTR_L4_TRE-LacO1Oid-81.5_R1|pENTR_L1_DD-rtTa-4xFF4_L2
193	pZDonor 1-GTW-2|pENTR_L4_TRE-LacO1Oid-59.5_R1|pENTR_L1_DD-rtTa-4xFF4_L2
194	pZDonor 1-GTW-2|pENTR_L4_TRE-LacO1Oid-81.5_R1|pENTR_L1_DD-rtTa3-4xFF4_L2
195	pZDonor 1-GTW-2|pENTR_L4_TRE-LacO1Oid-59.5_R1|pENTR_L1_DD-rtTa3-4xFF4_L2
196	pZDonor 2-GTW-3|pENTR_L4_UASrheo5x_R1|pENTR_L1_rtTA3-L2
197	pZDonor 2-GTW-3|pENTR_L4_UASgal4_R1|pENTR_L1-Kozak-EYFP-2A-TAL1-L2
198	pZDonor 2-GTW-3|pENTR_L4_TRE-tight_R1|pENTR_L1_LacI-mkate-mirFF4_L2
199	pZDonor 2-GTW-3|pENTR_L4_UASgal4_R1|pENTR_L1_LacI-mkate-mirFF4_L2
200	pZDonor 3-GTW-4|pENTR_L4_UAS-T1T1-72bp_R1|pENTR_L1_DD-mkate_L2
201	pZDonor 3-GTW-4|pENTR_L4_UAS-T1T1-72bp_R1|pENTR_L1_DD-EBFP2_L2
202	pZDonor 3-GTW-4|pENTR_L4_TRE-LacO1Oid-81.5_R1|pENTR_L1_DD-EYFP-4xFF4_L2
203	pZDonor 3-GTW-4|pENTR_L4_TRE-LacO1Oid-59.5_R1|pENTR_L1_DD-EYFP-4xFF4_L2
204	pZDonor 3-GTW-4|pENTR_L4_UAS-T1T1-72bp_R1|pENTR_L1_LacI-FF5-miRFF4_L2
205	pZDonor 4-GTW-5|pENTR_L4_TRE-tight_R1|pENTR_L1_DD-VP16Gal4_L2
206	pZDonor 4-GTW-5|pENTR_L4_TRE-tight_R1|pENTR_L1_mKate_L2
207	pZDonor 4-GTW-5|pENTR_L4_TRE-tight_R1|pENTR_L1_DD-EBFP2_L2
208	pZDonor 4-GTW-5|pENTR_L4_TRE-tight_R1|pENTR_L1_VP16Gal4_L2
209	pZDonor 4-GTW-5|pENTR_L4_UASgal4_R1|pENTR_L1_rtTA3-L2
210	pZDonor 5-GTW-6|pENTR_L4_UASgal4_R1|pENTR_L1_LacI-mkate-mirFF4_L2
211	pZDonor 5-GTW-6|pENTR_L4_Hef1a-LacO1Oid-12-11-09_R2|pENTR_L1_mKate-FF5-FF4_L2
212	pZDonor 5-GTW-6|pENTR_L4_TRE-tight_R1|pENTR_L1_DD-EBFP2_L2
213	pZDonor 5-GTW-6|pENTR_L4_TRE-tight_R1|pENTR_L1_EBFP2_L2
214	pZDonor 5-GTW-6|pENTR_L4_UASgal4_R1|pENTR_L1_EBFP2_L2
215	pZDonor 6-GTW-7|pENTR_L4_UASgal4_R1|pENTR_L1_mKate_L2
216	pLV_R4R2_GTW3|pENTR_L4_CMV5-CUO_R1-12-20-10|pENTR_L1_EYFP-FF6x4-L2
217	pZDonor_pEXPR_4-5_TRE-EYFP-2A-TAL1
218	pZDonor_pEXPR_2-3_TagBFP-2A-VP16Gal4
219	pZDonor 1-GTW-2|pENTR_L4_Hef1a_R1|pENTR_L1_rtTA3-2A-NLS-mKate2_L2
220	pZDonor 1-GTW-2|pENTR_L4_TRE-tight_R1|pENTR_L1_EYFP-2A-TAL1-L2
221	pZDonor 2-GTW-3|pENTR_L4_UASrheo5x_R1|pENTR_L1_DD-rtTa3-4xFF4_L2
222	pZDonor 2-GTW-3|pENTR_L4_UAS-T1T1-72bp_R1|pENTR_L1_LacI-FF5-miRFF4_L2
223	pZDonor 2-GTW-3|pENTR_L4_TRE-tight_R1|pENTR_L1_EBFP2_L2
224	pZDonor 3-GTW-4|pENTR_L4_CAG_R1|pENTR_L1_mKate-2A-Puro_L2
225	pZDonor 3-GTW-4|pENTR_L4_Hef1a_R1|pENTR_L1_rtTA3-2A-NLS-mKate2_L2

