A Novel Method for Mammalian Large Genetic Circuit Assembly and Delivery

By

Yinqing Li

B. S., Microelectronics

Fudan University, 2008

Submitted to the Department of Electrical Engineering and Computer Science

in partial fulfillment of the requirements for the degree of

Master of Science in Electrical Engineering and Computer Science

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

February 2012

Author .....
Department of Electrical Engineering and Computer Science
December 27, 2011

Certified by .....
Ron Weiss
Associate Professor
Department of Electrical Engineering and Computer Science
Department of Biological Engineering
Thesis Supervisor

Accepted by .....
Leslie A. Kolodziejski
Professor
Department of Electrical Engineering and Computer Science
Chairman, Department Committee on Graduate Students

A Novel Method for Mammalian Large Genetic Circuit Assembly and Delivery

By

Yinqing Li

Submitted to the Department of Electrical Engineering and Computer Science
on December 27, 2011, in partial fulfillment of the
requirements for the degree of
Master of Science in Electrical Engineering and Computer Science

Abstract

Genetic manipulation of mammalian cells provides a foundation for contemporary biological research both basic and applied. Existing methods for construction and introduction of large scale exogenous genetic information into mammalian cells and for creating stable cell lines are not efficient and suffer from limitations in terms of cost, speed, flexibility, and reliability.

In this thesis, a novel method is presented for the efficient construction and delivery of complex genetic circuits into mammalian cells. Multi-gene circuits are assembled with high efficiency from a validated modular library into single pieces. The assembled circuits can be used for transient expression and each individual circuit can be integrated into a cellular genome to create a stable cell line.

Genetic circuits were constructed that contain several expression units, including inducible control units and fluorescent markers. These circuits were delivered into Human Embryonic Kidney 293 (HEK293) cells for both transient and stable expression cases. Results show that the introduced genetic circuits performed as designed and that stable cell lines, each with the desired phenotype could be created efficiently. Several factors affecting the assembly efficiency and the performance of resulting circuits are also discussed.

Thesis Supervisor: Ron Weiss
Title: Associate Professor

Acknowledgements

First and foremost, I offer my sincere gratitude to my thesis advisor, Professor Ron Weiss, the Director of Synthetic Biology Center at MIT, for his invaluable support, help, and patience throughout this thesis.

Thanks to the people I work with: Dr. Patrick Guye, with whom I developed the initial idea and collaborated closely on this project and whose insights and knowledge helped lead this work in the right direction; Dr Liliana Wroblewska, for her generous help on FACS experiments and screening of Bxb1 cell lines; Xavier Duportet, who brought Bxb1 recombinase into this lab and created the first landing pad cell line; and all the current and past members of the Weiss group, for their help and friendship.

I would like to thank my family. I am grateful to my parents Prof. Fuli Li and Wenyi Zhang for their unconditional encouragement and love throughout my study. I would especially like to thank my fiancée Mangmang. She came to the East Coast right after she finished college to stay with me.

Finally, I would like to give thanks to my good friend and ex-roommate Nevan Clancy Hanumara for editing and proof-reading of this thesis. I would also like to thank Elizabeth Harris from MIT Writing Center for her valuable advices on English writing.

Without the help and assistance from all these people, this thesis is simply impossible.

Contents

List of Figures

List of Tables

Chapter 1

1. Introduction

Genetic manipulation of mammalian cells is a fundamental tool in basic and applied biological research. The mammalian cell can be thought of as a machine being run by a regulatory network of genes. A gene regulatory network is a collection of DNA segments (genes) that interact with each other (through their RNA and protein products) and with proteins in the cell, thereby governing the rates at which genes in the network are transcribed into mRNAs. These mRNAs encode proteins that fulfill cellular functions, including growth and division, signaling, regulation, and differentiation[1], [2]. Genetic circuits are subsets of larger gene regulatory networks consisting of functional clusters of genes that can modulate each other's expression through regulatory elements [3].

From the point of view of biologists, custom-building genetic circuits and testing them in living cells provides an unique opportunity to observe the behavior of biological components in different contexts, reveals the extent of our understanding of natural systems, and provides new insights in genetic control and molecular dynamics. One notable example was recent discoveries of stem cell differentiation pathways. Expression of certain stem cell-associated genes in fibroblasts can transform the cells to the pluripotent state [4], [5]. Similar studies have shown that fibroblast cells and stem cells can be reprogrammed into, for instance, pan-neurons [6], [7] or specific neuronal subtypes by ectopic gene expression [8], [9]. Work by Aviv [10] demonstrated that the hematopoietic cell differentiation process can be precisely controlled by exploiting one of the network motif regulators. In all of these experiments, master regulators governing cell differentiation circuitry were selected and perturbed, demonstrating the importance of genetic manipulation at the gene circuit level in answering basic biology questions.

For bioengineers, the emerging discipline of synthetic biology aims to rewire and reprogram organisms with novel phenotypes by assembling molecular "parts" (such as genetic elements) into biological "devices" (genetic circuits) that exhibit predictable behavior [11-13]. Applications include microbial production of chemical precursors, novel antibiotics and biofuels, and creation of synthetic attenuated viruses for use as vaccines [14], [15]. Synthetic biology also offers an alternative bottom-up approach to understanding biological networks and disease mechanisms, based on designing and constructing simple synthetic gene circuits [16]. Over the last decade, synthetic approaches

have provided key insights into the principles of gene circuit design [17-22]. To date, however, advances in synthetic biology are largely represented by synthetic bacterial systems [11], [20], [23], while eukaryotic synthetic biology efforts, especially in mammalian cells, are lagging behind, despite their indispensable potential value in gene-function analysis, drug discovery, the pharmaceutical manufacturing of protein therapeutics, and the implementation of gene-based therapies [12], [24-29]. As synthetic biology moves towards introducing more complex synthetic system into mammalian cells, tools, similar to the mature techniques employed in *Escherichia coli* (*E. coli*) genetics, have to be developed to facilitate the manipulation of mammalian genetic systems [11]. An essential element is the ability to physically design, construct and deliver complex genetic circuits which contain large numbers of expression units with regulatory elements.

1.1. Thesis Statement

In this thesis, we present a system to rapidly construct and deliver large mammalian genetic circuits. Firstly, Gateway recombination is employed to generate expression units out of mammalian genetic parts. Then a Gibson-based method is applied to rapidly assemble a genetic circuit consisting of these units. The assembly process can be easily automated, allowing the construction of a scalable circuit comprising many parts without human intervention. The assembled genetic circuit is then ready to be introduced into mammalian cells using a delivery method compatible with a carrier vector chosen. The achievement of this capability has the potential to greatly advance experimental mammalian genetics, and promote biological investigation as well as the synthetic engineering effort.

1.2. Previous Work

Current approaches to design, construct and deliver complex genetic circuits share one or more disadvantages, which are summarized in the following sections.

1.2.1. Genetic Circuit Design and Construction Methods

First of all, we want to introduce the metrics, shown in Table 1-1 with which we evaluate methods for genetic circuit design and construction.

Depending on the application, a genetic circuit could be a simple vector for constitutive co-expression of genes, or a circuitry with complex regulatory interaction. Circuit design

and construction methods should be driven to accommodate different practical needs. Once a circuit is designed, it needs to be constructed. In most cases it is difficult to directly synthesize the entire circuit as the cost associated with synthesis of a large construct is still prohibitive. The usual technique is to start with DNA molecules that contain sequences of interest. Then recombinant DNA technique is used to combine those pieces of desired sequences. This often involves physical transfer of genetic materials, e.g. a promoter being cloned from the original plasmid into a new circuit, with the potential for introducing DNA mutations. Mutations could make a circuit different from the desired. Therefore, construction fidelity is critical.

Modular design, using a verified library of parts, is a good way to ensure fidelity, reuse existing materials, standardize construction and minimize construction steps.

Scalability is another important metric. As we design more complex gene circuits, two pertinent questions are: Does the method provide the ability to build the circuit? Does the method allow us to add new parts to existing functional circuits? Employing scalable methods will yield good solutions to both. It is important to point out that mammalian gene expression is not as simple as that in *E. coli*. For a circuit to be persistently functional, chromatin modulating elements and other important genetic elements should be taken into account at the design step.

	Metric	Explanation
1	Design Flexibility	Independent expression regulation; Complex regulatory gene circuit
2	Construction Fidelity	Mutation free construction
3	Modularity	Library based method; Reuse of existing materials
4	Scalability	Construction of larger gene circuit
5	Ease of construction	Simple protocol
6	Expression Reliability	Gene expression as designed

Table 1-1: Metrics for evaluating genetic circuit design and construction methods

1.2.1.1. Gene Circuit Design

A basic part of a gene circuit is the promoter-gene pair or expression unit. In mammalian cells, transcription machinery binds the promoter and transcribes a gene into an mRNA, which is then translated to a protein. Similarly, microRNA can be transcribed, and processed for RNA interference.

A direct extension of a single promoter-gene pair to a multi-gene circuit is the concatenation of genes together under the same promoter to form a polycistronic construct. These genes are transcribed from a single promoter, but the mRNA product is then translated into separate products. IRES and 2A sequences serve this purpose [30], [31]. Internal ribosome entry site (IRES) sequences are able to associate with ribosomes in the middle of an mRNA and to initiate translation of the downstream gene. Co-expression of genes can be done using IRES, however the expression levels of the genes upstream and downstream of IRES are highly correlated and the genes placed downstream are expressed at a much lower level [30], [32], [33]. The 2A sequence, first found in the foot-and-mouth disease virus (FMDV), encodes a short polypeptide that mediates self-cleavage by a translational effect [31]. Ribosome-procession-arrest leads to release of the nascent polypeptide and re-initiation of translation at the next in-frame codon (ribosome skipping). In this way, discrete protein products are derived from a single promoter. Problems arises in both approaches as they do not have transcriptional regulation over individual downstream genes, thus the polycistronic method does not meet our needs for design flexibility in building complex gene circuits.

A complex genetic circuit that has multiple expression units can be designed onto several vectors instead, with each vector carrying one promoter-gene pair. Vectors can be co-transfected into cells for transient or stable expression, however, the stochastic stoichiometry of transfected vectors may lead to failure of the entire circuit because of a possible input-output mismatching between units [13]. The problem becomes even more severe when a large number of vectors need to be co-transfected for a large circuit. In addition, it is difficult to select the cells containing the entire genetic circuit due to the limited number of selection markers [34]. Alternatively, here we assemble multiple promoter-gene pairs into a single large vector. The genetic circuit on a large vector can be transiently transfected into cells or stably integrated in the chromosome. Only one selection marker is necessary for selecting a stable cell line baring the multi-gene construct.

1.2.1.2. Gene Circuit Construction Methods

In this section, we discuss common methods for assembly of multi-gene constructs.

Restriction Enzyme-based Cloning Methods

Promoters, genes, or parts are initially stored on separated vectors and can be joined to

form a gene circuit through a cloning process using restriction endonuclease (RE). RE are enzymes that can cleave DNA molecules only when specific DNA sequences are encountered, called restriction sites (RS). A cloning vector is treated with an RE to cleave the DNA at the site where a new genetic part will be inserted. The genetic part to be inserted can be amplified via a polymerase chain reaction (PCR) (an error-prone step) with a compatible restriction site added on both ends. It is then treated with a restriction endonuclease to generate ends compatible to those on the vector. The creation of the assembled DNA is a simple process where digested vector and insert are mixed at appropriate concentrations and exposed to a DNA ligase that covalently links the ends in a process called ligation. Ligation usually results in a DNA mixture containing randomly joined ends, out of which the desired product should be selected [35]. Restriction enzyme based cloning can be used for assembly of multi-gene circuits, such as biobricks [36], where DNA parts are modularized by removing certain RS inside and framing with those RSs. However, several rounds of PCR, restriction enzyme digestion, and ligation processes may be needed to build the full circuit, making it a time consuming and error-prone method. Besides, REs that are usually 6bp in length are frequently present in mammalian genetic sequences e.g. promoter, gene in many copies, thereby limiting its use for construction of complex mammalian gene circuits.

Type IIS Restriction Enzyme-based Cloning Methods

This cloning method is based on the special ability of type IIS restriction enzymes to cleave outside of their recognition sequence. Since actual cut sites are outside of the IIS restriction recognition site, the cut sites can be designed to create different overhangs so that ligation of multiple fragments in a single step feasible. Intermediate ligated products still contain restriction sites and should be digested until ligated properly. Examples using this method include Golden Gate cloning, and MoClo [37], [38]. Usually, restriction enzymes like BsaI and BsmBI that can create 4bp overhang are used to maximize the number of possible distinct overhangs. The disadvantage of this method is that it relies on enzymes that usually only have 6bp recognition sequences. As discussed in the previous section, it poses a difficulty for its application in mammalian gene circuit construction. In addition, it has been shown that ligation with 4bp overhangs is vulnerable to mismatching, and, therefore, the cloning efficiency decreases significantly when the number of input fragments increases[39].

Recombinase-based Cloning Methods

Recombinase-based cloning methods are easy to use and highly efficient as compared to restriction-ligation methods. These technologies take advantages of bacterial or viral site-specific recombinases such as phage λ integrase, phage P1 Cre, or yeast recombinase FLP [40], [41]. Recombinases catalyze a double-stranded DNA exchange between two specific DNA sequences called recombination sites. Gateway (Invitrogen) is one of the most popular recombination cloning method and is based on the site-specific recombination reactions of bacteriophage lambda in *E. coli* [42]. The recombination reactions can be represented as follows:

attBi x attPi ↔ attLi x attRi ($i \in n$ n is the number of orthogonal recombination sites)

The attBi x attPi reaction (BP reaction) is mediated by proteins Int and IHF.
The attLi x attRi reaction (LR reaction) is mediated by proteins Int, IHF, and Xis.

Only recombination sites belonging to the same set (same i, B to P, R to L) can be recombined [43]. Several mutations have been introduced into att sites to make the LR reaction irreversible and to optimize reaction specificity and efficiency. The genetic element of interest can be PCR amplified with attBi sites on the ends and inserted into a cloning vector carrying attPi (donor vector) in a BP reaction. The reaction products are a byproduct and a vector (entry vector) bearing the genetic part of interest flanked by attLi/attRi sites (crossover of attBi and attPi sites). This process can be easily extended to create a library of entry vectors. Entry vectors are sequencing verified and can be used in LR reactions for the assembly of gene circuits in a modular fashion with high fidelity and efficiency. Nevertheless, the commercial Gateway system only allows 4-way multisite gateway recombination, which greatly limits the complexity of the circuits that can be built. A time-consuming system has been proposed to assemble gene circuits hierarchically by alternating negative selections in successive rounds [44].

Another approach exploits Cre-mediated recombination [45] using a Cre/LoxP-based strategy, where each circuit-component contains a LoxP site and a distinct selection marker. Components are fused into one single vector by recombining their LoxP site in the presence of Cre. The desired product is selected by a combination of bacterial antibiotics and a strain-specific origin of replication. The maximum number of promoter-gene pairs is limited to 5, the number of available prokaryotic antibiotics.

In Fusion cloning method

Another method, In Fusion, use a patented enzyme to catalyze DNA ligation [46]. The 3'

and 5' regions of the homolog called extension are added onto the ends of inputs. The enzyme joins the ends of linear inputs that share the same extension. When linear inputs are combined, the In-Fusion enzymes recess the double-stranded extension into single-stranded DNA overhangs and fuse those regions to their complementary ends. In Fusion requires PCR [47] that makes In Fusion method susceptible to mutations. Besides, this method is not scalable in practice because of its high cost.

Gibson Method

Daniel Gibson described an enzymatic reaction mix composed of exonuclease, polymerase, and ligase [48]. In this reaction, the exonuclease first chews back DNA to expose long overhangs, and then compatible overhangs are annealed. Over-exposed regions are filled in by a polymerase and joints are repaired by a ligase. All these enzymatic reactions can be carried out in a single isothermal step in an optimized reaction buffer [49]. This enzyme mix was used to assemble DNA molecule as large as 900kb in vitro. However, the method requires PCR or chemically synthesized products as inputs which are either error prone or expensive. Restriction digested fragments can also be used, but just like other restriction enzyme based methods, this is not applicable to mammalian gene circuit assembly.

1.2.2. Delivery Methods

This section discusses current methods for introducing genetic circuits into mammalian cells.

1.2.2.1. Transient Transfection

Assembled circuits can be delivered into mammalian cells using chemical-based transfection or electroporation for transient expression. Chemical-based transfection can be divided into several types, mainly calcium phosphate [50], liposomes [51], and cationic polymers [52]. In all cases, a complex of chemical compound and nucleic acid is first formed and added onto cells. The nucleic acid is taken up by cells via different mechanisms. Liposomes are small membrane-bounded bodies that resemble a cell membrane. They can fuse with the cell membrane and deposit DNA. The complex comprising DNA and cationic polymers is imported into cells by endocytosis. The details of the calcium phosphate precipitation process are not entirely understood. Electroporation is another widely used method. It transiently increases the permeability of cell membrane

with short but strong electric pulses, during which time negatively charged DNA diffuses into the cells. Transfected DNA needs to be transferred into the nucleus for expression, usually during cell division. Nucleofector technology combines electroporation with special solution that can directly deliver DNA into the nucleus [53]. In a typical transient transfection experiment, the introduced genetic material is diluted out as cells proliferate since the vectors are not replenished *in vivo*.

1.2.2.2. Stable Cell Line

A stable cell line is one in which an exogenous gene circuit propagates with chromosome or is integrated into the chromosome. Cells can propagate without losing the introduced genetic materials, and thus have many advantages over transient expression. The established stable cell line can be used for long term experiments and be selected for continuous bioengineering or pharmaceutical production. To create a stable cell line, the process starts with transient transfection, but in this case, the genetic material either becomes an episomal plasmid or is integrated into the genomic DNA.

