Generating Inducible Vector Systems for Controlling Pluripotent Stem Cell Fate

by

Cesar Yamarte

A thesis submitted in conformity with the requirements for the degree of Master of Applied Science
Institute of Biomaterials and Biomedical Engineering
University of Toronto

© Copyright by Cesar Yamarte, 2012
Abstract

Transgenic manipulation of exogenous and endogenous gene expression in human embryonic stem cells (hESCs) is a powerful approach to decipher the genetic pathways dictating their developmental fate. Presently used genetic tools face limitations including leakiness in inducibility of expression, epigenetic silencing in long-term cell culture, low genomic integration efficiencies, small genetic cargo limit and lack of high-throughput cloning capabilities. To overcome these limitations, I have constructed R4-Integrase and piggyBac transposon genetic vector systems for stable transgene overexpression and knockdown in hESCs. Preliminary functional testing of the piggyBac vector system in HEK 293T and hESCs demonstrated vector inducibility as well as successful overexpression and knockdown of pluripotency factor OCT4. Concurrently, a cost-effective and high efficiency method for chemical transfection of hESCs was developed. Exogenous overexpression and knockdown of transcription factors in hESCs will aid in the elucidation of gene regulatory networks controlling pluripotency and developmental fate.
Acknowledgements

First I would like to thank my family. Although I have been far away from you all during this time, your determination has constantly reassured me that support and inspiration will be near no matter where I go. I would also like to thank Dr. Bill Stanford for the opportunity to learn about and to research one of the most interesting areas in science today. A special thanks to Dr. James Ellis, Dr. Cindi Morshead and Dr. Julie Audet for lending me a part of their schedules to evaluate my work. To lab members present and past, thank you for all the priceless advice that helped me chart the way to where I am today. I am sure I will see you all in the future. To my friends, thanks for the wild ride and for helping me to learn as much about life outside the lab as I did about science inside the lab.
Table of Contents

Acknowledgements ...................................................................................................................... iii
Table of Contents ........................................................................................................................ iv
List of Tables ................................................................................................................................... vii
List of Figures ................................................................................................................................. viii
List of Abbreviations ..................................................................................................................... x

Chapter 1 Introduction .................................................................................................................... 1
  1.1 Embryonic Stem Cells ........................................................................................................ 1
  1.2 Understanding Mammalian Development through ESCs ................................................. 2
  1.3 Regenerative Medicine ..................................................................................................... 3
    1.3.1 Immunogenicity ......................................................................................................... 4
    1.3.2 Cell availability ......................................................................................................... 4
    1.3.3 Tumorigenicity ......................................................................................................... 5
  1.4 Genetic manipulation of Human Embryonic Stem Cells .................................................... 6
    1.4.1 R4 Integrase Recombination ..................................................................................... 8
    1.4.2 The piggyBac Transposon ....................................................................................... 11
  1.5 Genetic Insulators & Silencing ......................................................................................... 12
  1.6 RNA Interference ............................................................................................................ 13
  1.7 Genetic Elements ............................................................................................................ 15
  1.8 Project Objectives and Hypothesis ................................................................................ 19

Chapter 2 Generation of modular, inducible, protein-tagged, selectable and epigenetically insulated overexpression/shRNA-knockdown vector systems based on R4 Integrase and piggyBac platforms ..................................................................................................................... 20
  2.1 Overview .................................................................................................................... 20
  2.2 Materials and Methods .................................................................................................. 21
    2.2.1 PCR Fusion of Genetic Elements ............................................................... 21
    2.2.2 Gateway® Cloning .............................................................................................. 23
    2.2.3 Sticky-End Ligation of shRNAs ................................................................. 25
    2.2.4 PCR ............................................................................................................... 26
    2.2.5 DNA Sequencing ............................................................................................ 26
    2.2.6 Restriction Digest Analysis of Constructs .............................................. 26
2.3 Results .......................................................................................................................... 27
  2.3.1 Assembly of Overexpression R4 Vector System .......................................................... 27
  2.3.2 Assembly of shRNA Knockdown R4 Vector System .................................................. 28
  2.3.3 Genetic Insulation and Gateway compatibility of piggyBac destination vector ........... 29
  2.3.4 Assembly of Overexpression piggyBac Vector System ............................................... 30
  2.3.5 Assembly of shRNA Knockdown piggyBac Vector System ........................................ 32