226	pZDonor 4-GTW-5|pENTR_L4_Hef1a_R1|pENTR_L1_VP16Gal4_L2
227	pZDonor 5-GTW-6|pENTR_L4_Hef1a_R1|pENTR_L1_mKate-2A-Puro_L2
228	pZDonor 5-GTW-6|pENTR_L4_Hef1a-LacO1Oid-12-11-09_R2|pENTR_L1_DD-EBFP2_L2
229	pZDonor 7-GTW-8|pENTR_L4_Hef1a_R1|pENTR_L1_EBFP2_L2
230	pZDonor 7-GTW-8|pENTR_L4_Hef1a_R1|pENTR_L1_Puro_L2
231	pZDonor 4-GTW-5|pENTR_L4_CAG_R1|pENTR_L1_rtTA3-2A-NLS-mKate2_L2
232	pZDonor 5-GTW-6|pENTR_L4_CAG_R1|pENTR_L1_rtTA3-2A-NLS-mKate2_L2
233	pZDonor 2-GTW-3|pENTR_L4_UASrheo5x_R1|pENTR_L1_LacI-FF5-miRFF4_L2
234	pZDonor 1-GTW-2|pENTR_L4_TRE-tight_R1|pENTR_L1_LacI-T2A-EBFP2-miRFF4_L2
235	pZDonor 2-GTW-3|pENTR_L4_Hef1a-LacO1Oid-12-11-09_R2|pENTR_L1_EYFP-FF4x4_L2
236	pZDonor 2-GTW-3|pENTR_L4_Hef1a_R1|pENTR_L1_VP16Gal4_L2
237	pZDonor 2-GTW-3|pENTR_L4_TRE-LacO1Oid-81.5_R1|pENTR_L1_EYFP-FF4x4_L2
238	pZDonor 2-GTW-3|pENTR_L4_TRE-tight_R1|pENTR_L1_DD-VP16Gal4_L2
239	pZDonor 2-GTW-3|pENTR_L4_UAS-T1T1-72bp_R1|pENTR_L1_mKate-FF5-FF4_L2
240	pZDonor 3-GTW-4|pENTR_L4_CAG_R1|pENTR_L1_rtTa3-T2A-VP16Gal4_L2
241	pZDonor 3-GTW-4|pENTR_L4_Hef1a-LacO1Oid-12-11-09_R2|pENTR_L1_mKate-FF5-FF4_L2
242	pZDonor 3-GTW-4|pENTR_L4_Hef1a_R1|pENTR_L1_RheoAct-2A-Rec-FF3_L2
243	pZDonor 3-GTW-4|pENTR_L4_TRE-LacO1Oid-81.5_R1|pENTR_L1_mKate-FF5-FF4_L2
244	pZDonor 4-GTW-5|pENTR_L4_CAG_R1|pENTR_L1_rtTA3-2A-Hygro_L2
245	pZDonor 4-GTW-5|pENTR_L4_CAG_R1|pENTR_L1_rtTa3-T2A-VP16Gal4_L2
246	pZDonor 4-GTW-5|pENTR_L4_Hef1a_R1|pENTR_L1_RheoAct-2A-Rec-FF3_L2
247	pZDonor 4-GTW-5|pENTR_L4_UASrheo5x_R1|pENTR_L1_DD-EBFP2_L2
248	pZDonor 5-GTW-6|pENTR_L4_CAG_R1|pENTR_L1_VP16Gal4-T2A-EBFP2_L2
249	pZDonor 1-GTW-2|pENTR_L4_minCMVLuxO7_R1|pENTR_L1_EYFP-2A-TAL1-L2
250	pZDonor 6-GTW-7|pENTR_L4_Hef1a_R1|pENTR_L1_65H4LuxRFmNLS_L2
251	pZDonor 1-GTW-2-TRE-LacO1Oid-81.5-rtTA3-FF4x4
252	pZDonor 1-GTW-2-UASrheo5x-LacI-FF5-miRFF4
253	pZDonor 1-GTW-2-UASrheo5x-LacI-T2A-EBFP2-FF5-miRFF4
254	pZDonor 4-GTW-5-TRE-LacO1Oid-81.5-DD-rtTa3-4xFF4
255	pZDonor 4-GTW-5-TRE-LacO1Oid-81.5-rtTA3-FF4x4
256	pZDonor 5-GTW-6-TRE-LacO1Oid-81.5-DD-EBFP2
257	pZDonor 5-GTW-6-TRE-LacO1Oid-81.5-EBFP2-4xFF4
258	pZDonor 7-GTW-8-Hef1a-VP16Gal4
259	pZDonor 7-GTW-8-Hef1a-VP16Gal4-T2A-EBFP2
260	pZDonor 5-GTW-6-TRE-LacO1Oid-81.5-mKate-FF5-FF4
261	pZDonor 5-GTW-6-TRE-LacO1Oid-81.5-DD-mkate
262	pZDonor 1-GTW-2-TRE-tight-LacI-FF5-miRFF4
263	pZDonor 2-GTW-3-Hef1a-Bleo
264	pZDonor 2-GTW-3-Hef1a-EBFP2
265	pZDonor 2-GTW-3-UASgal4-LacI-FF5-miRFF4
266	CAG-BxB1 integrase