When comparing methods for generating stable cell lines, a few factors are usually considered to be important [54]. (1) Specificity: Random integration may disrupt endogenous genetic elements, including important genes and regulatory sequences, resulting in unpredictable changes in cell phenotype. This can also have a profound impact on the gene circuit performance, for example, integration in heterochromatin region could result in silencing of the transgenes. (2) Ease of selection: Some methods may require less effort to select a clonal cell line with desired properties while other may be time consuming and labor intensive. (3) Stability: Transgenes should be maintained and be under little influence from the surrounding genomic sequences.

Non-Integrating Method:

Episomal Vector

The episomal vector or replicating plasmid method is an approach that does not require genomic integration. Several replicating elements taken from viruses including EBV and SV40 have been harnessed for keeping a transfected plasmid in an episomal state. In general, the problem with viral episomal vectors is the requirement of viral regulatory factors, which are often related to cell transformation, immunogenicity, and cytotoxicity [45], [54-56]. Alternatively, scaffold/matrix attachment regions (S/MAR) from the mammalian interferon-γ gene cluster have been incorporated into episomal vectors to

replace viral genes. Despite the fact that the vector was shown to be episomal for a short time without selection [57-59], long term loss of plasmid (5%) [60], integration in the host chromosome, and rearrangements within the episomal vectors in both viral and non-viral methods have been reported [61]. Moreover, vector copy number varies considerably among the same cell lines [57], and a cross different cell lines [61].

Integrating Methods:

Homologous Recombination-based Methods

In mammalian cells, homologous recombination (HR) is used in the repair of double strand break (DSB) in DNA. Repair machinery, with the guide of homologous sequences, finds and connects a broken DNA molecule to another DNA molecule via Holliday junctions. The junctions are resolved by enzymes and the two DNA molecules are linked. Homologous recombination is used in gene targeting, where integrating donor DNA contains the similar or identical sequences to the genomic locus where donor is to be integrated [62]. However, this process is quite inefficient, as the frequency of HR between donor and target sequences is low [63].

A big improvement in this field was the discovery that recombination could be stimulated several orders of magnitude by means of a single DSB in genomic DNA [64]. Chimeric endonuclease proteins have been proven versatile and effective for this purpose. The synthetic protein links a DNA-binding domain to a DNA-cleavage-domain. The DNA-binding domain recognizes a specific genomic location, and it brings in DNA-cleavage domain, the dimer of which cuts DNA. Zinc finger (ZF) assembly was traditionally used as a DNA-binding domain. Different sequences could be targeted by evolving and optimizing the ZF assembly [65-67], which is commercially available as CompoZr integration technology [68]. Recently, the discovery of TAL (transcription activator-like)proteins provides an alternative to ZF [69-71]. The TAL protein is composed of tandem repeats signified by their repeat variable di-residues (RVDs). There is a one-to-one correspondence between RVDs in the repeat domain and the DNA sequence the protein binds, which constitutes a binding-code [72], [73]. A TAL binding domain can be efficiently constructed using this binding-code to target a given sequence [39], [74].

Although promising, the HR based integration method often associates with non-homologous end joining (NHEJ), which could result in endogenous gene disruption and deletion. Even with the endo-nuclease targeting, random integration is still common from both HR and NHEJ [67], [69]. The case is even worse in the delivery of complex mammalian gene circuits as endogenous promoter sequence is often employed in a circuit.

Virus-based Methods

Viruses have evolved naturally to attack their host and delivery their genetic material into the host cell effectively. Replication incompetent viruses were engineered as vehicles to carry gene circuits on their genomes and transfer them into host cells. The most widely used viral vectors are derived from the Adeno-Associated Virus (AAV) Vector [75], and the Lentivirus vector [76] chosen for their relative simplicity in handling. However, delivery capacity with these viral vectors is restricted by their viral packaging size, usually 6kb for the AAV and 10kb for the Lentivirus [77]. A bigger concern with viral approach, in general, is that non-specific incorporation of DNA into the host chromosome (insertional mutagenesis) can disrupt cellular sequences, leading to undesired effects. Besides, the expression of virus factors during and after viral infection have implications in cytotoxicity and cell transformations[55].

Recombinase-based Methods

The performance of the methods mentioned above is more or less subject to unpredictable cellular genomic alternations. Recombinase based methods or more recent recombinase-mediated cassette exchange (RMCE) approaches have been proposed to enable non-disruptive insertion of genetic cassettes at pre-characterized genomic locus [78]. The idea behind these methods is similar to that of recombinase based assembly. Here, a pair of recombination sites are placed on a genetic circuit to be delivered and on a pre-characterized genomic target locus (landing pad site) respectively. The landing pad is inserted into a host cell chromosome prior to delivery. Upon expression of the recombinase a gene circuit can be inserted in the landing pad, or be exchanged into the landing pad (RMCE). Cre-LoxP and Flp-FRE systems are traditionally used for this purpose. Unlike those two, phiC31 catalyzes irreversible integration reaction, and is thus more efficient at transgene insertion or cassette exchange. In parallel to their benefits, caution is called for when using these recombinases. Pseudo integration sites for loxP, FRT, and phiC31 have been identified in the mouse genome. They can mediate recombination between the degenerate sequences although with lower efficiency [79-81]. Cytotoxicity was also reported [82] when recombinases were over-expressed. More recently, Mycobactriophage Bxb1 integrase has been shown capable of site-specific integration with higher efficiency and specificity compared to phiC31 [83], [84], yet more studies need to be done to elucidate their performance.

1.3. Thesis Organization

This thesis is organized as follows: Experimental materials and methods are listed in Chapter 2. Chapter 3 describes the flow of the newly proposed method for large genetic circuit construction and its important features. In addition, results of the experiment are given and discussed. In Chapter 4 several delivery methods are developed and tested individually for introducing large genetic circuits into mammalian cells . Additional experiment on genetic elements, called insulator sequences, are also presented. Conclusions and suggestions for future improvements are given in Chapter 5.

Chapter 2

2. Material and Methods

Molecular Biology

Restriction enzymes I-SceI endonuclease (R0694) and T4 DNA ligase (M0202) were purchased from New England Biolabs (NEB), MA. AccuPrime Pfx Supermix from Invitrogen (Life technologies, Carlsbad CA) was used for PCR amplification. Taq polymerase (M0273) and LongAmp Taq polymerase (M0323S) from NEB were used in PCR verification of the site specific integration. Enzymes used for Gibson assembly were purchases from vendors as described in the paper [1]. The E.cloni 10G (60080) bacteria strain from Lucigen was used for most cloning experiments except for the pJazz based system. A BigEasy v2.0 Linear Cloning System (pJazz) (43024, Lucigen Corp., Middletown WI) was purchased and used according to its manual. Antibiotic were used with following concentrations: 100 ug/mL ampicillin, 50ug/ml kanamycin, 25ug/ml chloramphenicol. Gel extraction was done with a Qiagen QIAquick Gel Extraction Kit. PCR purification was done using a Qiagen QIAquick PCR Purification Kit. A NucleoTraPCR from Macherey-Nagel was used for purification in some cases as indicated in the main text. Miniprep of DNA was accomplished using a Qiagen Qiaprep Spin Miniprep Kit. Some minipreps were automated on a Qiagen Qiacube using the same miniprep kit.

Gateway pENTR library

Gateway BP reactions were performed in accordance with the manufacturer's manual (Life technologies, Carlsbad CA) except for scaling down of the reactants used and reaction volumes. Briefly, to create the promoter library, promoter sequences of interest were amplified, digested, and inserted into pENTR_L4_L1 cut with compatible restriction enzymes. To create a gene library, genes of interest were amplified with an attB1 site added (GGGGACAAGTTTGTACAAAAAAGCAGGCTGA) in the forward primer and an attB2 site added (GGGGACCACTTTGTACAAGAAAGCTGGGTA) in the reverse primer. 10 fmol of the PCR product was mixed with 5 fmol of pDONR221_P1_P2 and incubated with 1ul of BP clonase II (11789-020, Life technologies, Carlsbad CA) for 1 hour. 1 ul of the reaction was transformed into ccdB sensitive competent *E. coli* cells. As an

alternative protocol, gene sequences of interest were amplified, digested, and inserted into pENTR_L1_L2 cut with compatible restriction enzymes.

Gateway LR reaction

Gateway LR reactions were performed in accordance with the manufacturer's manual (Life technologies, Carlsbad CA) except for scaling down of the reactants used and reactions' volumes. Briefly, 5 fmol of each of pENTR_L4-Promoter-R1, pENTR_L1-Gene-L2, and pZDonor_Seqn-GTW-Seqn+1 that contains Gateway cassette of pDEST_R4_R2 were mixed and incubated with 1ul of LR clonase II mix (11971-020, Life technologies, Carlsbad CA) for 16 hours. 1 ul of the reaction was transformed into ccdB sensitive competent *E. coli* cells.

Assembly

- **Digestion of Input Vectors**

70 fmol of each vector containing basic assembly units were pooled and digested in a total volume of 20 ul for three hours using 10 units of the restriction enzyme I-SceI (R0694, NEB Biolabs, MA). Subsequently, this digest was purified using the QIAGEN Qiaquick PCR purification kit and eluted into 30ul of warm EB buffer.

- **Digestion of the Adapter Vector**

210 fmol of the adapter vector required for proper circuit closure was digested in a total volume of 20 ul with 5 units each of restriction enzymes XbaI and XhoI (R0146, R0145, NEB Biolabs, MA) for three hours, purified using the QIAGEN Qiaquick PCR purification kit and eluted into 30 ul of warm EB buffer.

- **Digestion of the Carrier Vector**

2100 fmol of the carrier vector was digested in a total volume of 20 ul with 10 units of the restriction enzyme FseI (R0588, NEB Biolabs, MA) for one hour, purified using the QIAGEN Qiaquick PCR purification kit and eluted into 30 ul of warm EB buffer.

- **One-Step Assembly (Gibson Reaction)**

Isothermal Reaction Buffer (ISO) was prepared on ice and stored at -20C: 25% PEG-8000, 500 mM Tris-HCl pH 7.5, 50 mM MgCl2, 50 mM DTT, 1 mM each of each dATP, dCTP, dGTP, dTTP, and 5 mM NAD. Assembly Master Mix was prepared on ice and stored at -20C: 320 µl ISO buffer, 0.64 µl of 10 U/ µl T5 exonuclease (T5E4111K, Epicentre Biotechnologies, WI), 20 µl of 2 U/µl Phusion polymerase (F-530, NEB Biolabs, MA), 160 µl of 40 U/µl Taq ligase (M0208, NEB Biolabs, MA), bidistilled water to 1.2 ml [49]. 7 fmol of each part (digested and purified adapter vector, carrier vector and pool of assembly units) were combined in a 200 ul PCR reaction tube on ice and filled up to 5 ul of the total volume with bidistilled water. The mix was then added into 15 ul of Assembly Master Mix and the reaction was incubated at 50C for one hour. 2 ul of this reaction were then transformed into competent *E. coli* cells.

- **Hierarchical Assembly**

BAMs and adapter were assembled into a hierarchical carrier vector pJazz Swing. 70 fmol of the assembled vector was digested in a total volume of 20 ul for three hours using 10 units of the restriction enzyme I-SceI. Subsequently, this digest was purified using the QIAGEN Qiaquick PCR purification kit and eluted into 30 ul of warm EB buffer. 7 fmol of this digest were pooled with other 7fmol BAMs digest, 7 fmol adapter vector, and 7 fmol carrier vector. A one-step assembly protocol was applied in this mixture.

Microscope Imaging

Images were taken using a Zeiss Axiovert 200M epi-fluorescence microscope equipped with a 1344×1024 pixel cooled ORCA-ER CCD camera (Hamamatsu Corporation) and a 10x objective. Fluorescence images were analyzed in Zeiss Axiovision digital image processing software.

Flow Cytometry Measurement

Flow cytometry measurement was carried out on a BD LSR II in the Koch Institute Flow Cyctometry Core at MIT. Data was collected in BD FACSDiva software and analyzed in Flowjo (Tree Star, Inc. Ashland, OR).

Cell Culture

HEK293FT from Invitrogen was cultured according to its handling manual. Chemical DNA transfection was done using Qiagen SuperFect Transfection Reagent. Nucleofection

was carried out in a 4D-Nucleofector system from Lonza AG. All experiments were done in accordance with protocols from their original manufacturers. To induce Tet-On activation, Doxycycline from Clontech was added to culture media at 1ug/ml.

- **Creation of pCircuit-2 stable cell line**

300 ng pCircuit-2, 200 ng pCN and 6 ul Metafectene Pro (Biontex, Germany) were added to 70 ul of DMEM medium, incubated for 15 min. and added to 400'000 HEK293FT cell in 2 ml of DMEM medium. The supernatant was replaced the next morning and Hygromycin selection (200ng/ul) was started five days post-transfection and continued for the next 10 days. Subsequently, clonal cell lines were picked and expanded.

- **Creation of Bxb1 landing pad cell line**

The cell line was created by my colleagues Xavier Duportet, Dr. Patrick Guye, and was screened by Dr. Patrick Guye and Dr. Liliana Wroblewska. Briefly, a landing pad vector was integrated into HEK293FT cell genome by using CompoZr integration technology. The stable cells were selected with Hygromycin.

Chapter 3

3. Rapid Construction Method for Large Genetic Circuit

This Chapter describes the novel construction method developed herein for large genetic circuits consisting of multiple expression units.

3.1. Design of Complex Gene Circuit

3.1.1. Expression Unit

A complex gene circuit is built upon expression units. Each expression unit (EU) specifies how a gene will be expressed and regulated. The consortium of these EUs defines the complete function of a gene circuit. The combination of a promoter and a gene forms an EU."Promoter" refers to a non-coding sequence that can drive downstream expression. It may include not only a promoter sequence but also a promoter with a 5'UTR intron that contains important regulatory elements. Different promoters constitute a promoter library. Similarly, "gene" refers to a coding sequence or any other sequence that is expressed under the promoter. For example, cDNA, microRNA, and cDNA with 3' UTR regulatory elements are all under the name "gene" in this context. Different genes constitute a gene library and a gene chosen from the gene library can be paired with any promoter chosen from the promoter library. By specifying the EUs' function and their inter-relationship, a complex gene circuit is designed. To ensure that the gene circuit design is flexible and functional, not only do we leave the design flexibility at the EU level, but we also specify the fundamental structure of assembly so that EUs can be assembled together and behave reliably in mammalian cells. In this assembly method, an EU is embedded within a Basic Assembly Module (BAM).

3.1.2. Basic Assembly Module, Input Vector

A BAM consists of a chromatin-control related element (insulator), an EU, and a polyadenylation signal (polyAs) with a transcription termination sequence (Figure 3-1). These genetic elements are essential for reliable gene-circuit functions. A mammalian gene's expression can be affected by its genomic environment [85]. Firstly, inhibitory

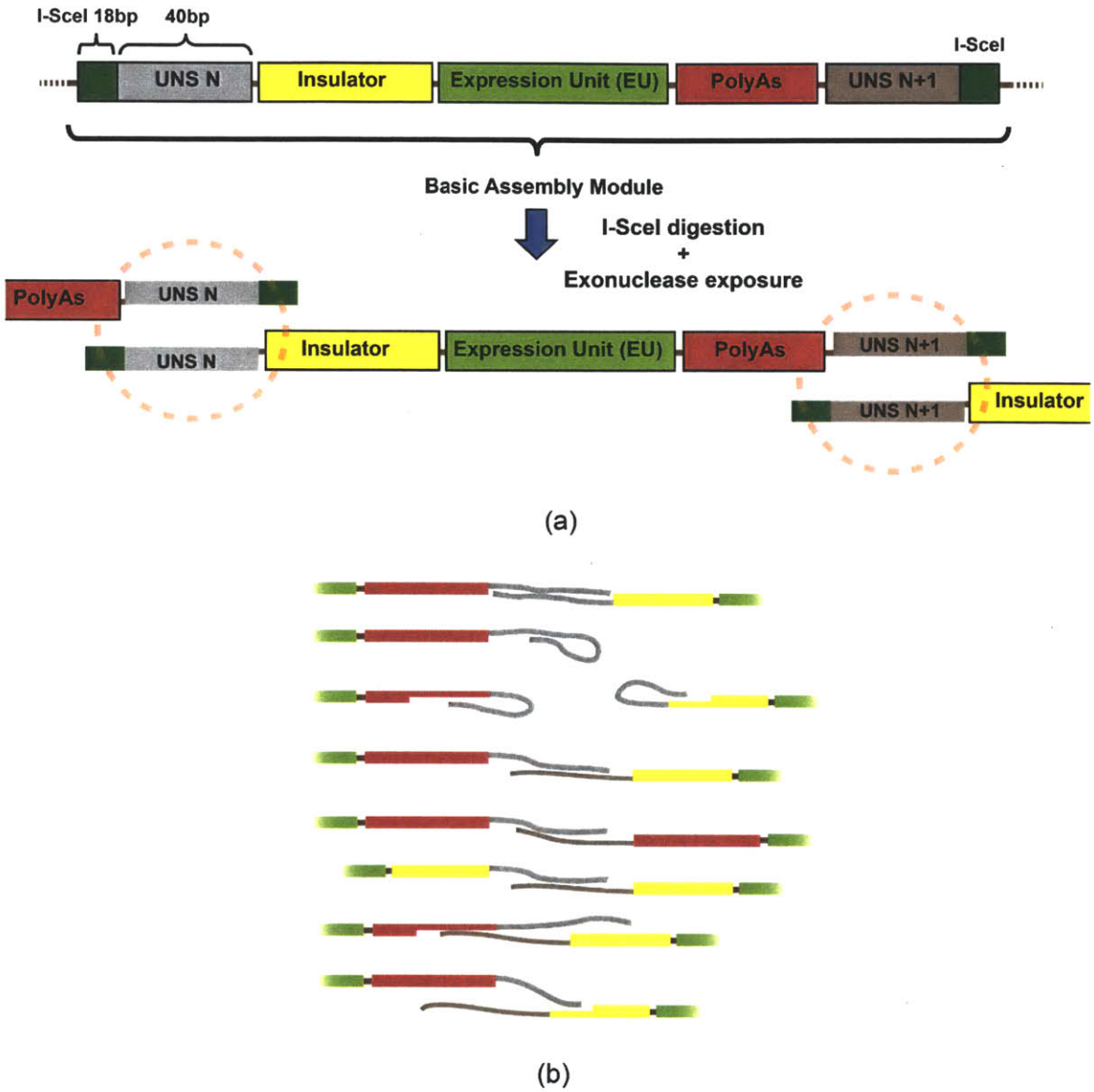


Figure 3-1: Basic assembly module and UNS mis-anneal

(a) A BAM consists of an insulator, an EU, and a polyadenylation signal (PolyAs) with transcription termination sequence. Unique nucleotide sequences (UNSs) guide BAMs assembly into a full circuit. (b) Assembly efficiency is affected by the quality of the UNSs; the annealing of non-complementary UNSs, and UNSs with other exposed single-stranded DNA results in faulty assembly. Unspecific annealing can occur in various ways including between different UNSs, UNS and the reverse complementary (rc) UNSs, UNSs and an insulator (exposed by exonuclease), UNSs and Polyadenylation Signal (exposed by exonuclease). Because the free leading head of UNSs can serve as a seeding region for unspecific annealing, even though the entire sequences may not have a high alignment score (Smith-Waterman), the alignment constraints are applied to the 7bp leading head on both sides of the UNSs.

chromatin can potentially shut down transgene expression regardless of its own regulation. Secondly, crosstalk from surrounding enhancers can exert additional regulations on transgene expression. Thirdly, transgene activation may be blocked by methylation. A chromatin-control insulator is included for reducing or minimizing interference between adjacent transcriptional units and methylation silencing. In this system, we selected the chicken DNase-I hypersensitive site 4 (cHS4) from β-globin locus as a chromatin insulator [86-90]. In addition to insulator, a polyadenylation signal is placed downstream of the EU. Most eukaryotic newly-made mRNAs are polyadenylated at 3' end during a process whereby polyadenylation signal is recognized by a protein complex and the mRNA tail is cleaved and poly(A) is synthesized at the end. The poly(A) tail is important for mRNA stability. We use a polyadenylation signal sequence taken from a Rabbit beta1-globin gene [91].