Chapter 3 Functional testing of vector systems in HEK 293T and hESCs ............................. 34
  3.1 Overview ..................................................................................................................... 34
  3.2 Materials and Methods .............................................................................................. 34
    3.2.1 The H9 Cell Line .................................................................................................... 34
    3.2.2 The H9 R4 Cell Line ............................................................................................ 35
    3.2.3 HEK 293T Cell Culture ........................................................................................ 35
    3.2.4 The HEK 293 R4 Cell Line .................................................................................... 35
    3.2.5 hESC Cell Culture ............................................................................................... 35
      3.2.5.1 Mouse embryonic fibroblast (MEF)-supported hESC culture .................. 36
      3.2.5.2 Feeder-free hESC culture with conditioned medium (CM) ....................... 36
      3.2.5.3 Feeder-free hESC culture with chemically defined E8 medium ............... 36
    3.2.6 Cell Transfection ................................................................................................... 36
      3.2.6.1 hESC Electroporation .................................................................................. 36
      3.2.6.2 Chemical Stemfect™ Transfection of H9 hESCs ........................................ 37
      3.2.6.3 Chemical Transfection of HEK 293T & HEK293 R4 Cells ....................... 38
    3.2.7 Cell Harvest for fluorescence analysis .................................................................... 38
      3.2.7.1 Cell Surface (SSEA-3 and TRA-1-60) Immunostaining ............................ 38
      3.2.7.2 OCT4 Immunostaining for flow Cytometry .............................................. 39
    3.2.8 OCT4 Immunostaining for in situ immunofluorescence imaging ......................... 39
    3.2.9 Flow Cytometry Analysis ..................................................................................... 39
    3.2.10 Immunoprecipitation (IP) .................................................................................. 39
    3.2.11 Immunoblotting ................................................................................................. 40
  3.3 Results ........................................................................................................................ 40
    3.3.1 R4 – Integrase Vector Systems ............................................................................. 40
      3.3.1.1 Testing of R4 Overexpression construct in HEK 293 R4 cells ................. 40
    3.3.2 PiggyBac Vector Systems .................................................................................... 42
List of Tables

Table 1.1. Characteristics of hESC transgene delivery systems ..................................................... 8

Table 1.2. Variants of the tetracycline inducible system .............................................................. 16

Table B.1. shRNA target sequences cloned into vector BP1-KD-GFP for knockdown experiments. .................................................................................................................................. 83
List of Figures

Figure 1.1. PhiC31 targeting and R4-Integrase retargeting of cells .............................................. 10

Figure 2.1. Three-step fusion PCR allows seamless fusion of two or more genetic cassettes. .... 23

Figure 2.2. Iterative workflow of Gateway cloning used in vector system assembly ............... 25

Figure 2.3. Gateway® recombination cloning scheme for the R4-Integrase overexpression construct assembly ........................................................................................................................ 27

Figure 2.4. Gateway® recombination cloning scheme for the R4-Integrase knockdown construct assembly ....................................................................................................................................... 29

Figure 2.5. Dual modification of the native piggyBac plasmid .................................................... 30

Figure 2.6. Gateway® recombination cloning scheme for the piggyBac overexpression construct assembly ........................................................................................................................................ 31

Figure 2.7. Gateway® recombination cloning scheme for the piggyBac knockdown construct assembly ........................................................................................................................................ 33

Figure 3.1. R4-Integrase retargeting of HEK 293 cells generates selectable reporter overexpression clones ................................................................................................................... 42

Figure 3.2. Inducibility in the overexpression system is dependent on doxycycline dose and the Crimson-Puromycin to TetR-KRAB-Zeocin repressor plasmid ratio ................................................. 44

Figure 3.3. Chemical transfection and drug selection of hESCs yields a highly transfected and pluripotent cell population ............................................................................................................ 46

Figure 3.4. Immunoblotting of whole cell and FLAG-immunoprecipitated lysates from transiently transfected HEK 293T cells ........................................................................................................................ 47

Figure 3.5. Drug selection optimization of piggyBac OCT4-knockdown vector transfections in hESCs................................................................................................................................. 49

Figure A.1. LR1 Crimson plasmid map ....................................................................................... 71

Figure A.2. pER4B-DEST plasmid map ....................................................................................... 71
Figure A.3. PB-MCSIIcHS4x2-DEST plasmid map................................................................. 72
Figure A.4. LRi1-Crimson-Puro plasmid map ................................................................. 73
Figure A.5. LRi2-GFP-Zeo plasmid map ...................................................................... 73
Figure A.6. LRi1-Oct4-Puro plasmid map ..................................................................... 73
Figure A.7. LRi2-KD-GFP-Neo plasmid map................................................................. 74
Figure A.8. Genetic elements used in cloning schemes in figures 2.2-2.6......................... 74

Figure B.1. Phase contrast (left) and fluorescence (right) images of HEK 293 R4 cells 24 hours after transfection with the R4 overexpression construct (Figure A.1) coding for E2-crimson with or without Integrase ...................................................................................................................... 75

Figure B.2. Neon™ electroporation of hESCs yields a highly transfected and pluripotent cell population .................................................................................................................................................................. 76

Figure B.3. Stemfect™ transfection of attached and suspended cells yields peripheral versus uniform distribution of transfection, respectively.............................................................................................................. 77

Figure B.5. Immunoblotting assay validates OCT4 antibody specificity............................. 79

Figure B.6. Mean values of chemical transfection and drug selection of hESCs............... 80

Figure B.7. Flow cytometry data for R4-Integrase retargeting of HEK 293 cells............... 81

Figure B.8. Flow cytometry graphs for overexpression inducibility assay. ......................... 82