267	pZDonor 1-GTW-2-Hef1a-DD-EYFP-2A-rtTA3-4xFF4
268	pZDonor 1-GTW-2-Hef1a-DD-rtTa3-NES-4xFF4
269	pZDonor 1-GTW-2-UAS-T1T1-72bp-DD-EBFP2-T2A-VP16Gal4
270	pZDonor 2-GTW-3-UAS-T1T1-72bp-mKate
271	pZDonor 3-GTW-4-CAG-rtTa3-T2A-VP16Gal4
272	pZDonor 3-GTW-4-CAG-rtTa3-T2A-VP16Gal4-E2A-mkate
273	pZDonor 3-GTW-4-Hef1a-65H4LuxRFmNLS
274	pZDonor 3-GTW-4-Hef1a-EBFP2-3xNLS
275	pZDonor 3-GTW-4-Hef1a-LacO1Oid-EYFP-2A-TAL1-4xFF4
276	pZDonor 3-GTW-4-TRE-tight-DD-EBFP2-T2A-VP16Gal4
277	pZDonor 4-GTW-5-CAG-rtTa3-T2A-VP16Gal4
278	pZDonor 4-GTW-5-Hef1a-Bleo
279	pZDonor 4-GTW-5-Hef1a-EBFP2-3xNLS
280	pZDonor 3-GTW-4-CPPT-UASgal4-mir223-destRFP-WPRE from Velia Siciliano
281	pZDonor 1-GTW-2-minCMVtet(TRE)-tTA-IRES2-GFP-WPRE from Velia Siciliano
282	pZDonor 2-GTW-3-minCMVtet(TRE)-mir223-mcherry-WPRE from Velia Siciliano
283	pZDonor 1-GTW-2-TRE-LacO1Oid-81.5-DD-EYFP-2A-rtTA3-4xFF4
284	pZDonor 1-GTW-2-TRE-LacO1Oid-81.5-DD-rtTa3-NES-4xFF4
285	pZDonor 1-GTW-2-UAS-T1T1-72bp-LacI-T2A-EBFP2-miRFF4
286	pZDonor 3-GTW-4-CAGop-EYFP-2A-TAL1-4xFF4
287	pZDonor 3-GTW-4-Hef1a-Bleo
288	pZDonor 4-GTW-5-CAG-rtTa3-T2A-VP16Gal4-E2A-mkate
289	pZDonor 4-GTW-5-Hef1a-LacO1Oid-EYFP-2A-TAL1-4xFF4
290	pZDonor 4-GTW-5-minCMVLoxO7-EYFP-2A-TAL1
291	pZDonor 5-GTW-6-CAG-rtTa3-T2A-VP16Gal4
292	pZDonor 5-GTW-6-CAG-rtTa3-T2A-VP16Gal4-E2A-mkate
293	pZDonor 1-GTW-2-TRE-tight-LacI-mkate-mirFF4
294	pZDonor 2-GTW-3-UAS-T1T1-72bp-EBFP2-4xFF4
295	pZDonor 3-GTW-4-Hef1a-LacO1Oid-TAL1
296	pZDonor 1-GTW-2-minCMVLuxO7-EYFP
297	pZDonor 1-GTW-2-TRE-tight-LmrA-VP16
298	pZDonor 2-GTW-3-CAGop-EYFP-FF4x4
299	pZDonor 2-GTW-3-pminCMV-LmrAx6-EYFP
300	pZDonor 2-GTW-3-UAS-T1T1-72bp-EBFP2
301	pZDonor 2-GTW-3-UAS-T1T1-72bp-EYFP
302	pZDonor 3-GTW-4-CAGop-EBFP2-4xFF4
303	pZDonor 3-GTW-4-CAGop-TAL1
304	pZDonor 3-GTW-4-Hef1a-LacO1Oid-EBFP2-4xFF4
305	pZDonor 3-GTW-4-UAS-LmrAx2-EYFP
306	pZDonor 4-GTW-5-minCMVLuxO7-TAL1
307	pZDonor 5-GTW-6-CAG-EBFP2

308	pZDonor 1-GTW-2-TRE-tight-LmrA
309	pZDonor 1-GTW-2-Hef1a-LmrA
310	pZDonor 1-GTW-2-Hef1a-LmrAVP16
311	CMV-iRFP
312	pZDonor 1-GTW-2-TRE-tight-EBFP2-T2A-LacI-FF5-mirFF4
313	pZDonor 1-GTW-2-UASrheo5x-DD-rtTa3-NES-4xFF4
314	pZDonor 5-GTW-6-Hef1a-LmrA-VP16
315	pZDonor 1-GTW-2-TRE-tight-LacI-FF5-miRFF4
316	pZDonor 1-GTW-2-TRE-tight-TAL1
317	pZDonor 2-GTW-3-UAS-LmrAx2-TAL1
318	pZDonor 2-GTW-3-UAS-T1T1-72bp-LmrA
319	pZDonor 3-GTW-4-UAS-LmrAx2-EBFP2
320	pZDonor 3-GTW-4-UAS-LmrAx2-EYFP
321	pZDonor 3-GTW-4-UAS-LmrAx2-TAL1
322	pZDonor 3-GTW-4-UAS-T1T1-72bp-EYFP
323	pZDonor 3-GTW-4-UAS-T1T1-72bp-LmrA
324	pZDonor 4-GTW-5-UAS-LmrAx2-EYFP
325	pZDonor 4-GTW-5-UAS-T1T1-72bp-EYFP
326	pZDonor 4-GTW-5-pminCMV-LmrAx6-iRFP
327	pZDonor_5'cHS4x2-3-GTW-4-CAGop-TAL1-4xFF4
328	pZDonor_5'cHS4x2-3-GTW-4-Hef1a-LacO1Oid-TAL1-4xFF4
329	CAGop-LmrA-4xFF4
330	Hef1a-LacO1Oid-LmrA-4xFF4
331	TRE-tight-DD-VP16Gal4-NES
332	UAS-LmrAx2-LacI-mirFF4
333	pLV-TRE-IRES2-EGFP
334	pLV-Ubc-IRES2-mCherry(1)
335	pLV-Ubc-IRES2-mCherry(2)
336	pLV-Ubc-VP16Gal4-IRES2-mCherry (1)
337	pLV-Ubc-VP16Gal4-IRES2-mCherry(2)
338	pGal4UAS-IRES2-EGFP
339	pIRES2-Bla
340	pPacI-Hef1a-LacO1Oid-EcoRI
341	pLV-Hef1a-TetO2-NheI-LacIKRAB-IRES2-EGFP
342	pLV-Hef1a-TetO2-NheI-IRES2-DsRed2
343	pLV-Hef1a-LacO1Oid-NheI-IRES2-EGFP
344	pLV-Hef1a-LacO1Oid-TetRKRAB-IRES2-mCherry
345	pLV-Gal4UAS-MyoD-IRES2-EGFP
346	pLV-TRE-Gal4VP16-IRES2-Puro
347	pPacI-Hef1a-CuO-EcoRI
348	pLV-Hef1a-LacO1Oid-CymR-IRES2-mCherry