Framing the BAM on each side is a homing enzyme I-SceI restriction site and an unique nucleotide sequence (UNS) inside the I-SceI site. The homing enzyme sites enable cutting out the BAMs from their circular vectors by I-SceI. Unlike typical restriction enzymes, I-SceI recognizes a 18bp restriction site. Because it is so long, this site is rarely present in mammalian promoters and genes. The UNSs on BAMs are exposed in the assembly process and they anneal only to their complementary UNSs, so that the whole genetic circuit can be assembled reliably from compatible BAMs in a predetermined fashion.

In the assembly process, a promoter and a gene are recombined to form the EU within the BAM. The physical vector containing the BAM is called the position vector before the insertion of a promoter-gene pair and after EU is formed, the vector is called the input vector as it is the input to the subsequent circuit assembly (Figure 3-4).

3.1.3. Design of Unique Nucleotide Sequences

Unique nucleotide sequences (UNSs) guide BAMs' assembly into a full circuit (Figure 3-1). Unspecific annealing can happen between non-compatible UNSs, UNS and the reverse complementary (rc) UNSs, UNSs and insulator (exposed by exonuclease), UNSs and Polyadenylation Signal (exposed by exonuclease). Therefore, it was necessary to design UNSs such that the probability of any unspecific annealing other than between complementary UNSs is low enough for assembly of 8 or more BAMs simultaneously. Computationally, the Smith–Waterman local alignment score was used as a measure for the annealing probability. The probability of unspecific annealing is proportional to the Smith–Waterman local alignment score of the involved sequences - the higher the alignment score, the higher the probability that two sequences will anneal.

```
procedure Stochastic Local Search for UNS Design
  input: Number of UNS, length of UNS, search pool size, set of constraints (C), crossover rate
         mutation rate λ
  output: Set of UNS that satisfy C
  S := ∅
  P := initial random pool of candidate UNSs
  for l := 1 to maxIterations do
    for i := 1 to maxTries do
      for j := 1 to size(P) do
        if P(j) satifsfies all constraints then
          add P(j) to S
          if S contains enough UNSs then
            return S
          end if
        end if
        for k := 1 to size(S) do
          AF(k) is increased by the number of alignments fails for UNS(k) ∈ S in current iter
        end for
        θ := inverse proportional to the # of constraints that candidate ∈ P violate
        for k := 1 to size(P) do
          select cU1, cU2 ∈ P with probability θ respectively
          ĉU := new candidates from crossover(cU1, cU2, β)
          add P̂ new candidates from ĉU by mutating bases with rate λ
        end for
        if P̂ has indentical candidates within P ∪ P̂ then
          remove identical candidates from P̂
          add random candidate to fill P̂
        end if
        P := P̂
      end for
    end for
    remove UNS ∈ S that has maximum AF for all UNSs ∈ S
  end for
  return S
end
```

Figure 3-2: Stochastic local search algorithm for iterative UNSs searching

Because the free leading ends of UNSs can serve as the seeding region for unspecific annealing, even though the entire sequences may not have high alignment scores, alignment constraints were applied to the 7bp leading head on both sides of UNSs

In addition to avoiding unspecific annealing, several other constraints needed to be met: Firstly, it was necessary to ensure that all complementary UNSs annealed easily within a certain temperature range. This condition was determined by the annealing temperature of UNSs. Secondly, secondary structure within the UNS had to be minimized. Secondary structure of UNSs can affect correctness and efficiency of annealing between two matching sequences and, hence, the efficiency of assembly. The GC content percentage of an UNS was set to be within a certain range (Table 3-1). Thirdly, the UNS needed to be free of some restriction sites so as to allow easy modification to the assembly vectors by restriction cloning. Table 3-1 lists the constraints used in the design of UNS and the parameters used for UNS searching.

Parameter Constraints	**Value**
Length of UNS	40
Length of UNS head region tested in alignment	7
Smith-Waterman alignment threshold for whole sequence	50
Smith-Waterman alignment threshold for sequence head region	25
Restriction sites	AgeI, XbaI
Secondary structure dG	-5 kcal/mol
Annealing temperature	65C - 75C
GC content percentage	40% - 80%
Crossover rate	0.9
Mutation rate	0.1
Random sequence pool size	10

Table 3-1: Parameters and constraints used for UNS design and searching

A crucial element of this research was the design of a stochastic local search algorithm for iterative UNSs searching and optimization. In each iteration, the algorithm starts with a pool of random sequences. Each sequence from the pool is aligned to the UNSs already found. The best sequence, with alignment scores lower than a threshold, is screened for its secondary structure, melting temperature, and restriction sites. If it meets all constraints, it is added as a new UNS to the pool of existing UNSs. The algorithm repeats the process until a desired number of UNSs are found. If no sequence from the random pool meets all requirements, a few sequences with better alignment scores are picked. A new search pool

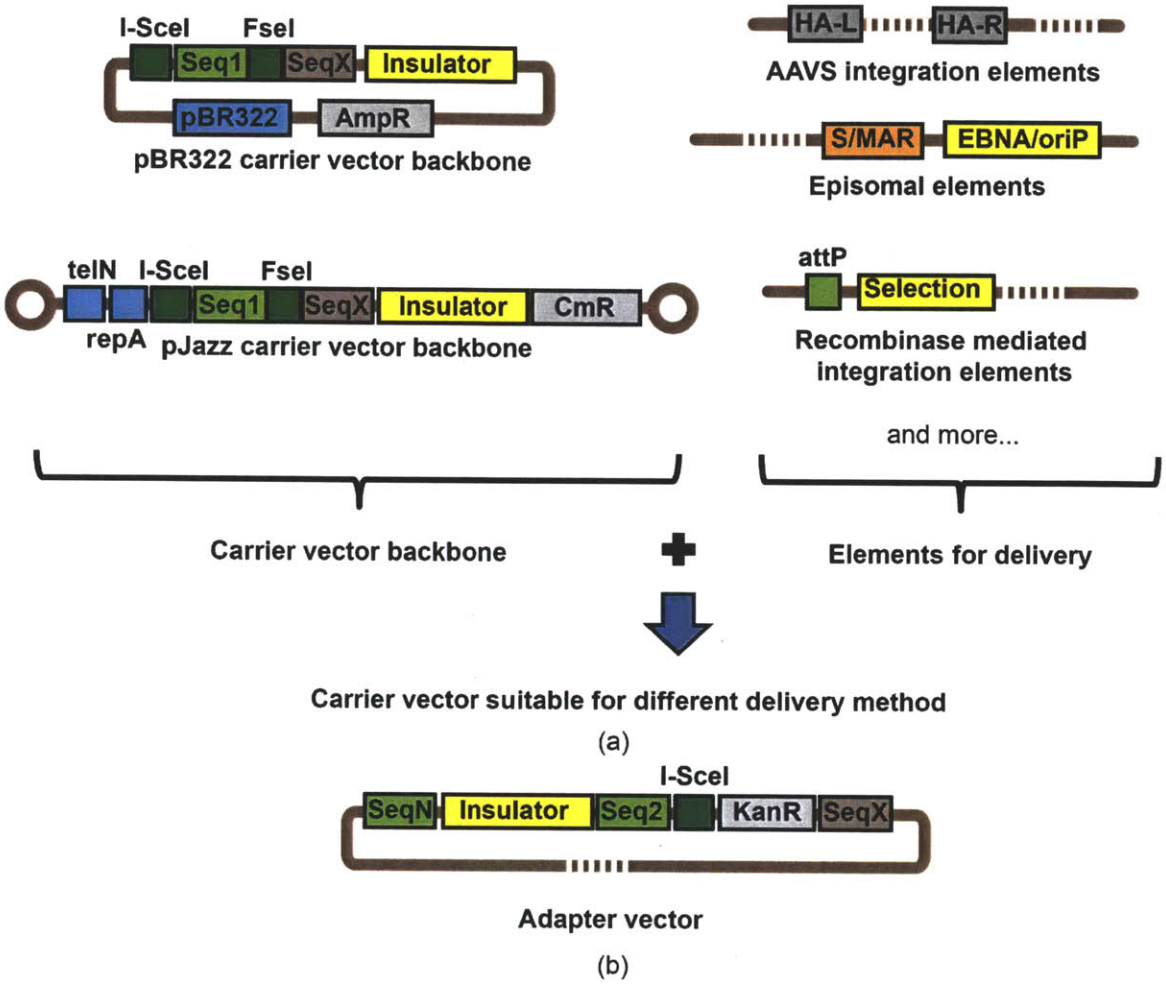


Figure 3-3: Carrier vector and adapter vector

(a) Carrier vector has Seq1-FseI-SeqX and components for replication, selection, and elements specific for different delivery methods. (b) Adapter vector contains SeqN, SeqX as well as second selection marker.

is then generated by recombining and mutating those selected sequences, and by adding newly generated random sequences, a process similar to a genetic algorithm. Subsequently, the search procedure repeats. If the iteration number exceeds a maximum, it is assumed that some of found UNSs are too restrictive for addition of a new sequence. In this case, the UNS that failed for most alignment-test failures will be removed, and a new search process started. A complete UNS design algorithm is shown in (Figure 3-2). 13 new UNS were designed with this algorithm. The quality of the design sequence is measured by the ΔG gap, the smallest energy gap between any pair of matching UNSs and mis-matching UNSs [92-94]. The ΔG gap for the designed sequence is -32 kcal/mol. In the current implementation, the first 9 UNSs were chosen and named Seq1 to Seq8 and SeqX.

3.1.4. Carrier Vector

A carrier vector has three roles: (1) to allow BAMs to assemble into it; (2) to recover the fully assembled gene circuit; (3) to assist in gene circuit delivery (Figure 3-3). Every carrier vector has a Seq1-FseI-SeqX cassette. FseI is chosen because it is only present once in the carrier vectors. FseI cuts open the cassette allowing exposure of UNS, Seq1 and SeqX. The carrier vector takes in all assembly inputs with Seq1 and SeqX annealed to the complementary UNSs on BAMs. Because the assembly reaction produces various intermediates, the fully assembled circuit needs to be selected and propagated to provide a large yield. Carrier vectors based on different replication mechanisms have distinct abilities to recover large gene circuits containing repetitive elements (e. g. chromatin regulatory elements, and polyadenylation signals on every input vectors). Firstly, the plasmid copy number affects large DNA molecules' stability [95], [96]. A high-copy-number plasmid is unsuitable for cloning large DNA fragments or DNA regions with secondary structure because inter-plasmid recombination is likely to happen, a result that would lead to loss of intact genetic circuit. On the other hand, a low copy number plasmid has low yield, meaning that significantly more bacterial culture is required to produce enough DNA for delivery. Secondly, the supercoiled status of the carrier vector is important. Plasmid supercoiling can induce secondary structures, bring homologous sequences close, favoring intra-plasmid sequence deletion and rearrangement [97], [98].

Prophage of coliphage N15 replicates as a linear dsDNA molecule with closed ends generated by a phage enzyme termed protelomerase (TelN). The linear cloning vector pJazz based on the phage genome showed improved stability for many types of DNA sequences, including AT-rich sequences, and sequences with short tandem repeats [97]. This study used carrier vectors based on pBR322 and pJazz.

In addition to differing in replication mechanism and selection markers, carrier vectors differ in the backbone sequences specific to distinct delivery methods (Table 3-2).

Delivery methods	Elements put on carrier vector backbone
Transient transfection	No additional sequence
Episomal persistence	EBNA/oriP or S/MAR transcription unit
Homologous recombination, e.g. CompoZr	Homologous sequences, e.g. AAVS locus
Recombinase based integration, e.g. Bxb1	Recombination site, e.g. Bxb1 attB site
Minicircle, e.g. minicircle mediated by Bxb1	Recombination site, e.g. Bxb1 attR site

Table 3-2: Carrier vectors contain genetic elements for distinct delivery methods

3.1.5. Adapter Vector

To assemble a full circuit, the UNSs on each BAM need to form a closed loop. An adapter vector containing the necessary UNSs to form a closed loop is added (Figure 3-4). For example, in order to construct a genetic circuit with 6 BAMs, an adapter vector which has Seq7 and SeqX is added, while the carrier vector always contains SeqX-Seq1. This adapter also helps increase the assembly efficiency. It is frequently observed that cut-open carrier vectors can re-ligate, which causes a significant drop in assembly efficiency. An adapter that has the second selection marker, e.g. Kanamycin resistance, is used with a second antibiotic, e.g. kanamycin, preventing the re-ligated carrier vector from propagating.

3.2. Construction Flow

3.2.1. One Step Assembly

A complete genetic circuit construction flow is shown in Figure 3-4.

The Mammalian promoters and the genes of interest are built into Gateway entry vector libraries and are sequence verified. To construct a full gene circuit, promoters and genes forming EUs are first picked up from libraries and then recombined into BAMs. The same promoters or genes can be reused in different promoter-gene pairs in a modular fashion. The recombination step is usually done in a Gateway LR reaction. As mentioned in Chapter 1, this is a recombinase based cloning step which does not involve a mutation-prone process. If the gene circuit to be built contains EUs that have been already constructed, input vectors harboring those EUs can be reused.