Figure B.9. Flow cytometry graphs for Stemfect™ chemical transfection assay............. 83
<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>BAC</td>
<td>bacterial artificial chromosome</td>
</tr>
<tr>
<td>BSA</td>
<td>bovine serum albumin</td>
</tr>
<tr>
<td>Dox</td>
<td>doxycycline</td>
</tr>
<tr>
<td>cDNA</td>
<td>complementary DNA</td>
</tr>
<tr>
<td>CDS</td>
<td>coding sequence</td>
</tr>
<tr>
<td>CM</td>
<td>conditioned medium</td>
</tr>
<tr>
<td>CMV</td>
<td>cytomegalovirus</td>
</tr>
<tr>
<td>ChIP-Seq</td>
<td>chromatin immunoprecipitation</td>
</tr>
<tr>
<td>sequencing</td>
<td></td>
</tr>
<tr>
<td>DMEM</td>
<td>Dulbecco’s modified eagle medium</td>
</tr>
<tr>
<td>dNTP</td>
<td>deoxyribonucleotide triphosphate</td>
</tr>
<tr>
<td>dsRNA</td>
<td>double stranded ribonucleic acid</td>
</tr>
<tr>
<td>FBS</td>
<td>fetal bovine serum</td>
</tr>
<tr>
<td>FGF</td>
<td>fibroblast growth factor</td>
</tr>
<tr>
<td>GAPDH</td>
<td>glyceraldehyde-3-phosphate dehydrogenase</td>
</tr>
<tr>
<td>OCT4</td>
<td>octamer-binding transcription factor4</td>
</tr>
<tr>
<td>FACS</td>
<td>fluorescence activated cell-sorted</td>
</tr>
<tr>
<td>ICM</td>
<td>inner cell mass</td>
</tr>
<tr>
<td>IF</td>
<td>immunofluorescence</td>
</tr>
<tr>
<td>IP</td>
<td>immunoprecipitation</td>
</tr>
<tr>
<td>iPSC</td>
<td>induced pluripotent stem cell</td>
</tr>
<tr>
<td>IRES</td>
<td>internal ribosomal entry site</td>
</tr>
<tr>
<td>IVS</td>
<td>intervening sequence</td>
</tr>
<tr>
<td>KRAB</td>
<td>Kruppel associated box</td>
</tr>
<tr>
<td>TR</td>
<td>terminal repeat</td>
</tr>
<tr>
<td>MEF</td>
<td>mouse embryonic fibroblast</td>
</tr>
<tr>
<td>mESC</td>
<td>mouse embryonic stem cell</td>
</tr>
<tr>
<td>miRNA</td>
<td>microribonucleic acid</td>
</tr>
<tr>
<td>mRNA</td>
<td>messenger ribonucleic acid</td>
</tr>
<tr>
<td>ODP</td>
<td>oligodendrocyte precursor</td>
</tr>
<tr>
<td>ORF</td>
<td>open reading frame</td>
</tr>
<tr>
<td>pBase</td>
<td>piggyBac transposase</td>
</tr>
<tr>
<td>PBS</td>
<td>phosphate buffered saline</td>
</tr>
<tr>
<td>PCR</td>
<td>polymerase chain reaction</td>
</tr>
<tr>
<td>PEI</td>
<td>Polyethylenimine</td>
</tr>
<tr>
<td>RISC</td>
<td>RNA Induced Silencing Complex</td>
</tr>
<tr>
<td>RNAi</td>
<td>ribonucleic acid interference</td>
</tr>
<tr>
<td>ROCKi</td>
<td>Rho-associated kinase inhibitor</td>
</tr>
<tr>
<td>RT</td>
<td>room temperature</td>
</tr>
<tr>
<td>rtTA</td>
<td>reverse tetracycline transactivator</td>
</tr>
</tbody>
</table>
SCI  spinal cord injury
SD  standard deviation
shRNA  short hairpin ribonucleic acid
siRNA  small interfering ribonucleic acid

SSEA-3  stage-specific embryonic antigen 3
ssRNA  single stranded ribonucleic acid
ZFN  zinc finger nuclease
Chapter 1
Introduction