349	pLV-Hef1a-CuO-LacIKRAB-IRES2-EGFP
350	pGal4UAS-CuOb-IRES2-EGFP
351	pLV-Gal4UAS-CuOb-IRES2-EGFP
352	pLV-Hef1a-LacO1Oid-CymR-IRES2-Bla NON functional
353	pLV-Hef1a-LacO1Oid-CymRKRAB-IRES2-mCherry NON functional
354	pLV-Hef1a-TetO2-Gal4VP16-IRES2-DsRed2
355	pLV-Ubc-LacIKRAB-IRES2-Hygro
356	pAD-CMV5-CymR
357	pAD-CMV5-CuOg-LacZ
358	pTagFP635C (CMV-mKate with C-terminus MCS)
359	pTagFP635N (CMV-mKate with N-terminus MCS)
360	pLV-minCMVLux07-rtTA-IRES2-DsRed2
361	pLV-Ubc-CymR-Hef1a-Cer-2A-Puro
362	pLV-minCMVLux07-IRES2-EGFP
363	pPG6600
364	pIRESBLEO3 midi
365	pLV-Hef1a-LacO1Oid-NheI-IRES2-EGFP
366	pLV-Hef1a-TetO2-NheI-IRES2-DsRed2
367	pLV-Tre-NgnIEYFP-2A-mkate-Ubc-Puro
368	pLV-Ubc-p65H4LuxRFmNLS-IRES2-Hygro
369	pLV-Hef1a-rtTa-his-IRES2-puro
370	#148
371	#149
372	pAD-CMV5-CuOs-LacZ
373	pAD-CMV5-CuOg-LacZ
374	pAD-CMV5-CymR
375	pAD-CMV5-cta
376	pAD-CMV5-rcta
377	pAD-CR5-LacZ
378	Nef7 from gene art
379	pIRESBLEO3 from clontech
380	pACT-MyoD
381	pBIND-id
382	pGL4.31
383	Nef7 midi
384	pLEIGW
385	PMT413
386	PMT1002
387	pBIND-id midi
388	pACT midi
389	pGL4.31 midi

References

1. Subramanian, S., *Synthetic Gene Networks for Differentiation of Mouse Embryonic Stem Cells*, in *Electrical Engineering*2008, Princeton University: Princeton NJ. p. 154.
2. Karig, D. and R. Weiss, *Signal-amplifying genetic circuit enables in vivo observation of weak promoter activation in the Rhl quorum sensing system.* Biotechnol Bioeng, 2005. **89**(6): p. 709-18.
3. Timothy S. Gardner, C.R.C.J.J.C., *Construction of a genetic toggle switch in Escherichia coli.* Letters To Nature, 2000. **403**: p. 339-342.
4. Elowitz, M.B. and S. Leibler, *A synthetic oscillatory network of transcriptional regulators.* Nature, 2000. **403**(6767): p. 335-8.
5. Hooshangi, S., S. Thiberge, and R. Weiss, *Ultrasensitivity and noise propagation in a synthetic transcriptional cascade.* Proc Natl Acad Sci U S A, 2005. **102**(10): p. 3581-6.
6. Basu, S., et al., *A synthetic multicellular system for programmed pattern formation.* Nature, 2005. **434**(7037): p. 1130-4.
7. Tom Knight, R.R., Leon Chan, Drew Endy, Reshma Shetty, Austin Che, *Idempotent Vector Design for the Standard Assembly of Biobricks*, in *MIT CSAIL*2002, Massachusetts Institute of Technology: Cambridge MA.
8. Gibson, D.G., et al., *Enzymatic assembly of DNA molecules up to several hundred kilobases.* Nat Methods, 2009. **6**(5): p. 343-5.

9. Barnea, G., et al., *The genetic design of signaling cascades to record receptor activation.* Proc Natl Acad Sci U S A, 2008. **105**(1): p. 64-9.

10. Antunes, M.S., et al., *Engineering key components in a synthetic eukaryotic signal transduction pathway.* Mol Syst Biol, 2009. **5**: p. 270.