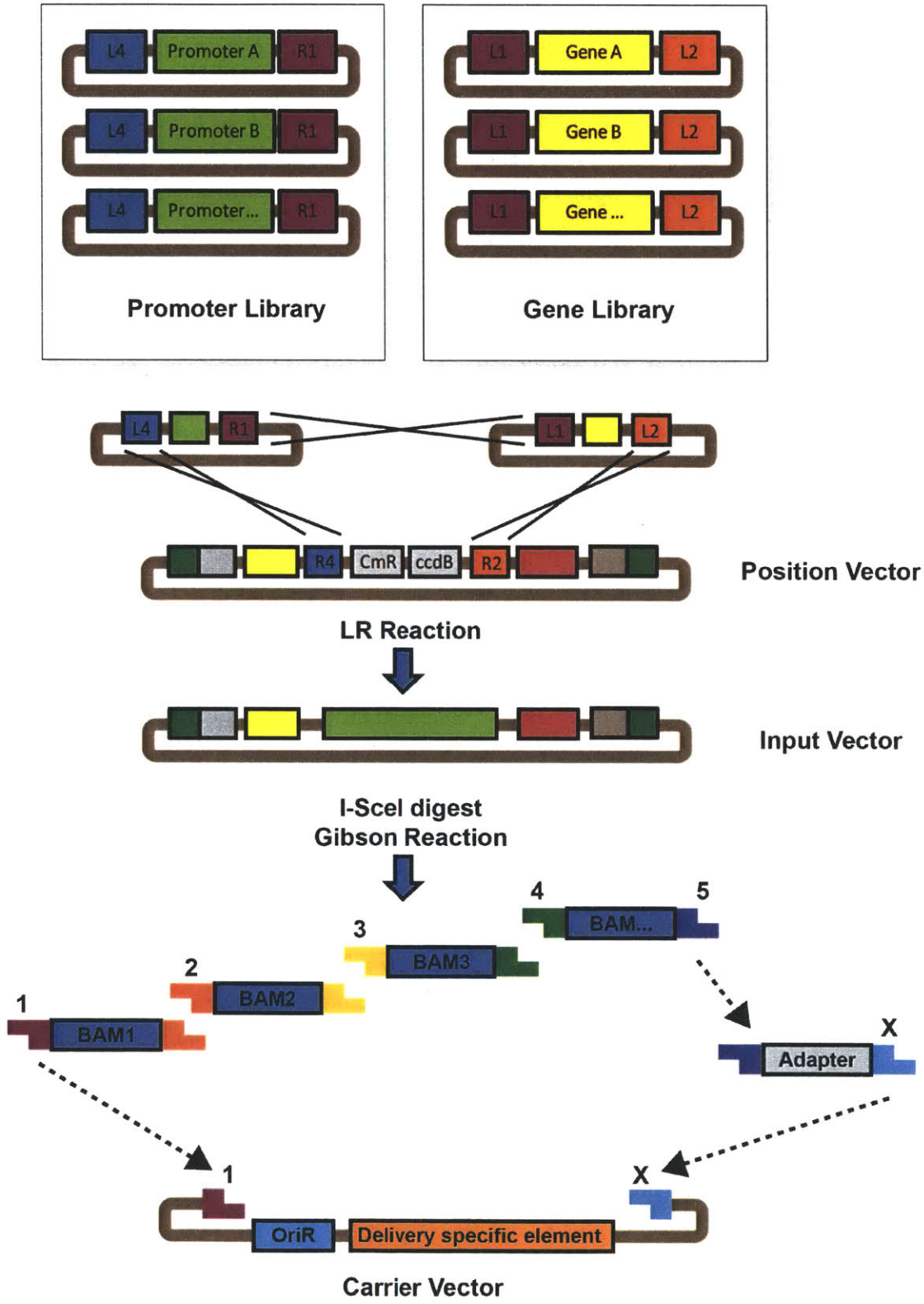


Figure 3-4: Assembly construction flow

In the second stage of assembly, a particular carrier vector is selected depending on how the circuit is to be delivered. This flexibility allows quick circuit tests of various delivery methods. Input vectors are digested with I-SceI to release BAMs. The chosen carrier vector and adapter are digested, leaving UNSs on the ends. Thereafter, two approaches can be applied to assemble the whole circuit in a single step. In one implementation, the I-SceI site residue and exposed UNSs on input vectors are removed by exploiting a special property of the residue site. The site sequence only contains three different types of nucleotides, namely C, A, T. Double G are placed between the I-SceI site and the UNS, and when the I-SceI digested BAMs are subject to T4 exonuclease in presence of only dGTP, 3' ends of DNA are chewed up to the GG leaving the UNS intact but removing the I-SceI site completely. After purifying the samples, assembly is carried out in a Gibson reaction (with non-error correcting polymerase) where UNSs are exposed, joined, and the dsDNA are repaired [48], [49]. In another implementation, the Gibson reaction is run directly on the digested inputs. The error-correcting polymerase in Gibson reaction mix can act as an exonuclease in 3'-5' direction removing the I-SceI residue site.

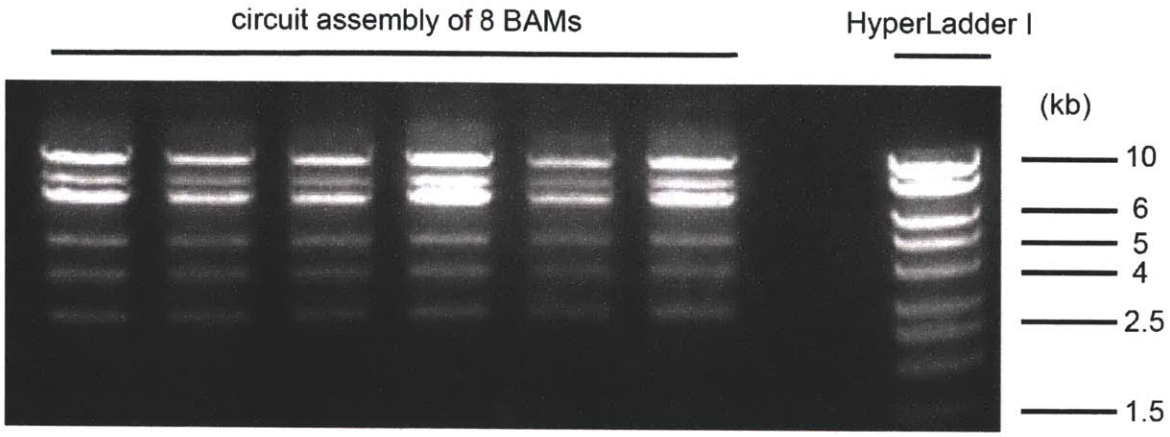
3.2.2. Hierarchical Assembly

In this research implementation, 7 position vectors were constructed resulting in the ability to assemble genetic circuits containing up to 7 EUs, corresponding to 14 gateway entry vectors. One way of going beyond 7 EUs is to construct more position vectors, which allow assembly of larger constructs in one-pot. While simultaneously combining multiple BAMs in a single step provides a high value, construction techniques also benefit from hierarchical assembly protocols that allows nested or multiple levels of construction from intermediate assemblies.

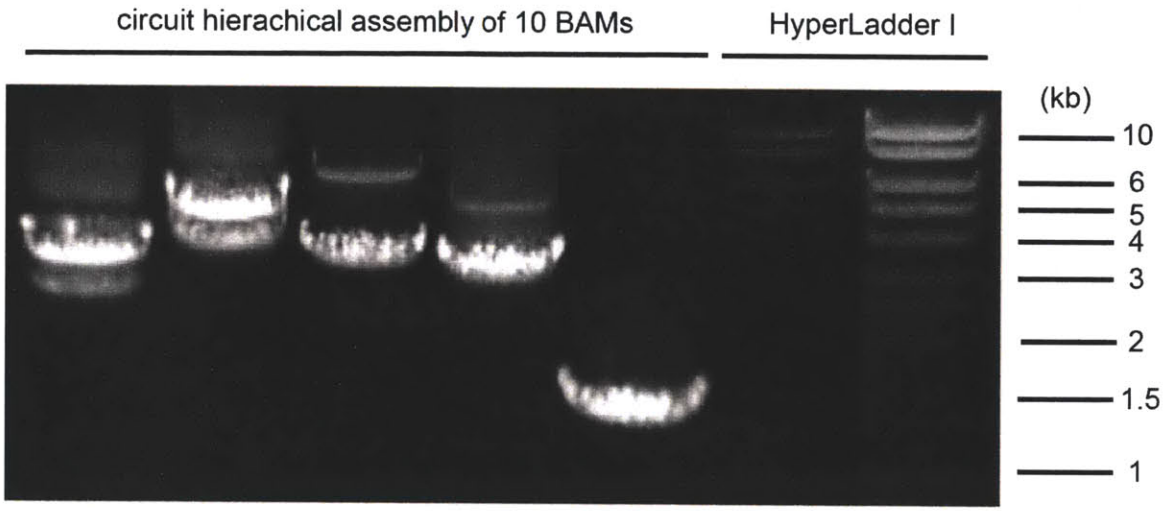
Hierarchical assembly reuses assembled circuits as inputs to nested assembly. An assembled circuit can be converted to a large assembly input flanked by Seq1 and Seq2. To facilitate this, one I-SceI site is placed upstream of Seq1 on the carrier vector and another I-SceI is placed downstream of Seq2 on the adapter vector (Figure 3-4). The assembled gene circuit can be readily digested by I-SceI to release the assembled BAMs and be used as an input filling in the first position in a nested assembly.

3.3. Assembly Efficiency Measurement

The assembly efficiency was measured by determining the ratio of bacterial colonies containing the correct vector out of the total screened colonies (Figure 3-5). Colonies were picked randomly and grown for 16 hours. Following the pelleting of the bacteria, the cells



(a)



(b)

Figure 3-5: Restriction digestion and PCR analysis of assembled circuit

The assembly efficiency was measured by determining the ratio of bacterial colonies containing the correct vector out of the total screened colonies. In cases where the restriction bands are too numerous to distinguish the correct ones from the incorrect ones, the sample was amplified with a set of primers designed for each BAM. (a) Assembly of 8 BAMs was digested with XbaI. Ethidium bromide stained gel of 8 minipreps showed the same restriction pattern. (b) Hierarchical assembly of 10 BAMs was amplified with a set of primers. Expected bands are 2653 and 3564 in lane 1, 3564 and 4757 in lane 2, 3220 and 6562 in lane 3, 5218 and 3218 in lane4, 1509 in lane 5. Hyperladder I lane was duplicated on the side and adjusted for light and contrast in Photoshop (Adobe System Inc.)

were lysed and plasmid DNA was extracted. Recovered plasmid DNA was digested, run on an agarose gel and checked for its restriction pattern. In cases where the restriction bands were too numerous to distinguish the correct ones from the incorrect ones, the sample was amplified with a set of primers designed for each BAM. The PCR products were run on an agarose gel and checked for their size. Correct minipreps were those that produced the expected restriction mapping pattern.

Using a library of sequence-verified parts, 8 input vectors were constructed. Assemblies were performed with an increasing number of input vectors to determine the assembly efficiency in relation to the number of assembled BAMs (proportional to circuit size). It was found that the addition of T4 exonuclease treatment consistently yielded lower efficiency. Therefore, the step for I-SceI residue removal was skipped. Gene circuits were assembled that were up to 43kb in size and composed of 7 EUs with almost 100% efficiency. The results are summarized in Table 3-3.

Assemblies using the pBR322-based carrier vector exhibited a decrease in efficiency for larger circuits (large number of parts and large size). It is described in the literature that the pBR322 backbone does not have the capacity for carrying large inserts (size > 20 kilobases) in stable fashion. To confirm this, a 25kb pBR322 plasmid was transformed back into bacteria, but it was not possible to recover any plasmid with the correct restriction pattern.

Several assembled vectors exhibiting incorrect restriction patterns were sequenced. These errors were rarely due to mismatching UNS, but rather due to annealing of sequences far upstream or far downstream of UNS. It is possible that the synergetic reaction of T5 exonuclease and error-correcting polymerase might create these joints. In this model, T5 exonuclease chewed up DNA further. Any mismatching annealing between extended exposed single strands would be an anchor place for the error-correcting polymerase. Its 3'-5' exonuclease activity would remove non-matching single strands and allow the dsDNA formation to complete. Since the probability for annealing of all complementary UNSs decreases with increasing number of inputs (BAMs), it is expected that more assembly failures like this type would happen with a larger circuit.

Next the hierarchical assembly using the linear pJazz-Swing carrier vector was tested. The above mentioned hierarchical assembly protocol was applied to a correctly assembled gene circuit in the pJazz-Swing carrier. Briefly, an assembled circuit was digested with I-SceI and purified using the Qiaquick PCR columns. It was then pooled with the other input vector, a carrier vector and an adapter as described for the non-hierarchical assembly. Following addition of the Gibson Assembly Master Mix and incubation for one hour at

Carrier vector			pBR322		pJazz		pJazz hierarchical	
parts	hierarchical	#inputs		size (bp)		size (bp)		size(bp)
1-3-X		3	12/12	8488				
1-4-X		4	3/3	16652				
1-5-X		5	10/12	23695	5/6	31664		
1-6-X		6	9/10	25160	5/5	35527		
1-7-X		7	5/7	32737	6/6	42280		
1-8-X		8	6/10	33709	9/10	44218		
(1-4-X) - 5-X		9	2/5	33337	5/5	43829		
	(1-6-X) - 3-X	9					5/6	38780
	(1-6-X) - 4-X	10					5/7	44448
	(1-6-X) - 6-X	12					5/10	54048
	(1-6-X) - 8-X	14					1/10	61392

Table 3-3: Assembly efficiency

Assembly efficiency measured by the fraction of the number of colonies that have the expected restriction pattern out of all colonies screened. 1-3-X means that 2 EUs and 1 adapter are assembled. (1-6-X) means 5 EUs and 1 adapter are assembled and the assembled circuit is reused as the first position in a nested assembly. x/y notates the screening. x: number of colonies that have the expted restriction patter. y: total number of colonies screened

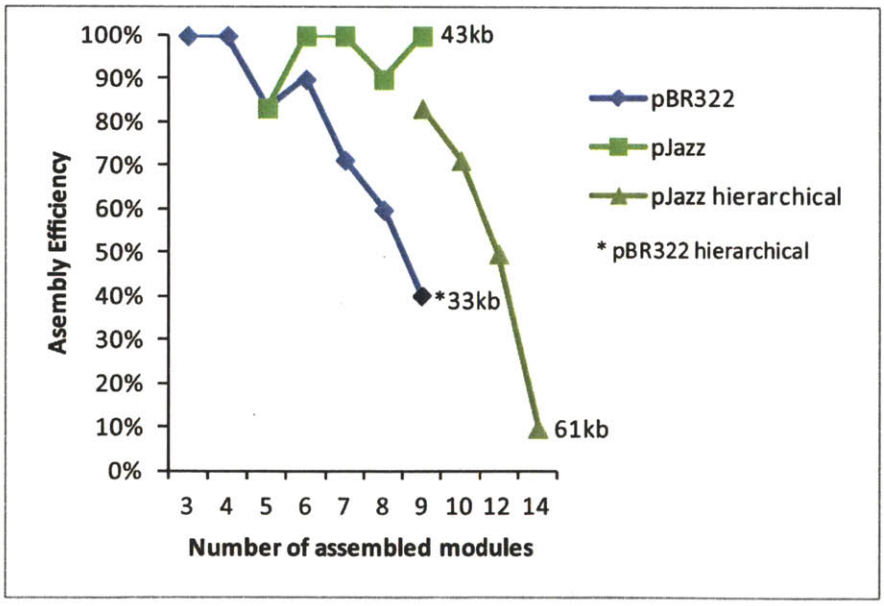


Figure 3-6: Assembly efficiency versus circuit complexity

Assembly efficiency data from Table 3-3 is plotted against the size and the number of assembled modules.

50C, 2 ul were transformed into competent *E. coli* cells. Colonies were randomly picked and plasmids were extracted after 16 hours of incubation.

A 61kb circuit consisting of 12 EUs was successfully assembled. From Figure 3-6, it can be seen that efficiency dropped quickly for hierarchical. In one possible failure mode, the original backbone from the first assembly is re-ligated to form the original assembled sample. Although the chance is low, it could be comparable to the probability of having all complementary UNSs perfectly annealed to their complementary UNSs when a large number of BAMs are to be assembled. In the hierarchal assembly, products with re-ligated samples were indeed found. Another probable reason for low large assembly efficiency is DNA damage during the purification procedure. The larger the circuit, the higher chance it will break during purification. This view is partly supported by the fact that assembly efficiency with gel extraction is much worse than with PCR purification which skips gel electrophoresis. The NucleoTraPCR purification kit, which uses silica beads and supports a larger sample size than the silica membrane used in PCR purification, was tried, however use of this kit resulted in more re-ligated samples. Presumably, this is because it has higher recovery rate for small pieces, such as the carrier backbone from the first assembly, than the large assembled BAM. If the backbone is not removed, it could lead to more re-ligated products of the first assembly in the nested assembly.

Chapter 4

4. Delivery Method for Large Genetic Circuit

This chapter describes and evaluates three different methods for introducing large gene circuits into mammalian cells.

4.1. Transient Transfection

Starting with transient transfection, the goal is to compare the transfection of an assembled circuit with co-transfection of several plasmids. The test circuit pCircuit-1, shown in Figure 4-1, consists of both constitutive EUs as well as inducible EU to mimic a real application circuit. The human elongation factor 1-a (hEF-1a) promoter constitutively expresses Enhanced Blue Fluorescent Protein (EBFP). The reverse tetracycline transactivator (rtTA3) and the Enhanced Yellow Fluorescent Protein (EYFP) are constitutively expressed under another hEF-1a promoter. The fluorescent protein TaqFP635 (mKate) is expressed under a tetracycline response element (TRE) promoter. When Doxycycline (Dox) is absent, the TRE promoter is inactive and no mKate is expressed. When Dox is added, rtTA3 binds the TRE promoter, which then drives mKate expression. The circuit was assembled and transfected into HEK293 cells. In comparison, the input vectors from which the test circuit was assembled were also co-transfected into cells.