1.1 Embryonic Stem Cells

Mammalian embryogenesis begins with multiple cell divisions of the fertilized egg to give rise to a multicellular structure known as the morula. Shortly after, the morula is polarized into an outer region, the trophectoderm, and an inner region defined as the inner cell mass (ICM) [1]. The human ICM is a developmentally specialized group of cells located in the inner region of the spherically-shaped, five-day old embryo known as the blastocyst. Under normal *in vivo* development, cells comprising the ICM differentiate into all the cell types making up the organism proper [2-3]. Embryonic stem cells (ESCs) are isolated *in vitro* from the ICM region of the mammalian blastocyst [4-6]. ESC isolation has been achieved either by immunosurgery whereby exposure of the blastocyst to antiserum against the cell’s species results in trophoblast cell death while leaving the ICM intact [7] or by mechanical dissociation of the ICM from the rest of the blastocyst [8]. The newly isolated ICM cells are cultured *in vitro* under established chemical and physical conditions to maintain their pluripotent state and are subsequently expanded into an ESC line. In culture, embryonic stem cells are characterized by their ability to self-renew as well as differentiate into any type of cell comprising the adult mammal. This latter characteristic is known as pluripotency and is the subject of much ESC research. Specifically, many efforts are geared towards deciphering the pathways governing the cell’s maintenance of pluripotency and developmental differentiation in order to better understand normal and abnormal pathways of human development [9]. In this respect, the transcription factors *OCT4*, *NANOG* and *SOX2* have been identified as central players regulating ESC pluripotency and are currently believed to be part of a larger gene regulatory network that continues to be
characterized [10-12]. Remarkably, the impact of pluripotency network research has been highlighted by discoveries such as the genetic reprogramming of somatic cells into ESC-like cells [13-14]. This technology has in turn allowed the derivation of patient-specific pluripotent cells and thus provided access to unprecedented cellular models of developmental diseases [15]. One other notable method to direct ESC fate involves the use of small molecules to drive differentiation towards specific developmental lineages [16-17]. Together, such approaches provide a promising future both for understanding mammalian development and for therapeutic uses of ESCs. This includes the screening of drugs to abrogate disease phenotypes or to generate specific cell populations for use in cell replacement therapies [18-19].

1.2 Understanding Mammalian Development through ESCs

Since their isolation in 1981 [5-6], mouse embryonic stem cells (mESCs) have served as a prevalent model for the study of human genetic diseases and embryonic development. The demonstration of homologous recombination of transgenic DNA in mESCs [20] paved the path for genomic engineering of the mouse. By transfecting mESCs with exogenous DNA flanked by sequences homologous to a genomic region of interest, researchers began creating numerous cellular models of human disease [21-24]. Subsequently, introducing the transgenic cells into mouse blastocysts gave rise to chimeric mice and propelled the study of human disease into the \textit{in vivo} realm [25]. Knockout mouse models are created by an out-of-frame insertion into the open reading frame (ORF) of the gene of interest to disrupt its translation and are used to study the effect of the absence of a gene’s protein product [22]. Knock-in models allow monitoring of the spatial expression pattern of a gene of interest through an in-frame insertion of a reporter gene into the ORF of the gene of interest [23]. Similarly, mESCs can be gene-targeted with precisely defined point mutations in order to study the mutation in conditional or hypomorphic
contexts [24]. Such transgenic alterations gave rise to numerous disease model cell lines which were, for the first time, both pluripotent and closely related to human development. Lastly, it is worth highlighting that central developmental hESC paradigms stem from mESC discoveries. Notably, roles of core pluripotency regulators OCT4 and NANOG [26-27] along with the systems biology approaches to expand the ESC pluripotency network [28-29] were established and developed in mESCs.

While research in mESCs has contributed to the study of human development and disease, the subsequent isolation of human ESC (hESC) lines by James Thomson [4] has unraveled important differences between the two species ranging from cell surface marker expression (SSEA-4- for mouse vs. SSEA-4+ for human) [4, 30], culture factor dependence (LIF for mouse vs. FGF2 for human) [31-32] and X-chromosome activation status (Xa for mouse vs. Xi for human) [33]. Interestingly, hESCs instead share the aforementioned traits with the more developmentally advanced mouse epiblast stem cells [34]. Given the molecular disparity between the cell lines, differences in gene expression profiles [35] and substantial divergence in embryogenesis between both species [36], it is of particular importance to augment mESC-derived data with hESC studies when studying human development and disease.

1.3 Regenerative Medicine

The most touted promise of stem cells has been their potential to repair or replace tissues degenerated by age or genetic disease. Although cell replacement therapies have been commonplace for much of the 20th century in the form of marrow and blood transfusions, such transplants arise from highly abundant and accessible tissue sources and are limited in the scope of diseases that they can target. With the isolation and vast developmental potency of stem cells as well as the ever-increasing sources [37] and mechanisms for their isolation [38], the number
of targets for stem cell-based regenerative medicine has drastically expanded beyond its hematopoietic beginnings. Nonetheless, the potential of the technology has been limited by issues of immunogenicity, cell availability and tumorigenicity.

1.3.1 Immunogenicity

As is the case in organ transplantations, donor-recipient histocompatibility represents a hurdle in allogeneic transplants of stem cells. While autologous transplantation may overcome the immunological barrier, this approach is plagued with its own disadvantages (discussed in 1.3.2 Cell Availability). Alternatively, the derivation of patient-specific iPSCs theoretically provides a histocompatible source of cells for autologous transplantation. While the immunogenicity of iPSCs and their derivatives remains to be fully characterized, some data have already been generated addressing this subject. Autologous transplantation of iPSC-derived cells has achieved symptom alleviation in immunocompetent sickle-cell mouse models with no immune reaction [39]. Conversely, syngeneic mouse transplantations of iPSCs have been shown to be susceptible to immunorejection [40]. Immunosuppressant regimes specific to cell transplants are currently under study as strategies to attenuate immune reactions [41]. Analogous to the approach employed with organ transplantation, the banking of cells encompassing a vast histocompatibility range [42-43] may be the most immediately feasible approach to address cell immunogenicity and supply challenges.