11. Kramer, B.P., W. Weber, and M. Fussenegger, *Artificial regulatory networks and cascades for discrete multilevel transgene control in mammalian cells.* Biotechnol Bioeng, 2003. **83**(7): p. 810-20.

12. Bellefroid, E.J., et al., *The evolutionarily conserved Kruppel-associated box domain defines a subfamily of eukaryotic multifingered proteins.* Proc Natl Acad Sci U S A, 1991. **88**(9): p. 3608-12.

13. Kramer, B.P., et al., *An engineered epigenetic transgene switch in mammalian cells.* Nat Biotechnol, 2004. **22**(7): p. 867-70.

14. Cho, K.H., et al., *Reverse engineering of gene regulatory networks.* IET Syst Biol, 2007. **1**(3): p. 149-63.

15. Lutz, R. and H. Bujard, *Independent and tight regulation of transcriptional units in Escherichia coli via the LacR/O, the TetR/O and AraC/I1-I2 regulatory elements.* Nucleic Acids Res, 1997. **25**(6): p. 1203-10.

16. Stricker, J., et al., *A fast, robust and tunable synthetic gene oscillator.* Nature, 2008. **456**(7221): p. 516-9.

17. Tigges, M., et al., *A tunable synthetic mammalian oscillator.* Nature, 2009. **457**(7227): p. 309-12.

18. Canton, B., A. Labno, and D. Endy, *Refinement and standardization of synthetic biological parts and devices.* Nat Biotechnol, 2008. **26**(7): p. 787-93.

19. Drew Endy, A.A., *BIOFAB: International open facility advancing biotechnology*, 2012.

20. Ellis, T., X. Wang, and J.J. Collins, *Diversity-based, model-guided construction of synthetic gene networks with predicted functions.* Nat Biotechnol, 2009. **27**(5): p. 465-71.

21. Rosenfeld, N., et al., *Accurate prediction of gene feedback circuit behavior from component properties.* Mol Syst Biol, 2007. **3**: p. 143.

22. Wang, B., et al., *Engineering modular and orthogonal genetic logic gates for robust digital-like synthetic biology.* Nat Commun, 2011. **2**: p. 508.

23. Weber, W. and M. Fussenegger, *Inducible product gene expression technology tailored to bioprocess engineering.* Curr Opin Biotechnol, 2007. **18**(5): p. 399-410.

24. Weber, W., et al., *Gas-inducible transgene expression in mammalian cells and mice.* Nat Biotechnol, 2004. **22**(11): p. 1440-4.

25. Chen, M.T. and R. Weiss, *Artificial cell-cell communication in yeast Saccharomyces cerevisiae using signaling elements from Arabidopsis thaliana.* Nat Biotechnol, 2005. **23**(12): p. 1551-5.

26. Morey, K.J., et al., *Developing a synthetic signal transduction system in plants.* Methods Enzymol, 2011. **497**: p. 581-602.

27. Anderson, J.C., et al., *BglBricks: A flexible standard for biological part assembly.* J Biol Eng, 2010. **4**(1): p. 1.

28. Li, Y., *A Novel Method for Mammalian Large Genetic Circuit Assembly and Delivery*, in *Electrical Engineering and Computer Science*2012, Massachusetts Institute of Technology: Cambridge MA. p. 75.

29. A.M. James Shapiro, M.D., Ph.D., Camillo Ricordi, M.D., Bernhard J. Hering, M.D., Hugh Auchincloss, M.D., Robert Lindblad, M.D., R. Paul Robertson, M.D., Antonio Secchi, M.D., Mathias D. Brendel, M.D., Thierry Berney, M.D., Daniel C. Brennan, M.D., Enrico Cagliero, M.D., Rodolfo Alejandro, M.D., Edmond A. Ryan, M.D., Barbara DiMercurio, R.N., Philippe Morel, M.D., Kenneth S. Polonsky, M.D., Jo-Anna Reems, Ph.D., Reinhard G. Bretzel, M.D., Federico Bertuzzi, M.D., Tatiana Froud, M.D., Raja Kandaswamy, M.D., David E.R. Sutherland, M.D., Ph.D., George Eisenbarth, M.D., Ph.D., Miriam Segal, Ph.D., Jutta Preiksaitis, M.D., Gregory S. Korbutt, Ph.D., Franca B. Barton, M.S., Lisa Viviano, R.N., Vicki Seyfert-Margolis, Ph.D., Jeffrey Bluestone, Ph.D., and Jonathan R.T. Lakey, Ph.D., *International Trial of the Edmonton Protocol for Islet Transplantation.* The new england journal of medicine, 2006. **355**(13): p. 1318-1330.

30. Mariette R. Atkinson, M.A.S., Jesse T. Myers, and Alexander J. Ninfa, *Development of Genetic Circuitry Exhibiting Toggle Switch or Oscillatory Behavior in Escherichia coli.* Cell, 2003. **113**: p. 597-607.

31. Gould, D.J. and Y. Chernajovsky, *Endogenous GATA factors bind the core sequence of the tetO and influence gene regulation with the tetracycline system.* Mol Ther, 2004. **10**(1): p. 127-38.

32. Rang, A. and H. Will, *The tetracycline-responsive promoter contains functional interferon-inducible response elements.* Nucleic Acids Res, 2000. **28**(5): p. 1120-5.