Transfection of the assembled circuit (AC) and co-transfection of input vectors (CT) showed different delivery rates. This was determined by the percentage of the cell population that co-expressed the two constitutive markers, EYFP and EBFP (Figure 4-1). The CT showed higher transfection rate than the AC: 66.8% versus 39.4%. This suggests that small plasmids (<9kb) are easier to deliver than a large construct (17kb). To characterize the cell population that actually received the gene circuit, the cell population without EYFP nor EBFP expression was removed from the total cell population in the analysis. The maximum and mean intensity of EYFP, EBFP, and mKate were comparable for both cases. It is interesting to note that the average intensity of mKate was slightly higher in the AC than in the CT despite the fact that the average intensity of EYFP in AC was only half of that in CT. The EYFP was bicistronically expressed with rtTA3, which

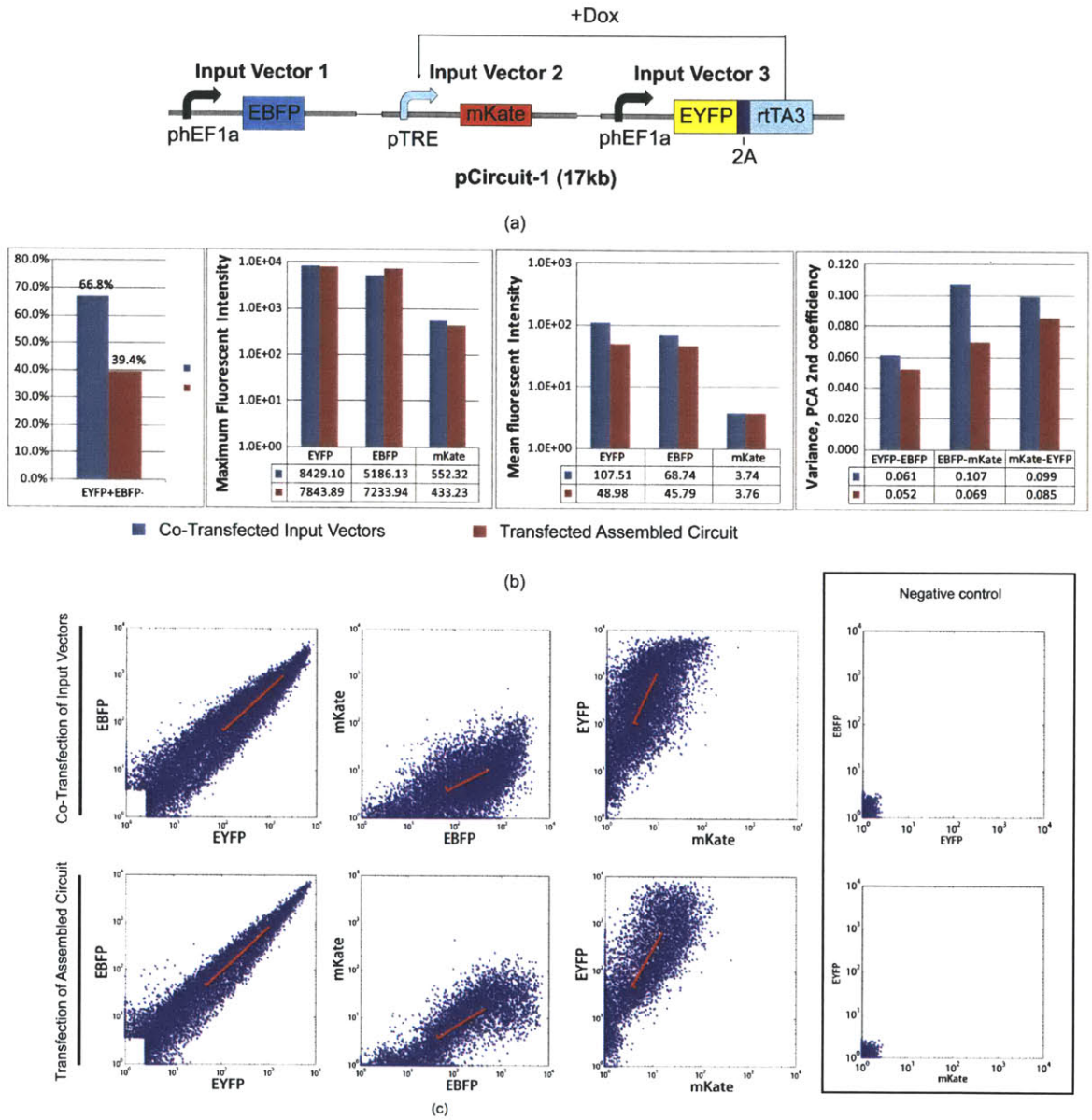


Figure 4-1: Transient transfection of assembled circuit

(a) The test circuit consisted of both constitutive EUs as well as inducible EU. EBFP, rtTA3 and EYFP are constitutively expressed under hEF-1a promoters. When Dox is added, rtTA3 binds TRE and drives mKate expression. (b) From left to right: The percentage of cell population that co-expressed the two constitutive markers, EYFP and EBFP; The maximum and mean intensity of EYFP, EBFP, and mKate; The strength of the PCA second component. (c) Cell population that has neither EYFP nor EBFP expression was removed from the total cell population. PCA was calculated on color pairs and its two components were drawn (red lines) with the length proportional to their strength respectively.

drove expression of mKate from the TRE promoter. This result suggests two possibilities: (1) the mKate input vector was transfected at much lower efficiency than rtTA3 input vector in the CT, so that the availability of mKate plasmid was the limiting factor; (2) rtTA3 was abundant but inducer Dox or VP16 activation machinery was the limiting factor. In either case, the AC showed an advantage over the CT in that it achieved a similar output (mKate) level without placing an extra burden on cells by resorting to over-production of rtTA3.

Next the correlation between each EU was investigated. It can be seen that each color pair showed good correlation in both the AC and the CT. However, the AC exhibited less variance than the CT. Principal component analysis (PCA) analysis was performed on each color pair. The first and second PCA components are drawn in Figure 4-1 with components' length proportional to their strength and it was clear that the CT showed larger variance (the strength of the second PCA component). The strength of the second component represented heterogeneity of gene expression. The large variance can be attributed to the stochastic stoichiometry of transfected input vectors in the CT in addition to the stochastic transcription-translation process.

4.2. Homologous Recombination Integration

Stable cell lines offer many advantages over transient transfected cell lines. The integrating circuit pCircuit-2 is shown in Figure 4-2. The red fluorescent protein mKate is constitutively expressed under a hEF-1a promoter. Activator rtTA3 coupled with a Hygromycin resistance (HygroR) are under a constitutive UbC promoter. The activator rtTA3 is used for TRE promoter induction and HygroR is an antibiotic resistance gene used for stable cell line selection. TRE promoter drives expression of EGFP upon Doxycycline induction. Stable cell lines were created using CompoZr integration technology [68]. Briefly, the circuit was assembled into a CompoZr integrating carrier vector that has an AAVS homologous sequence on its backbone. Integration of the circuit was mediated by ZF nuclease dimer that targeted AAVS integration site with high specificity.

Stable cell lines were selected with Hygromycin. Microscopy images are shown in Figure 4-2. mKate expression is constitutive and homogeneous. EGFP expression showed a distinct difference before and after Dox was added, indicating the tightness of TRE promoter regulation in the integrated assembled circuit. Notice that the EGFP expression is heterogeneous compared with mKate expression. This could be due to the stochastic nature of transcription on TRE promoter. One TRE promoter has 7 rtTA3 binding sites. Random rtTA3 binding events at these 7 sites may lead to the observed heterogeneity [99].

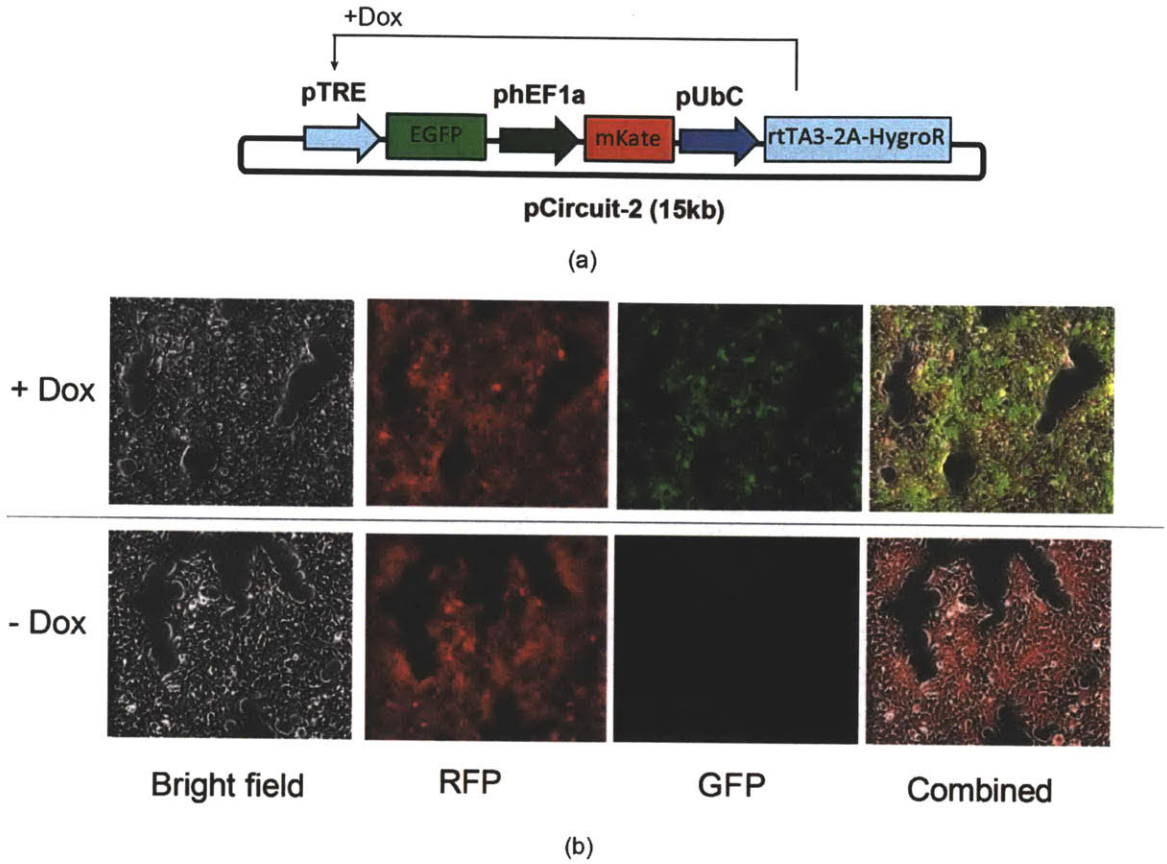


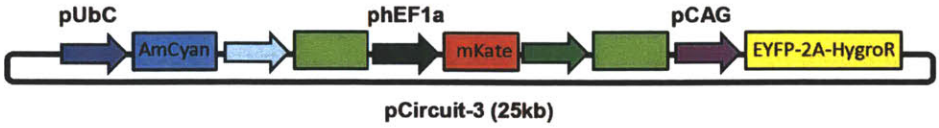
Figure 4-2: Integrate circuit using homologous recombination and ZF nuclease

(a) pCircuit-2 consists of three EUs. The red fluorescent protein mKate is constitutively expressed under hEF-1a promoter. Activator rtTA3 coupled with Hygromycin resistance (HygroR) are under constitutive UbC promoter. Activator rtTA3 is used for the TRE promoter induction and the HygroR is an antibiotic resistance gene used for stable cell line selection. TRE drives expression of EGFP upon Doxycycline induction (1ug/ml). The pCircuit-2 was assembled into the CompoZr integrating carrier vector that has AAVS homologous sequences on its backbone. (b) The pCircuit-2 was integrated into AAVS locus in HEK293 cells. Stable cell line was selected with Hygromycin. A single colony was expanded. Cells were split into two dishes. Fluorescent images were taken after one dish of cells was induced with Dox. No Dox was added to the control dish.

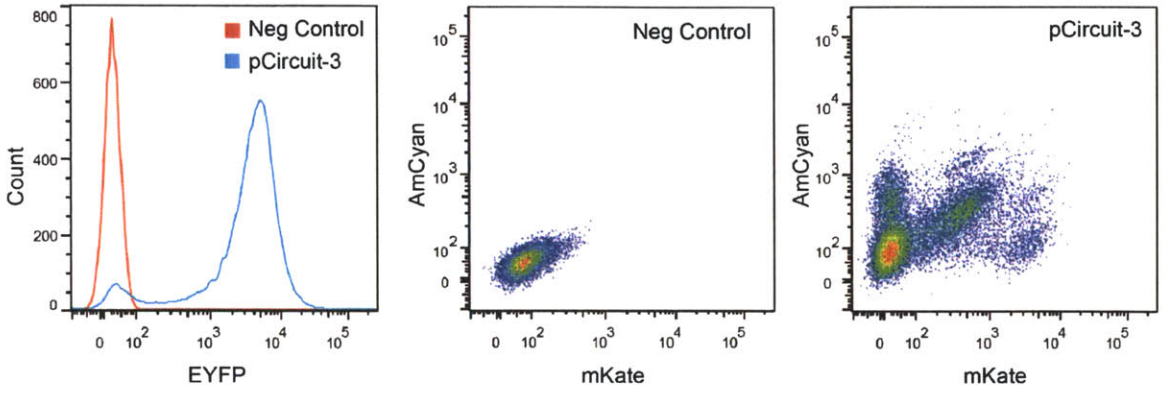
To test if stable cell lines could be selected after large circuit integration, a circuit with five modules and a size of 25kb was constructed, as shown in Figure 4-3. The circuit allowed Hygromycin selection and has three fluorescent reporters. We assembled the circuit into the CompoZr integrating carrier vector and integrated the circuit into HEK293FT cells.

Stable cell lines were selected on Hygromycin. Expectedly, the majority of cells expressed EYFP, which was bicistronically coupled to Hygromycin selection. However, surprisingly the fraction of cells that were positive for all constitutive color-markers comprised only a small portion of the total population (Figure 4-3). Non-fluorescent cells were removed from the analysis and the double-color positive (DCP) fraction was calculated for all fluorescent cells. This measure captures the imperfection of the circuit for cases where some EUs are working while others are not. CompoZr integration exhibited a much lower DCP fraction than transient transfection. This suggests that random integration or partial integration might be the causes and rules out other factors like the stochastic noise from transcription and translation. The CompoZR system utilizes HR for vector integration. Both HR and NHEJ could mediate (1) integration of circuit fragments; (2) random integration of circuit into silenced genomic loci; (3) both. The physical damage to circuit DNA during transfection or nucleofection [100] could exacerbate the case. Besides, gene silencing from methylation and heterochromatin could lead to loss of color expression.

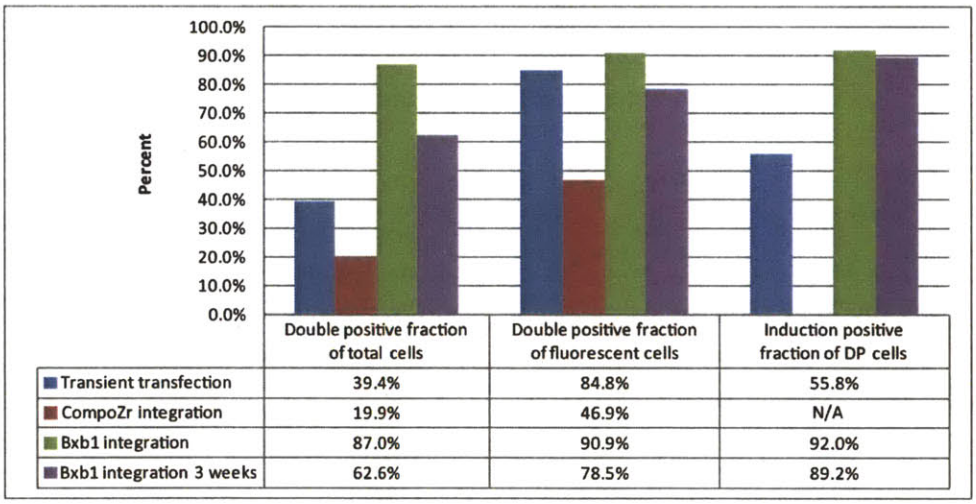
The experimental results suggested that in the case of large gene circuits integration via HR, it may be difficult to select a stable cell line that contains functioning circuit with just a single selection marker. The lower the percentage, the more clonal cells need to be screened for a one that contains a functional circuit in the specified locus. This means that screening for the presence of EU that is not coupled to any fluorescent marker or selection marker may prove to be difficult.



(a)



(b) (c)



	Double positive fraction of total cells	Double positive fraction of fluorescent cells	Induction positive fraction of DP cells
Transient transfection	39.4%	84.8%	55.8%
CompoZr integration	19.9%	46.9%	N/A
Bxb1 integration	87.0%	90.9%	92.0%
Bxb1 integration 3 weeks	62.6%	78.5%	89.2%

(d)

Figure 4-3: Integration of large genetic circuit using HR and ZF nuclease

(a) The AmCyan, mKate, and EYFP are constitutively expressed. EYFP is bicistronically coupled to Hygromycin Resistance gene. Two inducible EUs are placed between constitutive EUs. (b) Stable cell lines were selected on Hygromycin. There were 93.2% of total polyclonal cells expressing EYFP. (c) A large population only expressed either AmCyan or mKate but not both. (d) Statistics of the fraction of cells expressing two constitutive fluorescent markers (EYFP+EBFP+ for transient transfection (TT) test and Bxb1 integration (Bxb1) test, AmCyan+mKate+ for CompoZr integration test). In the second column, non-fluorescent cells were removed (EYFP-EBFP- for TT test and Bxb1 test, AmCyan-mKate- for CompoZr integration test). Induction refers to mKate expression upon addition of Dox.

4.3. Bxb1 Landing Pad Integration

In order to minimize random integration recombinase Bxb1 mediated integration was tested. A genetic circuit similar to pCircuit-1 was assembled into the Bxb1 integrating carrier vector instead of the CompoZr integrating carrier vector (Figure 4-4). The carrier vector contains Bxb1 attB site followed by a promoter-less Puromycin resistance (PuroR). The targeted cell line contains a Bxb1 landing pad in the genome. The landing pad has a constitutive promoter followed by a Bxb1 attP site and EYFP-2A-HygroR. Once inside the cell, the Bxb1 attB site on the carrier vector recombines with the Bxb1 attP site on the genome, thereby inserting the genetic circuit into the cell genome. PuroR then replaces EYFP-2A-HygroR and is expressed under the constitutive promoter. A stable line with successful integration can be selected with Puromycin.

Results showed that the EYFP/EBFP double positive (DP) cell population comprised 87% of the total population. Out of those DP cells, 92% cells were expressing inducible mKate. Of importance is that no color was linked to a selection marker (Figure 4-4). This suggested that clonal cell lines with intact circuit integrated could be easily screened out with a single selection marker. As for the remaining 13% cell population, gene silencing from methylation and heterochromatin spreading could account for the color loss. Non-specific integration and partial integration are also possible causes, which need to be further tested.

To test the circuit's long-term performance, polyclonal cells were cultured for an additional 3 weeks under Puro selection and then FACS analyzed. The fraction of DP cell population dropped from 87% to 62% and the mKate induction rate dropped from 92% to 89.2% (Figure 4-5). It was not tested whether this decrease came from the population overtaken by the non-DP cells or from long term gene silencing.