1.3.2 Cell availability

Whether autologous or allogeneic transplants are considered, a donor cell supply may in some cases be non-existent (chemotherapy), damaged (myocardial infarction), or inaccessible to isolation (neuronal precursor). When cells are available, cell quantity and developmental differentiation thresholds must be met to allow their efficient and safe transplantation. In turn, stringent protocols for expansion and differentiation of clinical-grade cells may be incompatible
with time-sensitive conditions like spinal cord injury (SCI). Lastly, for diseases of genetic origin, autologous transplants would be precluded as the patient would lack genetically normal cells for transplantation. While cell banking would address issues of availability, appropriately differentiated and histocompatible cell sources are needed. Fortunately, somatic stem cell sources and transplantation protocols are plentiful as evidenced by present clinical trials which currently number in the hundreds [44] and at least one mesenchymal stem cell-based product with regulatory approval for graft-versus-host disease treatments [45]. In parallel, hESC-derived cells including cardiomyocytes [46], pancreatic beta cells [47], osteoblasts [48] and oligodendrocyte precursors (ODPs) [18] represent potential sources of therapeutic material. Indeed, ODPs and retinal pigment epithelium cells comprise a small group of hESC-derived cells having been granted clinical trial status for SCI, macular dystrophy and macular degeneration therapies [49].

1.3.3 Tumorigenicity

While mouse teratoma assays are routinely used to ascertain the pluripotency of iPSCs and ESCs [50], the proliferative properties of teratomas also highlight the safety hurdles to be overcome in clinical uses of these and other stem cells. Specifically, protocols for the depletion of undifferentiated cells prior to implantation are of importance. One approach introduces an Oct4-driven suicide genetic cassette into the cell such that depletion is achieved through negative drug selection [51]. The genomic manipulation required for this method may preclude it from clinical use as genomic modifications themselves can increase the propensity of cells to uncontrolled proliferation. Antibody-based depletion protocols appear more advantageous as they are based on antigen expression patterns specific to a cell’s developmental status. Cell-surface antigens such as SSEA-3, SSEA-4, TRA-1-60 and more recently, SSEA-5 [52], are commonly used to define pluripotent cells. These markers, in combination fluorescence-
conjugated antibodies and flow assisted cell sorting, can be used for undifferentiated ex vivo cell depletion in a highly specific manner [52]. Lastly, the use of somatic stem cell sources reduces the risk of tumorigenicity given the rarity of somatic cell-induced tumor formation [53].