33. Griffith, L.G., et al., *In vitro organogenesis of liver tissue.* Ann N Y Acad Sci, 1997. **831**: p. 382-97.

34. Schaefer, A.L., et al., *Generation of cell-to-cell signals in quorum sensing: acyl homoserine lactone synthase activity of a purified Vibrio fischeri LuxI protein.* Proc Natl Acad Sci U S A, 1996. **93**(18): p. 9505-9.

35. Engebrecht, J., K. Nealson, and M. Silverman, *Bacterial bioluminescence: isolation and genetic analysis of functions from Vibrio fischeri.* Cell, 1983. **32**(3): p. 773-81.

36. Engebrecht, J., M. Simon, and M. Silverman, *Measuring gene expression with light.* Science, 1985. **227**(4692): p. 1345-7.

37. Miyashiro, T. and E.G. Ruby, *Shedding light on bioluminescence regulation in Vibrio fischeri.* Mol Microbiol, 2012.

38. Engebrecht, J. and M. Silverman, *Identification of genes and gene products necessary for bacterial bioluminescence.* Proc Natl Acad Sci U S A, 1984. **81**(13): p. 4154-8.

39. Dunlap, P.V. and J.M. Ray, *Requirement for autoinducer in transcriptional negative autoregulation of the Vibrio fischeri luxR gene in Escherichia coli.* J Bacteriol, 1989. **171**(6): p. 3549-52.

40. Dunlap, P.V. and E.P. Greenberg, *Control of Vibrio fischeri luminescence gene expression in Escherichia coli by cyclic AMP and cyclic AMP receptor protein.* J Bacteriol, 1985. **164**(1): p. 45-50.

41. Collins, C.H., F.H. Arnold, and J.R. Leadbetter, *Directed evolution of Vibrio fischeri LuxR for increased sensitivity to a broad spectrum of acyl-homoserine lactones.* Mol Microbiol, 2005. **55**(3): p. 712-23.

42. Schmitz, M.L. and P.A. Baeuerle, *The p65 subunit is responsible for the strong transcription activating potential of NF-kappa B.* EMBO J, 1991. **10**(12): p. 3805-17.

43. Maier, T., S. Jenni, and N. Ban, *Architecture of mammalian fatty acid synthase at 4.5 A resolution.* Science, 2006. **311**(5765): p. 1258-62.

44. Agrimi, G., et al., *Identification of the human mitochondrial S-adenosylmethionine transporter: bacterial expression, reconstitution, functional characterization and tissue distribution.* Biochem J, 2004. **379**(Pt 1): p. 183-90.

45. Reed, M.A., et al., *The type I rat fatty acid synthase ACP shows structural homology and analogous biochemical properties to type II ACPs.* Org Biomol Chem, 2003. **1**(3): p. 463-71.

46. Kim, J.H., et al., *High cleavage efficiency of a 2A peptide derived from porcine teschovirus-1 in human cell lines, zebrafish and mice.* PLoS One, 2011. **6**(4): p. e18556.

47. Zhang, H.B., L.H. Wang, and L.H. Zhang, *Detection and analysis of quorum-quenching enzymes against acyl homoserine lactone quorum-sensing signals.* Curr Protoc Microbiol, 2007. **Chapter 1**: p. Unit 1C 3.

48. Peretti, S., et al., *Cell death induced by the herpes simplex virus-1 thymidine kinase delivered by human immunodeficiency virus-1-based virus-like particles.* Mol Ther, 2005. **12**(6): p. 1185-96.

49. Knappik, A. and A. Pluckthun, *An improved affinity tag based on the FLAG peptide for the detection and purification of recombinant antibody fragments.* Biotechniques, 1994. **17**(4): p. 754-61.

50. Tie, Y., et al., *Critical differences in HIV-1 and HIV-2 protease specificity for clinical inhibitors.* Protein Sci, 2012. **21**(3): p. 339-50.

51. Ryoichi Arai, H.U., Atsushi Kitayama, Noriho Kamiya and Teruyuki Nagamune, *Design of the linkers which effectively separate domains of a bifunctional fusion protein.* Protein Engineering, 2011. **14**(8): p. 529-532.

52. Briggs, J.A. and H.G. Krausslich, *The molecular architecture of HIV.* J Mol Biol, 2011. **410**(4): p. 491-500.

53. Sakuragi, J., *[What's going on post-budding?].* Uirusu, 2011. **61**(1): p. 91-8.

54. Lai, J.F., et al., *Doxycycline- and tetracycline-regulated transcriptional silencer enhance the expression level and transactivating performance of rtTA.* J Gene Med, 2004. **6**(12): p. 1403-13.

55. Tian, X., et al., *An improved tet-on system for gene expression in neurons delivered by a single lentiviral vector.* Hum Gene Ther, 2009. **20**(2): p. 113-23.

56. Buonaguro, L., et al., *Immunogenicity of HIV Virus-Like Particles in Rhesus Macaques by intra-nasal administration.* Clin Vaccine Immunol, 2012.

57. Pegu, P., et al., *Induction of strong anti-HIV cellular immunity by a combination of Clostridium perfringens expressing HIV gag and virus like particles.* Curr HIV Res, 2011. **9**(8): p. 613-22.

58. Zhang, X., et al., *Design and immunogenicity assessment of HIV-1 virus-like particles as a candidate vaccine.* Sci China Life Sci, 2011. **54**(11): p. 1042-7.