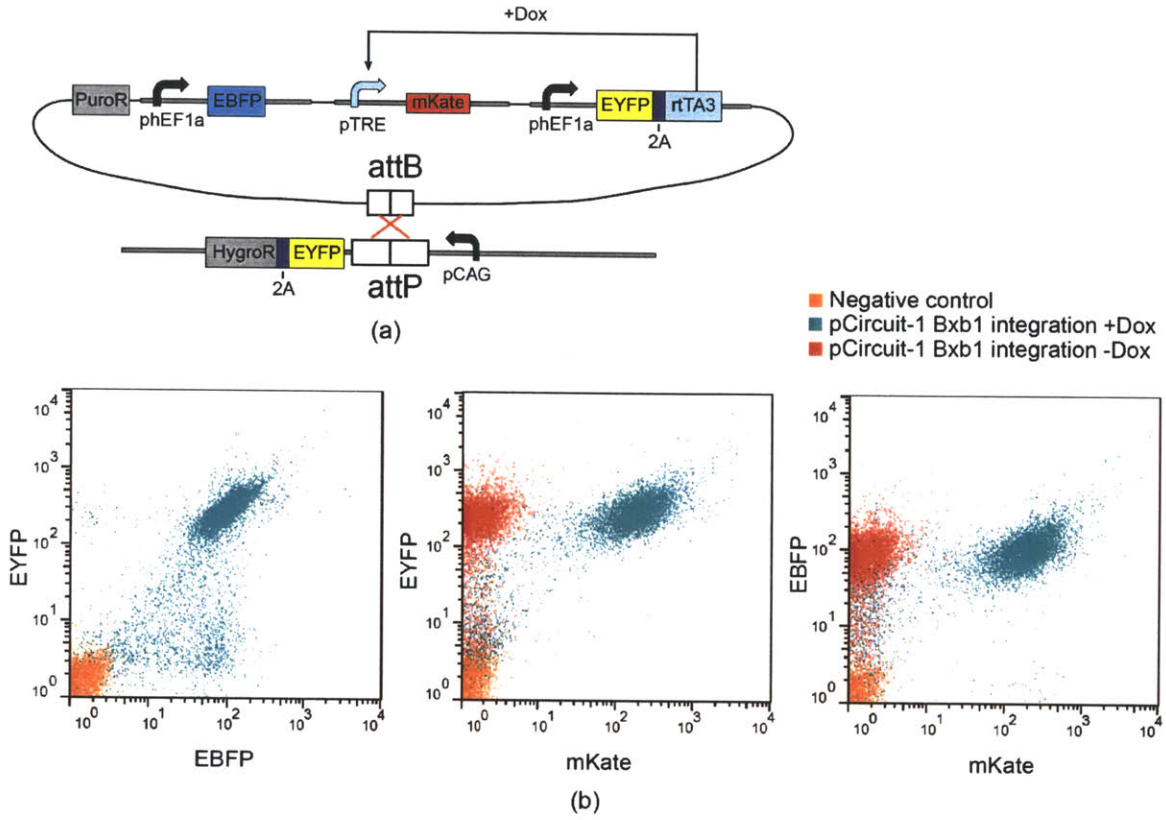


Figure 4-4: Integration of circuit using Bxb1 recombinase and landing pad cell line.

(a) The test circuit is similar to the pCircuit-1 except that it was assembled into the Bxb1 integrating carrier vector. The Bxb1 attB site on the carrier vector crossovers with the attP site on the genome. The Bxb1 integrase links half the attB site to half the attP site and does the same thing to another pair of half attB and attP sites, thereby inserting the circuit into the host genome. (b) Stable cell lines were selected with Puromycin and induced by Dox.

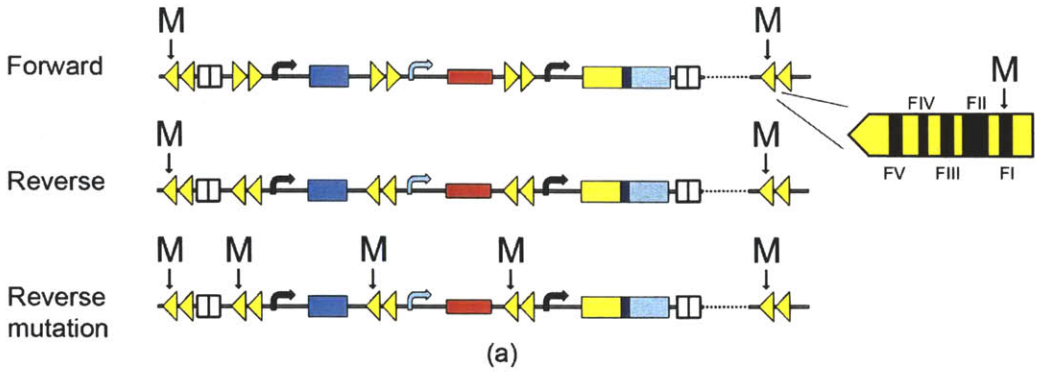
4.4. More Tests on cHS4

Mammalian gene expression is subject to extensive epigenetic regulation. Chromosomal position effects and transcription interference are particularly troublesome for transgene expression. Integrated gene circuits can have highly variable or even silenced expression due to surrounding heterochromatin spreading along with de novo methylation, a phenomenon known as chromosomal position-effect [101]. Transgene expression can also be altered by regulatory elements situated in close proximity, for instance, in adjacent EUs [88].

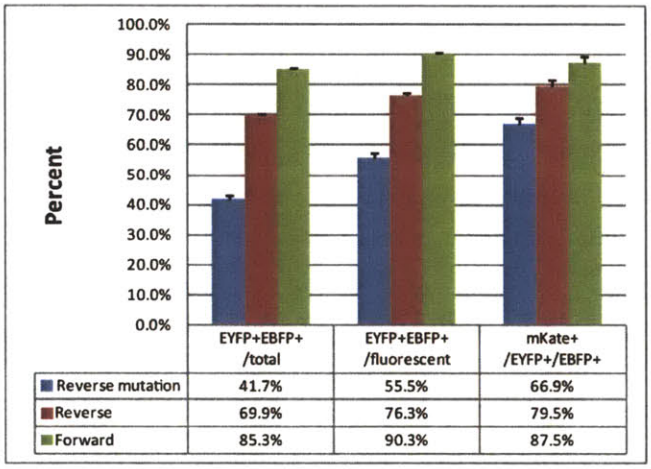
The insulator element has two primary activities: (1) enhancer blocking; (2) position-effect prevention that alleviates repressive histone modification [102], [103]. The most characterized chromatin insulator is cHS4, which stands for the DNase-I hypersensitive site, found within the chicken β-globin locus control region (LCR). Many studies support the finding that much of its insulation activity lies in a 250-bp core fragment and the insulating activity is boosted when tandem copies of cHS4 cores are used [86], [101], [104], [105]. The core region has been dissected into 5 footprinted regions (FI-FV). FII and FVI are CTCF and USF1 binding sites. These two insulator proteins can recruit histone-modifying complexes and can cooperatively mediate the suppression of transcriptional interference [88], [89], [106]. While CTCF cannot bind to a methylated site, VEZF1 has been found associated with FI, FII, and FV and mediated protection from methylation.

This study was interested in investigating the cHS4 effect on a large integrated genetic circuit. Specifically, it is desirable to test the effect of insulator directionality and the effect of a mutated VEZF1 binding site on the performance of a gene circuit with multiple EUs, because previous reports showed inconsistency in cHS4 use [101], [104], [106-109]. A test was designed with three different setups: in between each adjacent EU (Figure 4-5) (1) 2x cHS4 cores placed in its forward direction (as was used in the previous section); (2) 2x cHS4 cores placed in its reverse direction; (3) 2x cHS4 cores placed in reverse direction with a point mutation in the FI site [110] (Genbank Access Number JF330265, AF515846).

Three circuits were assembled using three sets of position vectors. They differed only in the insulator. The assembled circuits were delivered and integrated into HEK293FT cells mediated by Bxb1 recombinase. Stable cell lines were selected and FACS analyzed (Figure 4-5). The results showed that insulators in the forward direction gave the highest DP percentage (EYFP+EBFP+), followed by reverse insulators. Insulators in the reverse

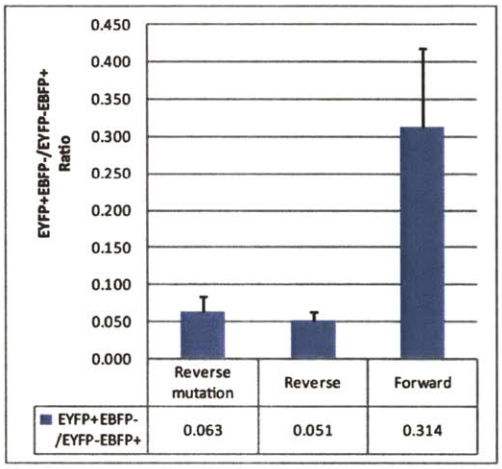


(a)



	EYFP+EBFP+ /total	EYFP+EBFP+ /fluorescent	mKate+ /EYFP+/EBFP+
Reverse mutation	41.7%	55.5%	66.9%
Reverse	69.9%	76.3%	79.5%
Forward	85.3%	90.3%	87.5%

(b)



	Reverse mutation	Reverse	Forward
EYFP+EBFP- /EYFP-EBFP+	0.063	0.051	0.314

(c)

Figure 4-5: Insulator cHS4 test

(a) Three circuits were constructed using three sets of position vectors, which differ in insulator sequences. Forward: 2x cHS4 cores were placed in the same direction of the EUs. Reverse: 2x cHS4 cores were placed in the opposite direction of the EUs. Reverse mutation: 1x cHS4 and another 1x cHS4 core with a mutation in the FI site were placed in the opposite direction of the EUs. A pair of 2x cHS4 cores with a mutation in the FI site flanks inserted circuit on the landing pad. M: a single mutation. (b) Statistics of the fraction of cell that expressed both constitutive fluorescent markers (EYFP+EBFP+). In the second column, non-fluorescent cells were removed (EYFP-EBFP-). Induction refers to mKate expression upon addition of Dox. (c) The ratio of the fraction of EYFP-EBFP+ versus the fraction of EYFP-EBFP+ in the total polyclonal cell population.

direction with a mutation in the FI had only 41.7%. Similarly, the forward cHS4 gave the highest mKate induction percentage. Note that EYFP+EBFP- fraction is much lower than EYFP-EBFP+ fraction in all cases, and the forward cHS4 exhibited more severe imbalance between EYFP and EBFP expression. Interestingly, the EYFP expression unit is the last one on the circuit, with only reverse cHS4 insulating it from the chromosomal sequence in all three cases. The results suggest that the directionality affected insulator performance and forward cHS4 gave best insulation activity. In order to be conclusive about the directionality effect, further tests are need to rule out the possibility that the integration quality (partial integration) are different for three cases and the differences may contribute to the observed effects.

Chapter 5

5. Conclusion and Future Work

The development of DNA construction and delivery methods of large genetic circuits not only provides valuable tools for investigating biological questions, but also serves as an experimental foundation for engineering useful synthetic biological systems. Traditional methods suffer from one or multiple problems. This thesis presents a new experimental technique to enable rapid assembly and delivery of large mammalian gene circuits. This method developed herein enables construction of gene circuits in a modular and efficient manner whereby: (1) EUs are built reliably upon sequence verified libraries; (2) necessary elements such as insulators and polyadenylation signals are incorporated with the EUs into BAMs; (3) multiple BAMs are assembled efficiently, either in a one-step reaction or in a hierarchical process; (4) circuits are embedded into carrier vectors that best suit delivery mechanisms. This method has become a routine technique in the MIT Weiss laboratory and over a hundred circuits were reliably and efficiently assembled over the course of a year by five people. Very large circuits consisting of 11 EUs with sizes up to 61kb were assembled successfully.

Assembled circuits were delivered into cells in various ways. Transient transfection showed that when transferred, an assembled circuit was less noisy than co-transfected circuit components. Bxb1 Integration experiments demonstrated that stable cell lines harboring intact functional circuits could be readily screened. The cell lines exhibited decent and stable transgene expressions. Results were presented on the effects of insulator directionality and the FI site mutant. To date, these were the first tests of insulators in the context of large gene circuit and forward cHS4 was shown to work best.

As a next step to improve assembly efficiency, tests of protocols with a T4 exonuclease followed by a non-error-correction polymerase treatment are planned. Since annealing between non-matching DNA sequences is a major cause for efficiency drop and the error-correcting polymerase can facilitate this non-matching annealing, it is expected that high efficiency could be maintained for larger constructs when T4 exonuclease and non-error-correcting polymerase are used.

The pJazz carrier has superior assembly capability, however, as a linear piece of DNA it cannot be used for Bxb1 mediated integration because the crossover of recombination sites

would split a linear piece into two parts. Current work focuses on developing a circularization method for pJazz in order to use it with Bxb1 mediated recombination delivery. This circularization would not only make pJazz useful for Bxb1 integration, but also remove the prokaryotic backbone sequence that could otherwise stimulate methylation and gene silencing [111].

It also critical to verify the Bxb1 integration specificity and it is very likely that insulator performance is dependent on integration quality. Without this information, it would be hard to be conclusive about the effects of insulator configurations.

Finally, it is also planned to screen for EYFP+EBFP+mKate+ triple positive cell lines from Bxb1 integration experiments. Those cell lines will be used for a long term test, the result of which would provide more information on the long term silencing of integrated circuits.

Appendix A

A. Plasmids Construction

pLV_CAG_CN-2A-CN

The cDNA encoding the two zinc finger nucleases for the AAVS1 locus separated by a 2A tag (Efficient targeting of expressed and silent genes in human ESCs and iPSCs using zinc-finger nucleases, Nature Methods 2009) was synthesized (GeneArt, Germany) and PCR-amplified with attB1/attB2 tags using the primers oPG608b/oPG609b. Upon gel-extraction, the PCR-product was recombined into a pENTR_L1_L2 vector using BP clonase (Life Technologies, Carlsbad CA), yielding the pENTR_L1_CN-2A-CN_L2 vector. In a next step, pENTR_L1_CN-2A-CN_L2 and pENTR_L1_CAG_L2 were recombined into pLV_R4R2_GTW using the LR clonase II plus (Life Technologies, Carlsbad CA), resulting in pLV_CAG_CN-2A-CN.

pGTW4b

The Gateway selection cassette was amplified from pLenti6 (Life technologies, Carlsbad CA) using the oligos oPG452/453, cut with XhoI and MluI and inserted into XhoI/MluI-cut pAAVS1-CAGGS-eGFP (Efficient targeting of expressed and silent genes in human ESCs and iPSCs using zinc-finger nucleases, Nature Methods 2009), resulting in GTW4b.

pGTW4c

The tandem repeated cHS4 core insulators were extracted from pNI-CD (G. Felsenfeld, NIH, MD) using EcoRI and KpnI and ligated into MfeI/KpnI-cut GTW4b.

pZDonor_Seq1-FseI-SeqX

XbaI-Seq1-FseI-SeqX-MfeI was annealed from oligos o-YQ303, o-YQ304 which gave XbaI and MfeI compatible overhangs. XbaI-Seq1-FseI-SeqX-MfeI was inserted into pGTW4c cut with XbaI/MfeI.

pZDonor-L_ISceI-Seq1-FseI-SeqX

MfeI-PciI was annealed from oligos o-YQ353, o-YQ354, which gave MfeI and PciI compatible overhangs. It was then inserted into pZDonor_ISceI-Seq1-FseI-SeqX cut with MfeI/PciI.

pZDonor_ISceI-Seq1-FseI-SeqX

XbaI-ISceI-Seq1-FseI was annealed from oligos o-YQ331, o-YQ332 which gave XbaI and FseI compatible overhangs. XbaI-ISceI-Seq1-FseI was inserted into pZDonor_Seq1-FseI-SeqX cut with XbaI/FseI.

pZDonor I-SceI L

HindIII-ISceI-XhoI was annealed from oligos o-YQ205, o-YQ206, which gave HindIII and XhoI compatible overhangs. It was then inserted into pGTW4c cut with XbaI/MfeI.

pZDonor I-SceI R

AgeI-ISceI-XhoI was annealed from oligos o-YQ207, o-YQ208, which gave AgeI and XhoI compatible overhangs. It was then inserted into pZDonor I-SceI L cut with AgeI/XhoI.

pHL_SA-2A-Puro_1-X_HR

The splice acceptor, puromycin gene and polyadenylation sequences were amplified from pAAVS1-SA-2A-Puro (Efficient targeting of expressed and silent genes in human ESCs and iPSCs using zinc-finger nucleases, Nature Methods 2009) using oPG880/oPG881, digested with NheI and ligated into XbaI-cut pGTW4c, creating pHL_SA-2A-Puro_1-X_HR.

pHL_SA-2A-Bla_1-X_HR

The bleomycin resistance gene was amplified with a splice acceptor and a 2A tag from pLenti6 using oPG950/oPG953 and then oPG952/oPG953, digested with BglII and XbaI and ligated into BglII/XbaI-cut pHL_SA-2A-Puro_1-X_HR.

pJazz-PreBop, pJazz-BeBop

Subsequently we amplified the HL_SA-2A-Puro_1-X_HR or the HL_SA-2A-Bla_1-X_HR cassette from pHL_SA-2A-Puro_1-X_HR respectively pHL_SA-2A-Bla_1-X_HR using oPG450f/oPG451, digested it with NotI and cloned it into the NotI-cut pJazz vector (Lucigen, WI), yielding pJazz-PreBop (Puromycin resistance) and pJazz-BeBop (Basticidin resistance).

pJazz-Swing

A 304 bp cassette was digested with NheI and FseI from pZDonor-L_ISceI-Seq1-FseI-SeqX and ligated into the NheI/FseI cut pJazz-PreBop, yielding pJazz-Swing

pZDonor Position Vectors

cHS4-Gateway selection cassette-rabbit polyadenylation signal was amplified from pGTW4c with oligos listed in the table below. It was a nested PCR. The first step used the first set of oligos and the product was used as the template for the second step PCR with the second set of oligos. The PCR products were digested with AgeI/PacI and ligated into pZDonor I-SceI R cut with AgeI/PacI. The vectors were selection Cm. The vectors were diluted into 10fmol for Gateway reaction.

pZDonor Position Vectors	Oligos
pZDonor 1-GTW-2	o-YQ209, o-YQ211; o-YQ210, o-YQ212
pZDonor 2-GTW-3	o-YQ213, o-YQ215; o-YQ214, o-YQ216
pZDonor 3-GTW-4	o-YQ230, o-YQ240; o-YQ231, o-YQ241
pZDonor 4-GTW-5	o-YQ232, o-YQ242; o-YQ233, o-YQ243
pZDonor 5-GTW-6	o-YQ234, o-YQ244; o-YQ235, o-YQ245
pZDonor 6-GTW-7	o-YQ236, o-YQ246; o-YQ237, o-YQ247
pZDonor 7-GTW-8	o-YQ238, o-YQ248; o-YQ239, o-YQ249

Table_ A-1: Oligos for construction of position vectors

pKan_cHS4-HAR-Kan-SeqX

Kan resistance was amplified from pDONR221 P1P2 vector (Life Technologies, Carlsbad CA) with oligos o-YQ313, o-YQ314. pUC origin of replication was amplified with o-YQ321, o-YQ322. They were digested with XbaI/XhoI and ligated. cHS4-HAR cassette was digested from pGTW4c with XbaI/SacII and inserted into the ligated vector cut with XbaI/SacII.

pKan_cHS4-Seq2-ISceI-HAR-Kan-SeqX

SalI-Seq2-ISceI-EagI was annealed from oligos o-YQ333, o-YQ334, which gave SalI and EagI compatible overhangs. It was then inserted into pKan_cHS4-HAR-Kan-SeqX cut with SalI/EagI.