1.4 Genetic manipulation of Human Embryonic Stem Cells

An important goal of current hESC research is to identify the genetic regulators controlling the cell’s developmental fate. More specifically, our laboratory aims to understand early human developmental pathways by deciphering physical and regulatory interactions of transcription factors hypothesized to function in maintaining pluripotency or driving differentiation of the cell. A systems biology approach was previously used in mESCs, yielding validated drafts of transcriptional regulatory networks controlling pluripotency and differentiation [28]. By transgenically perturbing expression of transcriptional network factors most closely regulated by the core pluripotency genes in the mouse, i.e. Oct4, Sox2 and Nanog, we similarly hope to deduce the regulatory network of hESCs. Our current approach entails the knockdown and overexpression of candidate transcription factors while monitoring subsequent changes in key aspects of cellular biology. These changes include developmental fate, global gene expression, and physical interactions between the transcription factors and genomic regulatory regions. In mESCs, optimized transfection and culture methods readily yield knock-in [54], knockout [22], knockdown [28] and gene-targeted [55] cells. In contrast, genetic manipulation of hESCs is difficult relative to mESCs on at least two fronts. Firstly, hESCs show increased sensitivity to transfection treatments such as single-cell dissociation, to electroporation and to transfection reagents. Secondly, hESCs exhibit low propensity towards homologous recombination (HR), with mESC-optimized HR protocols in hESCs initially yielding approximately 1 successful recombination event for every $10^7$ cells transfected [56]. To increase hESC survival in culture,
parameters such as media, extracellular attachment matrices and dissociation reagents are continuously being optimized [57-59]. Apoptosis-inhibiting small molecules such as neurotrophins [60], and inhibitors of mitogen-activated protein kinase (MEKi), glycogen synthase kinase (GSKi) are also reported to increase survival after single-cell dissociation [61]. Among these, pre- and post-transfection incubation of the cells in the Rho-associated kinase inhibitor (ROCKi) Y-27632 [58] has been the most widely adopted method. To maximize nuclear uptake of exogenous DNA, advanced transfection methods such as electroporation, nucleofection [62-66] and nanomaterial-based delivery [67] also continue to be developed. Lastly, the use of vector systems known to enhance or actively effect genomic integration of transgenes such as Bacterial Artificial Chromosomes (BACs) [68], Cre/loxP cell lines [69], Zinc Finger Nucleases (ZFNs) [70] and lentiviruses [71] has helped to at least partially overcome the hurdle of low rates of endogenous HR in hESCs. Nonetheless, the low integration efficiencies of BACs, the labor-intensive generation of Cre/loxP cell lines, the genetic insert size limit of lentiviral vectors and the cost-prohibitive nature of ZFN design currently render these methods impractical for high-throughput screening projects aiming to decipher the transcription factor networks controlling stem cell fate. For such a purpose, we have developed vector systems exploiting the integration specificity of the R4 Integrase system or the high integration efficiency of the piggyBac transposase system. Table 1.1 provides a comparative look at gene delivery systems for hESCs.
<table>
<thead>
<tr>
<th>Vector System</th>
<th>Max. Cargo Size</th>
<th>Integration Location</th>
<th>Transfection Efficiency</th>
<th>Silencing</th>
<th>Usage Limits</th>
<th>Integration Reversibility</th>
<th>Copy #/cell</th>
</tr>
</thead>
</table>
| ePiggyBac [72] | 18kb           | • TTAA tetranucleotides  
• Bias to transcriptional start sites | • Transgene size-dependent  
• 72.1% for 18kb transgene | • Insertion site-dependent | • None | • Catalyzed by transposase | • 1-12 (transposase dose dependent) |
| loxP/Cre      | >50kb [73]     | • loxP targeting plasmid is random.  
• Transgene targeting is at loxP site | • In loxP-primed cell lines:  
• 50% [69]  
• 90% [74] | • Insertion site-dependent [69, 74] | • loxP-primed cell lines [69, 74] | • Catalyzed by Cre + HIV-TAT [69, 74] | • 1 to 2 [69] |
| PiggyBac      | At least 11kb [75]  
14.3kb in mESCs [76] | • TTAA tetranucleotides  
• Bias to transcriptional start sites and intragenic regions [75] | • 5000 per 10⁶ transfected [72] | • Insertion site-dependent [75] | • None | • Catalyzed by transposase [75]  
• No footprint | • 1 to 5 [75] |
| R4-Integrase   | At least 13kb [77] | • 13q32.3 locus [78] | • 10-20 per 10⁶ transfected [78] | • No silencing in random differentiation [79] | • Silencing in specialized (dopaminergic neuron) differentiation [79] | • PhiC31-primed cell lines  
• Irreversible [80] | • 3-6 [78] |
| Sleeping Beauty | 10 kb (sandwich conformation in HeLa cells) [81] | • TA dinucleotides [82]  
• No bias | • 5% [82-83] | • Insertion site-dependent | • None | Applicable in all cell types:  
• Catalyzed by transposase  
• TA dinucleotide footprint after excision | • 3-6 [83] |

Table 1.1. Characteristics of hESC transgene delivery systems. All data is for hESCs unless otherwise noted.

1.4.1 R4 Integrase Recombination

The irregularity of random DNA integration when generating transgenic cell lines raises important technical and safety issues. Firstly, epigenetic variance in different areas of the genome can result in varying levels of expression and transgene silencing between one given transfected cell and another. Secondly, transgene integration in or near oncogenes presents an obvious safety hurdle in any technologies with clinical potential given the uncontrolled cell proliferation observed upon activation of oncogenes. In this regard, site-specific DNA integration systems from bacteriophage have been used to control the location of transgene integration. Among the integration systems identified, those endogenous to Streptomyces
temperate phages PhiC31 and R4 stand unique as they mediate recombination between two specific yet non-identical sequences [84-85]. Recombination between attachment sequences originally found in the bacterial host (attB) and the phage (attP) are catalyzed by PhiC31 and R4 integrases in a non-reversible manner whereby each integrase uniquely recognizes its own attB and attP sites. The identification of similar attachment sites in the human genome, termed pseudo-att sites, has enabled the use of the integrases for genomic modification of hESCs. In particular, a pseudo-attP site at the 13q32.3 locus has been identified in hESCs as permissive to transgene expression and refractive to epigenetic silencing [78]. This locus was targeted with an R4-attB-containing plasmid to generate the H9 R4 cell line used in this project (Figure 1.1). The combination of the R4 and PhiC31 integrases comprises the recombination system used in our project and is rooted in two components: recombination sites (pseudo attP, R4-attB and R4-attP) and Integrases (PhiC31 and R4). In a two-step process, a cell-line of interest is first primed for R4 site-specific integration and secondly retargeted with an expression plasmid containing the transgene of interest. In the first step, PhiC31 Integrase catalyzes recombination between a naturally occurring genomic pseudo-attP site and an attB site on a donor plasmid. This donor plasmid contains an active Hygromycin resistance cassette, a promoterless Zeocin resistance cassette as well as an R4-attP recombination site which, upon genomic integration of the plasmid, becomes a recombination landing pad for incoming expression constructs [86-87]. Subsequently, the Hygromycin resistance cassette allows the selection of an R4-primed cell population. In the second step, transgenes of interest are first cloned into an expression construct whose backbone contains an orphan EF1α promoter and an R4-attB site that allows recombination with the R4-attP site in the genome of the R4-primed cell. Upon successful genomic integration of this transgene-containing expression construct, the EF1α promoter is recombined such that it becomes immediately adjacent to the promoterless Zeocin resistance
cassette. Consequently, the EF1α promoter activates Zeocin resistance, allowing for selection of successfully retargeted cells.