59. Tanaka, K., et al., *Epitope-tagging of a functional thyrotropin receptor: detection of the native receptor on intact cells.* Biochem Biophys Res Commun, 1996. **228**(1): p. 21-8.

60. Kari, B., et al., *Characterization of monoclonal antibodies reactive to several biochemically distinct human cytomegalovirus glycoprotein complexes.* J Virol, 1986. **60**(2): p. 345-52.

61. Fujiwara, K., et al., *A single-chain antibody/epitope system for functional analysis of protein-protein interactions.* Biochemistry, 2002. **41**(42): p. 12729-38.

62. Pfahl, M. and M. Hendricks, *Interaction of tight binding repressors with lac operators. An analysis by DNA-footprinting.* J Mol Biol, 1984. **172**(4): p. 405-16.

63. Baldauf, S.L., M.A. Cardani, and R.A. Bender, *Regulation of the galactose-inducible lac operon and the histidine utilization operons in pts mutants of Klebsiella aerogenes.* J Bacteriol, 1988. **170**(12): p. 5588-93.

64. Rotman, B., *On the rate limiting step in downhill transport via the LacY permease of Escherichia coli.* J Supramol Struct, 1977. **7**(1): p. 29-35.

65. Thompson, M.R., *Escherichia coli heat-stable enterotoxins and their receptors.* Pathol Immunopathol Res, 1987. **6**(2): p. 103-16.

66. Manhart, M.D., et al., *Evidence for pseudomonas exotoxin A receptors on plasma membrane of toxin-sensitive lm fibroblasts.* Infect Immun, 1984. **45**(3): p. 596-603.

67. Bau, M.Y. and R.K. Draper, *Ricin intoxicates End4 mutants that have an aberrant Golgi complex.* J Biol Chem, 1993. **268**(27): p. 19939-42.

68. Ulevitch, R.J., *Recognition of bacterial endotoxins by receptor-dependent mechanisms.* Adv Immunol, 1993. **53**: p. 267-89.

69. Daub, H., et al., *Role of transactivation of the EGF receptor in signalling by G-protein-coupled receptors.* Nature, 1996. **379**(6565): p. 557-60.

70. Ferguson, S.S., et al., *Role of beta-arrestin in mediating agonist-promoted G protein-coupled receptor internalization.* Science, 1996. **271**(5247): p. 363-6.

71. Shenker, A., *G protein-coupled receptor structure and function: the impact of disease-causing mutations.* Baillieres Clin Endocrinol Metab, 1995. **9**(3): p. 427-51.

72. Zheng, H., et al., *Selective recruitment of G protein-coupled receptor kinases (GRKs) controls signaling of the insulin-like growth factor 1 receptor.* Proc Natl Acad Sci U S A, 2012. **109**(18): p. 7055-60.

73. Shoichet, B.K. and B.K. Kobilka, *Structure-based drug screening for G-protein-coupled receptors.* Trends Pharmacol Sci, 2012.

74. Bourret, R.B., et al., *Protein phosphorylation in chemotaxis and two-component regulatory systems of bacteria.* J Biol Chem, 1989. **264**(13): p. 7085-8.

75. Gross, R., B. Arico, and R. Rappuoli, *Families of bacterial signal-transducing proteins.* Mol Microbiol, 1989. **3**(11): p. 1661-7.

76. Ninfa, A.J., *Protein phosphorylation and the regulation of cellular processes by the homologous two-component regulatory systems of bacteria.* Genet Eng (N Y), 1991. **13**: p. 39-72.

77. Kleiman, J.H. and B. Moss, *Characterization of a protein kinase and two phosphate acceptor proteins from vaccinia virions.* J Biol Chem, 1975. **250**(7): p. 2430-7.

78. Zhou, R., et al., *Characterization of protein kinase A-mediated phosphorylation of ezrin in gastric parietal cell activation.* J Biol Chem, 2003. **278**(37): p. 35651-9.

79. Reiter, C.E., et al., *Characterization of insulin signaling in rat retina in vivo and ex vivo.* Am J Physiol Endocrinol Metab, 2003. **285**(4): p. E763-74.

80. Shenoy, A.R., *His kinase or mine? histidine kinases through evolution.* J Biosci, 2000. **25**(4): p. 317-22.

81. Besant, P.G., E. Tan, and P.V. Attwood, *Mammalian protein histidine kinases.* Int J Biochem Cell Biol, 2003. **35**(3): p. 297-309.

82. Steeg, P.S., et al., *Histidine kinases and histidine phosphorylated proteins in mammalian cell biology, signal transduction and cancer.* Cancer Lett, 2003. **190**(1): p. 1-12.

83. Crepin, S., et al., *The Pho regulon and the pathogenesis of Escherichia coli.* Vet Microbiol, 2011. **153**(1-2): p. 82-8.

84. Bezy, R.P. and D.M. Kehoe, *Functional characterization of a cyanobacterial OmpR/PhoB class transcription factor binding site controlling light color responses.* J Bacteriol, 2010. **192**(22): p. 5923-33.

85. Jacob Beal, R.W., Fusun Yaman, Noah Davidsohn, and Aaron Adler, *A Method for Fast, High-Precision Characterization of Synthetic Biology Devices*, in *MIT-CSAIL-TR-2012-008*2012, Massachusetts Institute of Technology: Cambridge, MA.