Adapter Vectors

The adapter vectors were constructed by annealing oligos listed in the table below. The annealed oligos had XbaI and MfeI compatible overhangs. They were inserted into pKan_cHS4-Seq2-ISceI-HAR-Kan-SeqX cut with XbaI/MfeI.

Adapter Vectors	Oligos
pKan_Seq3-cHS4-Seq2-ISceI-HAR-Kan-SeqX	o-YQ335, o-YQ336
pKan_Seq4-cHS4-Seq2-ISceI-HAR-Kan-SeqX	o-YQ337, o-YQ338
pKan_Seq5-cHS4-Seq2-ISceI-HAR-Kan-SeqX	o-YQ339, o-YQ340
pKan_Seq6-cHS4-Seq2-ISceI-HAR-Kan-SeqX	o-YQ341, o-YQ342
pKan_Seq7-cHS4-Seq2-ISceI-HAR-Kan-SeqX	o-YQ349, o-YQ350
pKan_Seq8-cHS4-Seq2-ISceI-HAR-Kan-SeqX	o-YQ351, o-YQ352

Table_ A-2: Oligos for construction of adapter vectors

Oligos

oPG950	GTGACGTGGAGGAGAATCCCGGCCCTAGGCTCGAGATGGCCAAGCCTTTGTCTCAAGAAG
oPG952	GGGAATTCAGATCTGGCAGCGGAGAGGGCAGAGGAAGTCTTCTAACATGCGGTGACGTGGAGGAGAATCC
oPG880	GGGAATTCCTAGCTAGCAAGCTTCTGACCTCTTCTCTTCCTCC
oPG881	GCTCTATGGGTCGACAGTACTAAGCTTGCTAGCTAGGGAATTCCC
oPG450f	GGGAATTCGCGGCCGCGGCCATTACGGCCTGCTTTCTCTGACCAGCATTCTCTCC
oPG451f	GGGTTCCTGGCTCTGCTCTCCACGGCCGCCTCGGCCGCGGCCGCGAATTCCC
oPG452	CACCGCTCGAGACAATTGCATCGATGGTACCGTATCGATGTCGACGTTAACGCTAGTG
oPG453	CTCGTTCAGCTTTCTTGTACAAAGTGGTACGCGTGAATTCCCATG
oPG608b	GGGGACAAGTTTGTACAAAAAAGCAGGCTGACTGCCACCATGGCCCCCAAGAAGAAGCGG
oPG609b	GAAATCAATTTCTGAGGCGCGCCTACCCAGCTTTCTTGTACAAAGTGGTCCCC
o-YQ205	AGCTTATTACCCTGTTATCCCTAACCGGTCAGGACACCAC
o-YQ206	TCGAGTGGTGTCCTGACCGGTTAGGGATAACAGGGTAATA
o-YQ207	CCGGTCAGGACACCATTAATTAATAGGGATAACAGGGTAATC
o-YQ208	TCGAGATTACCCTGTTATCCCTATTAATTAATGGTGTCCTGA
o-YQ209	ATACTTGAACCGGTTTACCGAGCTCTTATTGGTTTTCAAACTTCATTGACTGTGCCAAGC
o-YQ210	GAGCTCTTATTGGTTTTCAAACTTCATTGACTGTGCCAAGCTTCTGACCTCTTCTCTTCC

o-YQ211	TAAGTAGATTAATTAAGGATCTTAAAAACATTATACAATACTACAAGCATAAAAACGCAC
o-YQ212	AACATTATACAATACTACAAGCATAAAAACGCACCAAGCTTAGTACTGTCGATCCGCTGG
o-YQ213	ATACTTGAACCGGTGCGTTTTTATGCTTGTAGTATTGTATAATGTTTTTAAGATCCAAGC
o-YQ214	TTTATGCTTGTAGTATTGTATAATGTTTTTAAGATCCAAGCTTCTGACCTCTTCTCTTCC
o-YQ215	TAAGTAGATTAATTAAGGCGTATAAAACATCTGGATAAGACGAGAGATTGGGTATTAGAC
o-YQ216	ACATCTGGATAAGACGAGAGATTGGGTATTAGACCAAGCTTAGTACTGTCGATCCGCTGG
o-YQ230	ATACTTGAACCGGTCTAATACCCAATCTCTCGTCTTATCCAGATGTTTTATACGCCAAGC
o-YQ231	ACCCAATCTCTCGTCTTATCCAGATGTTTTATACGCCAAGCTTCTGACCTCTTCTCTTCC
o-YQ232	ATACTTGAACCGGTGAATTCCCTTATGTGAGTGTAAAAGGCAGGCGAGTTTGTCCCAAGC
o-YQ233	CCCTTATGTGAGTGTAAAAGGCAGGCGAGTTTGTCCCAAGCTTCTGACCTCTTCTCTTCC
o-YQ234	ATACTTGAACCGGTTGCTTGCAAAAGCAGTAATTGGAAAGCACTCTCAAAGAATCCAAGC
o-YQ235	GCAAAAGCAGTAATTGGAAAGCACTCTCAAAGAATCCAAGCTTCTGACCTCTTCTCTTCC
o-YQ236	ATACTTGAACCGGTAGATAAGTTGATTTAGCCATAAAATATTGTTTCCGTGACCCCAAGC
o-YQ237	AGTTGATTTAGCCATAAAATATTGTTTCCGTGACCCCAAGCTTCTGACCTCTTCTCTTCC
o-YQ238	ATACTTGAACCGGTTCTGAGTCACGGCTTCATTGGCATTCCGTACAACGAACGTCCAAGC
o-YQ239	GTCACGGCTTCATTGGCATTCCGTACAACGAACGTCCAAGCTTCTGACCTCTTCTCTTCC
o-YQ240	TAAGTAGATTAATTAAGGGACAAACTCGCCTGCCTTTTACACTCACATAAGGGAATTCAC
o-YQ241	CGCCTGCCTTTTACACTCACATAAGGGAATTCACCAAGCTTAGTACTGTCGATCCGCTGG
o-YQ242	TAAGTAGATTAATTAAGGATTCTTTGAGAGTGCTTTCCAATTACTGCTTTTGCA

	AGCAAC
o-YQ243	AGAGTGCTTTCCAATTACTGCTTTTGCAAGCAACCAAGCTTAGTACTGTCGATC CGCTGG
o-YQ245	AAACAATATTTTATGGCTAAATCAACTTATCTACCAAGCTTAGTACTGTCGATC CGCTGG
o-YQ246	TAAGTAGATTAATTAAGGACGTTCGTTGTACGGAATGCCAATGAAGCCGTGAC TCAGAAC
o-YQ247	TGTACGGAATGCCAATGAAGCCGTGACTCAGAACCAAGCTTAGTACTGTCGAT CCGCTGG
o-YQ248	TAAGTAGATTAATTAAGGGGATGCATGGTGTTGTTTTACCGCTATAGGCTCTCT GAGGAC
o-YQ249	GGTGTTGTTTTACCGCTATAGGCTCTCTGAGGACCAAGCTTAGTACTGTCGATC CGCTGG
o-YQ313	ATACTTGATCTAGAAATATTCGCGAGACCGCGGGCAGCTCTGGCCCGTGTCTC
o-YQ314	TAAGTAGACTCGAGGGTTAGGCGACTGTTATAACTTACCTCTGTAATACTAGTG ATACCGTCTGACGCTCAGTGGAACGACG
o-YQ321	TAAGTAGATCTAGAGGTACGCGTATTAATTGCGTTGCGCT
o-YQ322	ATACTTGACTCGAGAGTTACGCGTCGTTCCACTGAGC
o-YQ331	CTAGAATTACCCTGTTATCCCTAGGTTTACCGAGCTCTTATTGGTTTTCAAACTT CATTGACTGTGCCGGCCGG
o-YQ332	CCGGCACAGTCAATGAAGTTTGAAAACCAATAAGAGCTCGGTAAACCTAGGGA TAACAGGGTAATT
o-YQ333	TCGACGGTGCGTTTTTATGCTTGTAGTATTGTATAATGTTTTTAAGATCCTAGGG ATAACAGGGTAATC
o-YQ334	GGCCGATTACCCTGTTATCCCTAGGATCTTAAAAACATTATACAATACTACAAG CATAAAAACGCACCG
o-YQ335	CTAGAGGTCTAATACCCAATCTCTCGTCTTATCCAGATGTTTTATACGCCC
o-YQ336	AATTGGGCGTATAAAACATCTGGATAAGACGAGAGATTGGGTATTAGACCT
o-YQ337	CTAGAGGTGAATTCCCTTATGTGAGTGTAAAAGGCAGGCGAGTTTGTCCCC
o-YQ338	AATTGGGGACAAACTCGCCTGCCTTTTACACTCACATAAGGGAATTCACCT
o-YQ339	CTAGAGGTTGCTTGCAAAAGCAGTAATTGGAAAGCACTCTCAAAGAATCCC
o-YQ340	AATTGGGATTCTTTGAGAGTGCTTTCCAATTACTGCTTTTGCAAGCAACCT
o-YQ341	CTAGAGGTAGATAAGTTGATTTAGCCATAAAATATTGTTTCCGTGACCCCC
o-YQ342	AATTGGGGGTCACGGAAACAATATTTTATGGCTAAATCAACTTATCTACCT
o-YQ349	CTAGAGGTTCTGAGTCACGGCTTCATTGGCATTCCGTACAACGAACGTCCC
o-YQ350	AATTGGGACGTTCGTTGTACGGAATGCCAATGAAGCCGTGACTCAGAACCT
o-YQ351	CTAGAGGTCCTCAGAGAGCCTATAGCGGTAAAACAACACCATGCATCCCCC

o-YQ352	AATTGGGGGATGCATGGTGTTGTTTTACCGCTATAGGCTCTCTGAGGACCT
o-YQ353	AATTGAGGGGATAACGCAGGAAAGAA
o-YQ354	CATGTTCTTTCCTGCGTTATCCCCTC

Table_ A-3: Oligos used in this work

Bibliography

[1] N. Nandagopal and M. B. Elowitz, "Synthetic Biology: Integrated Gene Circuits," *Science*, vol. 333, no. 6047, pp. 1244-1248, Sep. 2011.

[2] E. Davidson and M. Levin, "Gene regulatory networks.," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 102, no. 14, p. 4935, Apr. 2005.

[3] D. R. Scannell and K. Wolfe, "Rewiring the transcriptional regulatory circuits of cells.," *Genome biology*, vol. 5, no. 2, p. 206, Jan. 2004.

[4] A. Rolletschek and A. M. Wobus, "Induced human pluripotent stem cells: promises and open questions.," *Biological chemistry*, vol. 390, no. 9, pp. 845-9, Sep. 2009.

[5] K. Takahashi and S. Yamanaka, "Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors.," *Cell*, vol. 126, no. 4, pp. 663-76, Aug. 2006.

[6] T. Vierbuchen, A. Ostermeier, Z. P. Pang, Y. Kokubu, T. C. Südhof, and M. Wernig, "Direct conversion of fibroblasts to functional neurons by defined factors.," *Nature*, vol. 463, no. 7284, pp. 1035-41, Feb. 2010.

[7] Z. P. Pang et al., "Induction of human neuronal cells by defined transcription factors.," *Nature*, pp. 3-7, May. 2011.

[8] M. Caiazzo et al., "Direct generation of functional dopaminergic neurons from mouse and human fibroblasts.," *Nature*, pp. 10-15, Jul. 2011.

[9] B. Tursun, T. Patel, P. Kratsios, and O. Hobert, "Direct conversion of C. elegans germ cells into specific neuron types.," *Science (New York, N.Y.)*, vol. 331, no. 6015, pp. 304-8, Jan. 2011.

[10] N. Novershtern et al., "Densely interconnected transcriptional circuits control cell states in human hematopoiesis.," *Cell*, vol. 144, no. 2, pp. 296-309, Jan. 2011.

[11] K. a Haynes and P. a Silver, "Eukaryotic systems broaden the scope of synthetic biology.," *The Journal of cell biology*, vol. 187, no. 5, pp. 589-96, Nov. 2009.

[12] M. Tigges and M. Fussenegger, "Recent advances in mammalian synthetic biology-design of synthetic transgene control networks.," *Current opinion in biotechnology*, vol. 20, no. 4, pp. 449-60, Aug. 2009.

[13] P. E. M. Purnick and R. Weiss, "The second wave of synthetic biology: from modules to systems.," *Nature reviews. Molecular cell biology*, vol. 10, no. 6, pp. 410-22, Jun. 2009.

[14] E. Wimmer, S. Mueller, T. M. Tumpey, and J. K. Taubenberger, "Synthetic viruses: a new opportunity to understand and prevent viral disease.," *Nature biotechnology*, vol. 27, no. 12, pp. 1163-72, Dec. 2009.

[15] A. S. Khalil and J. J. Collins, "Synthetic biology: applications come of age.," *Nature reviews. Genetics*, vol. 11, no. 5, pp. 367-79, May. 2010.

[16] W. Weber and M. Fussenegger, "Emerging biomedical applications of synthetic biology," *Nature Reviews Genetics*, vol. 13, no. 1, pp. 21-35, Nov. 2011.

[17] T. S. Bayer, "Using synthetic biology to understand the evolution of gene expression.," *Current biology : CB*, vol. 20, no. 17, pp. R772-9, Sep. 2010.

[18] J. B. Xavier, "Social interaction in synthetic and natural microbial communities.," *Molecular systems biology*, vol. 7, p. 483, Apr. 2011.

[19] T. K. Lu, A. S. Khalil, and J. J. Collins, "Next-generation synthetic gene networks.," *Nature biotechnology*, vol. 27, no. 12, pp. 1139-50, Dec. 2009.

[20] D. Sprinzak and M. B. Elowitz, "Reconstruction of genetic circuits.," *Nature*, vol. 438, no. 7067, pp. 443-8, Nov. 2005.

[21] S. Mukherji and A. van Oudenaarden, "Synthetic biology: understanding biological design from synthetic circuits.," *Nature reviews. Genetics*, vol. 10, no. 12, pp. 859-71, Dec. 2009.

[22] N. Nandagopal and M. B. Elowitz, "Synthetic Biology: Integrated Gene Circuits," *Science*, vol. 333, no. 6047, pp. 1244-1248, Sep. 2011.

[23] D. A. Drubin, J. C. Way, and P. A. Silver, "Designing biological systems.," *Genes & development*, vol. 21, no. 3, pp. 242-54, Feb. 2007.

[24] Z. Xie, L. Wroblewska, L. Prochazka, R. Weiss, and Y. Benenson, "Multi-Input RNAi-Based Logic Circuit for Identification of Specific Cancer Cells," *Science*, vol. 333, no. 6047, pp. 1307-1311, Sep. 2011.

[25] V. M. Rivera et al., "A humanized system for pharmacologic control of gene expression," *Nature Medicine*, vol. 2, no. 9, pp. 1028-1032, Sep. 1996.

[26] T. Clackson, "Regulated gene expression systems.," *Gene therapy*, vol. 7, no. 2, pp. 120-5, Jan. 2000.

[27] A. Auricchio et al., "Constitutive and regulated expression of processed insulin following in vivo hepatic gene transfer.," *Gene therapy*, vol. 9, no. 14, pp. 963-71, Jul. 2002.

[28] W. Weber et al., "A synthetic mammalian gene circuit reveals antituberculosis compounds.," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 105, no. 29, pp. 9994-8, Jul. 2008.

[29] M. Gossen and H. Bujard, "Studying gene function in eukaryotes by conditional gene inactivation.," *Annual review of genetics*, vol. 36, pp. 153-73, Jan. 2002.

[30] P. de Felipe, "Polycistronic viral vectors.," *Current gene therapy*, vol. 2, no. 3, pp. 355-78, Oct. 2002.

[31] G. Luke, H. Escuin, P. De Felipe, and M. Ryan, "2A to the fore - research, technology and applications.," *Biotechnology & genetic engineering reviews*, vol. 26, pp. 223-60, Jan. 2010.

[32] H. Mizuguchi, Z. Xu, A. Ishii-Watabe, E. Uchida, and T. Hayakawa, "IRES-dependent second gene expression is significantly lower than cap-dependent first gene expression in a bicistronic vector.," *Molecular therapy : the journal of the American Society of Gene Therapy*, vol. 1, no. 4, pp. 376-82, Apr. 2000.