**Figure 1.1. PhiC31 targeting and R4-Integrase retargeting of cells.** Source: Jump-In™ TI™ Gateway® Targeted Integration System Manual (Life Technologies).
1.4.2 The piggyBac Transposon

Originally identified in the *Trichoplusia ni* moth, the piggyBac transposon is a genetic element consisting of a ~2.4kb transposase-coding sequence flanked by two terminal repeat (TR) sequences [88]. *In vivo*, the transposase protein catalyzes the excision of its own DNA sequence out of the genome at the TRs followed by insertion of the cut sequence into other genomic TTAA tetranucleotide sequences. As such, the transposase coding sequence is continuously spliced in and out of the host genome in a cut-and-paste manner when in the presence of the transposase protein [89].

Since its discovery, the piggyBac transposon has become a widely used tool for gene manipulation in mammalian cells. For transgenic applications, including the system used in this study, the transposase DNA sequence and flanking TRs are cloned into separate plasmids which are simultaneously used for transfection as part of a two-or-more-vector system. In the first plasmid, the transposase-coding sequence normally found between the TRs is replaced by the transgene of interest. The second plasmid (pBase or helper plasmid) encodes the transposase open reading frame but lacks TRs. By co-transfecting the target cell population with both plasmids, the transposase mediates the transgene’s insertion into random TTAA tetranucleotide genomic sequences. The piggyBac possesses features which have made it an attractive vehicle for transgene delivery. First, the piggyBac’s insertion/excision mechanism is fully reversible, permitting transgene removal when necessary [75]. The excision process leaves no genetic footprint [75, 89-90], potentially permitting a future use of piggyBac in gene or cell replacement therapy. Second, the piggyBac’s 14.3kb cargo size capacity [76] allows the study of relatively large mammalian genes in sophisticated overexpression or knockdown vector systems which themselves harbor several large regulatory cassettes. Impressively, new-generation piggyBac systems have successfully transposed cargo capacity from 18kb in hESCs [72] to 100kb in
mESCs as a piggyBac-BAC hybrid as well as in a currently available commercial inducible vector system [91]. Third, as all steps of transposition are solely catalyzed by piggyBac transposase [89], dependency on the host cell’s machinery is minimized which in turn allows use of the system across many species and cell types. Such features have made piggyBac the vector used in recent high-impact studies including the delivery of transcription factors to reprogram human embryonic fibroblasts into iPSCs [92] and to direct differentiation of hESCs to neural and epithelial cells [72].

1.5 Genetic Insulators & Silencing

A common observation in genetic manipulation of hESCs is the silencing of exogenous genetic elements [79]. This genetic silencing is related to naturally occurring methylation of DNA regulatory regions and deacetylation of DNA-binding proteins like histones [93]. Such chemical modifications to the histone protein structure are used by the cell to regulate gene activity by altering the protein-DNA interactions which in turn dictate the accessibility of transcriptional machinery to genomic regulatory regions and thus elicit enhancement or silencing of neighboring genes.

To circumvent the cell’s epigenetic silencing effect in transgenic applications, hESC genomic regions displaying low or no silencing effects have been identified. Such expression-permissive regions serve as attractive sites for site-directed integration of transgenes through platforms such as R4-Integrase and Cre recombinase [69, 78]. Nonetheless, the extensive epigenetic reorganization that occurs in a differentiating stem cell renders even these regions susceptible to silencing [79]. Consequently, DNA sequences such as the 5’ region of the chicken β-globin locus are routinely incorporated into expression constructs as an additional level of genetic insulation. This 1.2 kb stretch of sequence, better known as chicken DNAse I hypersensitive 4 (cHS4), has
been shown to mitigate expression interference from neighboring promoters and to insulate genes from the chromatin silencing effects of the cell [94]. A heterochromatin-spreading barrier model has been suggested as the mechanism of cHS4 insulating activity and is mediated by the sequence’s recruitment of nucleosome destabilization proteins like histone acetyltransferases (HATs) in conjunction with blocking the binding of histone methylation proteins [95]. In particular, the proteins USF1 and USF2 bind to five sites known as ‘Footprint I-V’ within a 250bp stretch in cHS4 and have been shown to recruit histone acetylators like CREB-binding protein [94-95]. Not surprisingly, histone modifications associated with transcriptional activation such as lysine acetylation (H3K9/K14), di-methylation of H3K4 and tetra-acetylation of histone 4 are commonly observed around cHS4 sequences. As the 1.2 kb cHS4 element shows poor insulating capacity in its singular form [96-97], the cHS4 element is included in our constructs as a 2.4kb doublet.