86. Leisner, M., et al., *Rationally designed logic integration of regulatory signals in mammalian cells.* Nat Nanotechnol, 2010. **5**(9): p. 666-70.

87. Engler, C. and S. Marillonnet, *Generation of families of construct variants using golden gate shuffling.* Methods Mol Biol, 2011. **729**: p. 167-81.

88. Wing, K.D., *RH 5849, a nonsteroidal ecdysone agonist: effects on a Drosophila cell line.* Science, 1988. **241**(4864): p. 467-9.

89. Saez, E., et al., *Identification of ligands and coligands for the ecdysone-regulated gene switch.* Proc Natl Acad Sci U S A, 2000. **97**(26): p. 14512-7.

90. Wyborski, D.L., J.C. Bauer, and P. Vaillancourt, *Bicistronic expression of ecdysone-inducible receptors in mammalian cells.* Biotechniques, 2001. **31**(3): p. 618-20, 622, 624.

91. Ai, H.W., et al., *Exploration of new chromophore structures leads to the identification of improved blue fluorescent proteins.* Biochemistry, 2007. **46**(20): p. 5904-10.

92. Fischbach, M. and C.A. Voigt, *Prokaryotic gene clusters: a rich toolbox for synthetic biology.* Biotechnol J, 2010. **5**(12): p. 1277-96.

93. Strauss, W.M., *Transfection of mammalian cells via lipofection.* Methods Mol Biol, 1996. **54**: p. 307-27.

94. Bleris, L., et al., *Synthetic incoherent feedforward circuits show adaptation to the amount of their genetic template.* Mol Syst Biol, 2011. **7**: p. 519.

95. Baek, J.H., Y.J. Kang, and S.Y. Lee, *Transcript and protein level analyses of the interactions among PhoB, PhoR, PhoU and CreC in response to phosphate starvation in Escherichia coli.* FEMS Microbiol Lett, 2007. **277**(2): p. 254-9.

96. Bachhawat, P., et al., *Mechanism of activation for transcription factor PhoB suggested by different modes of dimerization in the inactive and active states.* Structure, 2005. **13**(9): p. 1353-63.

97. Vlassov, A.V., et al., *Exosomes: Current knowledge of their composition, biological functions, and diagnostic and therapeutic potentials.* Biochim Biophys Acta, 2012.

98. Casadaban, M.J. and S.N. Cohen, *Analysis of gene control signals by DNA fusion and cloning in Escherichia coli.* J Mol Biol, 1980. **138**(2): p. 179-207.

99. Miyoshi, H., et al., *Development of a self-inactivating lentivirus vector.* J Virol, 1998. **72**(10): p. 8150-7.

100. Mayer, M.P., *A new set of useful cloning and expression vectors derived from pBlueScript.* Gene, 1995. **163**(1): p. 41-6.

101. Miller M, Hafner M, Sontag E, Davidsohn N, Subramanian S, et al. *Modular Design of Artificial Tissue Homeostasis: Robust Control through Synthetic Cellular Heterogeneity.* PLoS Comput Biol (2012)

102. Antero Salminen, Kai Kaarniranta, *Control of p53 and NF-κB signaling by WIP1 and MIF: Role in cellular senescence and organismal aging.* Cellular Signalling, Volume 23, Issue 5, May 2011, P. 747-752

103. Sharon L. Paige, Sean Thomas, Cristi L. Stoick-Cooper, Hao Wang, Lisa Maves, Richard Sandstrom, Lil Pabon, Hans Reinecke, Gabriel Pratt, Gordon Keller, Randall T. Moon, John Stamatoyannopoulos, Charles E. Murry, *A Temporal Chromatin Signature in Human Embryonic Stem Cells Identifies Regulators of Cardiac Development.* Cell, 28 September Vol.151.1, , P. 221-232,

104. K. Ampatzis, P. Makantasi, C.R. Dermon, *Cell proliferation pattern in adult zebrafish forebrain is sexually dimorphic.* Neuroscience, 2011 P.171-176

105. W C Fuqua, S C Winans, and E P Greenberg *Quorum sensing in bacteria: the LuxR-LuxI family of cell density-responsive transcriptional regulators.* J.Bacteriol, January 1994 Vol. 176 P.269-275

106. Mei-zhen Wang, Xin Zheng, Hong-zhen He, Dong-sheng Shen, Hua-jun Feng, *Ecological roles and release patterns of acylated homoserine lactones in Pseudomonas sp. HF-1 and their implications in bacterial bioaugmentation.* Bioresource Technology, December 2012Vol. 125, , P. 119-126

107. Mergener K, Fäcke M, Welker R, Brinkmann V, Gelderblom HR, Kräusslich HG. *Analysis of HIV particle formation using transient expression of subviral constructs in mammalian cells.* Virology. 1992 Jan, Vol.186 P. 25-39.

108. Personal communication with Brynne Stanton in the Voigt lab March 2012.

109. Personal communication with Shimyn Slomovic in the Collincs lab April 2012.

110. Jacob Beal, Ron Weiss, Douglas Densmore, Aaron Adler, Evan Appleton, Jonathan Babb, Swapnil Bhatia, Noah Davidsohn, Traci Haddock, Joseph

Loyall, Richard Schantz, Viktor Vasilev, and Fusun Yaman, *An End-to-End Workflow for Engineering of Biological Networks from High-Level Specifications,* ACS Synthetic Biology, online publication July 2012.