[33] A. Paquin, D. E. Jaalouk, and J. Galipeau, "Retrovector encoding a green fluorescent protein-herpes simplex virus thymidine kinase fusion protein serves as a versatile suicide/reporter for cell and gene therapy applications.," *Human gene therapy*, vol. 12, no. 1, pp. 13-23, Jan. 2001.

[34] S. C. Makrides, "Components of vectors for gene transfer and expression in mammalian cells.," *Protein expression and purification*, vol. 17, no. 2, pp. 183-202, Nov. 1999.

[35] J. Russell, David W.; Sambrook, *Molecular cloning: a laboratory manual.* Cold Spring Harbor Laboratory, 2001.

[36] T. Knight, "Idempotent Vector Design for Standard Assembly of Biobricks."

[37] E. Weber, C. Engler, R. Gruetzner, S. Werner, and S. Marillonnet, "A modular cloning system for standardized assembly of multigene constructs.," *PloS one*, vol. 6, no. 2, p. e16765, Jan. 2011.

[38] C. Engler, R. Gruetzner, R. Kandzia, and S. Marillonnet, "Golden gate shuffling: a one-pot DNA shuffling method based on type IIs restriction enzymes.," *PloS one*, vol. 4, no. 5, p. e5553, Jan. 2009.

[39] F. Zhang, L. Cong, S. Lodato, S. Kosuri, G. M. Church, and P. Arlotta, "Efficient construction of sequence-specific TAL effectors for modulating mammalian transcription," *Online*, no. January, pp. 1-6, 2011.

[40] J. L. Hartley, G. F. Temple, and M. A. Brasch, "DNA cloning using in vitro site-specific recombination.," *Genome research*, vol. 10, no. 11, pp. 1788-95, Nov. 2000.

[41] E. Magnani, L. Bartling, and S. Hake, "From Gateway to MultiSite Gateway in one recombination event.," *BMC molecular biology*, vol. 7, no. 1, p. 46, Jan. 2006.

[42] Life Techonologies, *GATEWAY™ Cloning Technology Instruction Manual.* .

[43] Y. Sasaki et al., "Evidence for high specificity and efficiency of multiple recombination signals in mixed DNA cloning by the Multisite Gateway system.," *Journal of biotechnology*, vol. 107, no. 3, pp. 233-43, Feb. 2004.

[44] Q.-J. Chen, H.-M. Zhou, J. Chen, and X.-C. Wang, "A Gateway-based platform for multigene plant transformation.," *Plant molecular biology*, vol. 62, no. 6, pp. 927-36, Dec. 2006.

[45] A. Kriz et al., "A plasmid-based multigene expression system for mammalian cells.," *Nature communications*, vol. 1, no. 8, p. 120, Jan. 2010.

[46] Clontech, "In-Fusion® PCR Cloning System." [Online]. Available: http://bioinfo.clontech.com/infusion/.

[47] B. Zhu, G. Cai, E. Hall, and G. Freeman, "In-Fusion™ assembly: seamless engineering of multidomain fusion proteins, modular vectors, and mutations," *BioTechniques*, vol. 43, no. 3, pp. 354-359, Sep. 2007.

[48] D. G. Gibson et al., "Complete chemical synthesis, assembly, and cloning of a Mycoplasma genitalium genome.," *Science (New York, N.Y.)*, vol. 319, no. 5867, pp. 1215-20, Feb. 2008.

[49] D. G. Gibson et al., "Enzymatic assembly of DNA molecules up to several hundred kilobases," *Nature Methods*, vol. 6, no. 5, pp. 12-16, 2009.

[50] S. Bacchetti, "Transfer of the Gene for Thymidine Kinase to Thymidine Kinase-Deficient Human Cells by Purified Herpes Simplex Viral DNA," *Proceedings of the National Academy of Sciences*, vol. 74, no. 4, pp. 1590-1594, Apr. 1977.

[51] P. L. Felgner et al., "Lipofection: a highly efficient, lipid-mediated DNA-transfection procedure.," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 84, no. 21, pp. 7413-7, Nov. 1987.

[52] S. Nimesh, N. Gupta, and R. Chandra, "Cationic polymer based nanocarriers for delivery of therapeutic nucleic acids.," *Journal of biomedical nanotechnology*, vol. 7, no. 4, pp. 504-20, Aug. 2011.

[53] Lonza, "The Nucleofector™ Technology – a Revolution in Transfection."

[54] A. D. Adamson, D. Jackson, and J. R. E. Davis, "Novel approaches to in vitro transgenesis.," *The Journal of endocrinology*, vol. 208, no. 3, pp. 193-206, Mar. 2011.

[55] C. E. Thomas, A. Ehrhardt, and M. A. Kay, "Progress and problems with the use of viral vectors for gene therapy.," *Nature reviews. Genetics*, vol. 4, no. 5, pp. 346-58, May. 2003.

[56] D. a Jackson, S. Juranek, and H. J. Lipps, "Designing nonviral vectors for efficient gene transfer and long-term gene expression.," *Molecular therapy : the journal of the American Society of Gene Therapy*, vol. 14, no. 5, pp. 613-26, Nov. 2006.

[57] A. C. W. Jenke et al., "Nuclear scaffold/matrix attached region modules linked to a transcription unit are sufficient for replication and maintenance of a mammalian episome.," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 101, no. 31, pp. 11322-7, Aug. 2004.

[58] C. Piechaczek, C. Fetzer, a Baiker, J. Bode, and H. J. Lipps, "A vector based on the SV40 origin of replication and chromosomal S/MARs replicates episomally in CHO cells.," *Nucleic acids research*, vol. 27, no. 2, pp. 426-8, Jan. 1999.

[59] M. Conese, C. Auriche, and F. Ascenzioni, "Gene therapy progress and prospects: episomally maintained self-replicating systems.," *Gene therapy*, vol. 11, no. 24, pp. 1735-41, Dec. 2004.

[60] D. Schaarschmidt, J. Baltin, I. M. Stehle, H. J. Lipps, and R. Knippers, "An episomal mammalian replicon: sequence-independent binding of the origin recognition complex.," *The EMBO journal*, vol. 23, no. 1, pp. 191-201, Jan. 2004.

[61] K. Van Craenenbroeck, P. Vanhoenacker, and G. Haegeman, "Episomal vectors for gene expression in mammalian cells.," *European journal of biochemistry / FEBS*, vol. 267, no. 18, pp. 5665-78, Sep. 2000.

[62] and J. D. Harvey Lodish, Arnold Berk, S Lawrence Zipursky, Paul Matsudaira, David Baltimore, "Molecular Cell Biology." W. H. Freeman, 2000.

[63] D. Carroll, "Genome Engineering With Zinc-Finger Nucleases," *Genetics*, vol. 188, no. 4, pp. 773-782, Aug. 2011.

[64] F. Paques and P. Duchateau, "Meganucleases and DNA Double-Strand Break-Induced Recombination: Perspectives for Gene Therapy," *Current Gene Therapy*, vol. 7, no. 1, p. 18, 2007.

[65] F. Le Provost, S. Lillico, B. Passet, R. Young, B. Whitelaw, and J.-L. Vilotte, "Zinc finger nuclease technology heralds a new era in mammalian transgenesis.," *Trends in biotechnology*, vol. 28, no. 3, pp. 134-41, Mar. 2010.

[66] Y.-G. Kim, "Construction of a Z-DNA-specific restriction endonuclease," *Proceedings of the National Academy of Sciences*, vol. 94, no. 24. pp. 12875-12879, 25-Nov-1997.

[67] D. Hockemeyer et al., "Efficient targeting of expressed and silent genes in human ESCs and iPSCs using zinc-finger nucleases.," *Nature biotechnology*, vol. 27, no. 9, pp. 851-7, Sep. 2009.

[68] Sigma-Aldrich, "CompoZr® ZFN technology from Sigma-Aldrich." [Online]. Available: http://www.compozrzfn.com/. [Accessed: 22-Sep-2011].

[69] D. Hockemeyer et al., "Genetic engineering of human pluripotent cells using TALE nucleases.," *Nature biotechnology*, vol. 29, no. 8, pp. 731-734, Jul. 2011.

[70] T. Li et al., "TAL nucleases (TALNs): hybrid proteins composed of TAL effectors and FokI DNA-cleavage domain.," *Nucleic acids research*, vol. 39, no. 1, pp. 359-72, Jan. 2011.

[71] M. Christian et al., "TAL effector nucleases create targeted DNA double-strand breaks," *Genome*, 2010.

[72] J. Boch et al., "Breaking the code of DNA binding specificity of TAL-type III effectors.," *Science (New York, N.Y.)*, vol. 326, no. 5959, pp. 1509-12, Dec. 2009.

[73] M. J. M. and A. J. Bogdanove, "A Simple Cipher Governs DNA Recognition by TAL Effectors," *Science*, no. December, pp. 50011-50011, 2009.

[74] T. Cermak et al., "Efficient design and assembly of custom TALEN and other TAL effector-based constructs for DNA targeting.," *Nucleic acids research*, vol. 39, no. 12, p. e82, Jul. 2011.

[75] P. C. Hendrie and D. W. Russell, "Gene targeting with viral vectors.," *Molecular therapy : the journal of the American Society of Gene Therapy*, vol. 12, no. 1, pp. 9-17, Jul. 2005.

[76] A. S. Cockrell and T. Kafri, "Gene delivery by lentivirus vectors," *Molecular Biotechnology*, vol. 36, no. 3, pp. 184-204, Apr. 2007.

[77] M. A. Kay, J. C. Glorioso, and L. Naldini, "Viral vectors for gene therapy: the art of turning infectious agents into vehicles of therapeutics.," *Nature medicine*, vol. 7, no. 1, pp. 33-40, Jan. 2001.

[78] S. Turan et al., "Recombinase-mediated cassette exchange (RMCE): traditional concepts and current challenges.," *Journal of molecular biology*, vol. 407, no. 2, pp. 193-221, Mar. 2011.

[79] T. W. Chalberg et al., "Integration specificity of phage phiC31 integrase in the human genome.," *Journal of molecular biology*, vol. 357, no. 1, pp. 28-48, Mar. 2006.

[80] B. Thyagarajan, M. J. Guimarães, A. C. Groth, and M. P. Calos, "Mammalian genomes contain active recombinase recognition sites.," *Gene*, vol. 244, no. 1-2, pp. 47-54, Feb. 2000.

[81] M.-C. Birling, F. Gofflot, and X. Warot, "Site-specific recombinases for manipulation of the mouse genome.," *Methods in molecular biology (Clifton, N.J.)*, vol. 561, pp. 245-63, Jan. 2009.

[82] A. Loonstra et al., "Growth inhibition and DNA damage induced by Cre recombinase in mammalian cells.," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 98, no. 16, pp. 9209-14, Jul. 2001.

[83] J. P. Russell, D. W. Chang, A. Tretiakova, and M. Padidam, "Phage Bxb1 integrase mediates highly efficient site-specific recombination in mammalian cells," *BioTechniques*, vol. 40, no. 4, pp. 460-464, 2006.

[84] J. Huang, P. Ghosh, G. F. Hatfull, and Y. Hong, "Successive and Targeted DNA Integrations in Drosophila Genome by Bxb1 and {Phi}C31 Integrases.," *Genetics*, vol. 189, no. 1, pp. 391-5, Jun. 2011.

[85] M. Gaszner and G. Felsenfeld, "Insulators: exploiting transcriptional and epigenetic mechanisms.," *Nature reviews. Genetics*, vol. 7, no. 9, pp. 703-13, Sep. 2006.

[86] J. H. Chung, M. Whiteley, and G. Felsenfeld, "A 5' element of the chicken beta-globin domain serves as an insulator in human erythroid cells and protects against position effect in Drosophila.," *Cell*, vol. 74, no. 3, pp. 505-14, Aug. 1993.

[87] J. Dickson et al., "VEZF1 elements mediate protection from DNA methylation.," *PLoS genetics*, vol. 6, no. 1, p. e1000804, Jan. 2010.

[88] K. Yahata et al., "cHS4 insulator-mediated alleviation of promoter interference during cell-based expression of tandemly associated transgenes.," *Journal of molecular biology*, vol. 374, no. 3, pp. 580-90, Nov. 2007.

[89] F. Recillas-Targa et al., "Position-effect protection and enhancer blocking by the chicken beta-globin insulator are separable activities.," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 99, no. 10, pp. 6883-8, May. 2002.

[90] T. H. Kim et al., "Analysis of the vertebrate insulator protein CTCF-binding sites in the human genome.," *Cell*, vol. 128, no. 6, pp. 1231-45, Mar. 2007.

[91] J. Lanoix and N. H. Acheson, "A rabbit beta-globin polyadenylation signal directs efficient termination of transcription of polyomavirus DNA.," *The EMBO journal*, vol. 7, no. 8, pp. 2515-22, Aug. 1988.

[92] M. R. Shortreed et al., "A thermodynamic approach to designing structure-free combinatorial DNA word sets.," *Nucleic acids research*, vol. 33, no. 15, pp. 4965-77, Jan. 2005.

[93] M. Andronescu, Z. C. Zhang, and A. Condon, "Secondary structure prediction of interacting RNA molecules.," *Journal of molecular biology*, vol. 345, no. 5, pp. 987-1001, Feb. 2005.

[94] A. Condon and H. H. Hoos, "RNAsoft : a suite of RNA secondary structure prediction and design software tools," *Computer*, vol. 31, no. 13, pp. 3416-3422, 2003.

[95] T. Feng et al., "Increased efficiency of cloning large DNA fragments using a lower copy number plasmid.," *BioTechniques*, vol. 32, no. 5, pp. 992, 994, 996 passim, May. 2002.

[96] T. P. Cunningham, R. C. Montelaro, and K. E. Rushlow, "Lentivirus envelope sequences and proviral genomes are stabilized in Escherichia coli when cloned in low-copy-number plasmid vectors.," *Gene*, vol. 124, no. 1, pp. 93-8, Feb. 1993.

[97] R. Godiska et al., "Linear plasmid vector for cloning of repetitive or unstable sequences in Escherichia coli.," *Nucleic acids research*, vol. 38, no. 6, p. e88, Apr. 2010.

[98] F. Malagón and A. Aguilera, "Genetic stability and DNA rearrangements associated with a 2 x 1.1-Kb perfect palindrome in Escherichia coli.," *Molecular & general genetics : MGG*, vol. 259, no. 6, pp. 639-44, Oct. 1998.

[99] Z. Qu et al., "Homogeneity and long-term stability of tetracycline-regulated gene expression with low basal activity by using the rtTA2S-M2 transactivator and insulator-flanked reporter vectors.," *Gene*, vol. 327, no. 1, pp. 61-73, Feb. 2004.

[100] A. Golberg and B. Rubinsky, "The Effect of Electroporation Type Pulsed Electric Fields on DNA in Aqueous Solution," *Technology in Cancer Research and Treatment*, vol. 9, no. 4, pp. 423-430, 2010.

[101] D. W. Emery, E. Yannaki, J. Tubb, and G. Stamatoyannopoulos, "A chromatin insulator protects retrovirus vectors from chromosomal position effects.," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 97, no. 16, pp. 9150-5, Aug. 2000.

[102] A. M. Bushey, E. R. Dorman, and V. G. Corces, "Chromatin insulators: regulatory mechanisms and epigenetic inheritance.," *Molecular cell*, vol. 32, no. 1, pp. 1-9, Oct. 2008.

[103] M. Amouyal, "Gene insulation. Part II: natural strategies in vertebrates.," *Biochemistry and cell biology = Biochimie et biologie cellulaire*, vol. 88, no. 6, pp. 885-98, Dec. 2010.

[104] J. H. Chung, a C. Bell, and G. Felsenfeld, "Characterization of the chicken beta-globin insulator.," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 94, no. 2. pp. 575-80, 21-Jan-1997.

[105] H. Hanawa, M. Yamamoto, H. Zhao, T. Shimada, and D. a Persons, "Optimized lentiviral vector design improves titer and transgene expression of vectors containing the chicken beta-globin locus HS4 insulator element.," *Molecular therapy : the journal of the American Society of Gene Therapy*, vol. 17, no. 4, pp. 667-74, Apr. 2009.

[106] a C. Bell, a G. West, and G. Felsenfeld, "The protein CTCF is required for the enhancer blocking activity of vertebrate insulators.," *Cell*, vol. 98, no. 3, pp. 387-96, Aug. 1999.

[107] V. J. Mutskov, C. M. Farrell, P. a Wade, A. P. Wolffe, and G. Felsenfeld, "The barrier function of an insulator couples high histone acetylation levels with specific protection of promoter DNA from methylation.," *Genes & development*, vol. 16, no. 12, pp. 1540-54, Jun. 2002.

[108] E. Yannaki, J. Tubb, M. Aker, G. Stamatoyannopoulos, and D. W. Emery, "Topological constraints governing the use of the chicken HS4 chromatin insulator in oncoretrovirus vectors.," *Molecular therapy : the journal of the American Society of Gene Therapy*, vol. 5, no. 5 Pt 1, pp. 589-98, May. 2002.

[109] P. Ciana, "Engineering of a Mouse for the in Vivo Profiling of Estrogen Receptor Activity," *Molecular Endocrinology*, vol. 15, no. 7, pp. 1104-1113, Jul. 2001.

[110] S. A. Rankin, A. M. Zorn, and D. R. Buchholz, "New doxycycline-inducible transgenic lines in Xenopus.," *Developmental dynamics : an official publication of the American Association of Anatomists*, vol. 240, no. 6, pp. 1467-74, Jun. 2011.

[111] M. a Kay, C.-Y. He, and Z.-Y. Chen, "A robust system for production of minicircle DNA vectors.," *Nature biotechnology*, vol. 28, no. 12, pp. 1287-9, Dec. 2010.