1.6 RNA Interference

One of the cellular mechanisms to regulate endogenous gene expression is RNA interference (RNAi). While RNAi occurs by several pathways, its ultimate effect is the decrease in cellular levels of a given protein. Endogenous ribonucleic acids known as small interfering RNAs (siRNAs) and microRNAs (miRNAs) mediate two of the prominent pathways for regulating gene cellular expression. Briefly, the mechanism of siRNA-mediated RNAi is as follows: the cellular presence of endogenous or exogenously introduced double stranded RNA (dsRNA) is recognized by the ribonuclease Dicer. Dicer cuts dsRNA into 20-25bp double stranded molecules with 2-3 nucleotide overhangs. The overhangs allow recruitment of a protein complex known as the RNA Induced Silencing Complex (RISC). RISC in turn separates dsRNA into single stranded RNA (ssRNA) and uses the antisense strand to locate and anneal to a messenger
RNA (mRNA) encoding the target gene. Once the RISC-ssRNA-mRNA pairing occurs, the mRNA is degraded by RISC resulting in a subsequent decrease of target protein levels. Alternatively, regulation of gene expression can be mediated by miRNAs. MiRNAs are endogenously produced RNA molecules which are processed into 21bp sequences of dsRNA by proteins including Dicer, Drosha and DGCR8. In addition to mediating mRNA degradation, miRNAs regulate gene expression predominantly by inhibition of translation. The latter mechanism occurs through RISC-mediated binding of the 3' untranslated regions of nascent mRNAs by the fully processed and single stranded miRNAs [98]. Importantly, ChIP-Seq analyses have revealed binding of miRNA promoters by pluripotency nodes Sox2, Oct4 and Nanog, thus linking miRNA to the regulation of ESC pluripotency and differentiation [99]. Since its identification in *C. elegans* [100], RNAi has been widely used in molecular genetic studies by improving on the traditional knockout gene approach to perform loss-of-function assays [101-102] as well for directed differentiation of hESCs [72, 103-104]. The resource-intensive nature of the knockout approach combined with the notoriously low efficiency of HR in hESCs lends support to the use of RNAi as an efficient alternative to achieve suppression of protein expression.

As a laboratory tool, RNAi is most commonly applied in one of two ways: as synthetic dsRNA oligonucleotides or as part of a vector that allows *de novo* synthesis of short hairpin RNAs (shRNA). While potent in achieving gene knockdown [105], cell transfection with synthetic dsRNA yields transient knockdown effects with maximum knockdown times ranging from 48 to 72 hours post-transfection [105-106]. In contrast, prolonged or permanent knockdown can be achieved by stable transfection of an shRNA-coding vector. shRNA-based interference involves cloning a short (20-30bp) stretch of the target gene’s DNA coding sequence downstream of an appropriate RNA polymerase III promoter such as H1. By stable or transient transfection of such
a vector and as part of a doxycycline-inducible vector system, the shRNA-based approach chosen for our project is designed to effect temporal and modular control of gene knockdown activity.

1.7 Genetic Elements

A number of genetic elements were used in our construct design plans (Figures 2.2-2.6), each providing a specific function:

The **EF1α** promoter. In its native state, this pol II promoter drives the constitutive expression of elongation factor 1 alpha protein which itself promotes aminoacyl-tRNA binding to the ribosome as part of cellular peptide synthesis [107]. Consistent with the ubiquitous nature of translation, the EF1α promoter drives robust transgene expression in a number of human cell types [108] and in long-term hESC culture [109-111]. As part of our overexpressor vector systems, EF1α is directly fused to a pair of tetracycline operators (TetO2) conferring tetracycline-inducible transgene overexpression whereby activation occurs in the presence of tetracycline and repression occurs in its absence.

The **TetR-KRAB** is a fusion between the Tn10-derived prokaryotic tetracycline repressor (TetR) DNA binding domain and the Krupel-associated box (KRAB) transcription repressor domain of the human Kox1 zinc finger protein [112]. This fusion protein binds the tet operator (TetO2) sequences found adjacent to the EF1α promoter and its mechanism of repression consists of heterochromatin protein recruitment. The TetR-KRAB as used in our tetracycline inducible vector systems is functionally equivalent to the more commonly known tTS repressor [113] while having stronger DNA binding properties that allow tighter repression. The tTS protein is itself a fusion between a doxycycline-dependent DNA binding domain (t) and a transcription regulation domain (TS). Mutagenesis of the two domains has given rise to various tetracycline-
inducible systems comprised of variant fusion proteins differing in their DNA-binding properties in the presence of doxycycline and their effect on transcription (Table 1.2).

<table>
<thead>
<tr>
<th>DNA-binding Domain</th>
<th>Transcription Regulation Domain</th>
</tr>
</thead>
<tbody>
<tr>
<td>t (tet)</td>
<td>TA (transcriptional activator)</td>
</tr>
<tr>
<td>DNA-</td>
<td>tTA –</td>
</tr>
<tr>
<td>binding</td>
<td>- DNA binding inhibited by Dox</td>
</tr>
<tr>
<td>Domain</td>
<td>- Activates transcription</td>
</tr>
<tr>
<td></td>
<td>- Tet-Off system</td>
</tr>
<tr>
<td>rt (reverse tet)</td>
<td>TS (transcriptional silencer)</td>
</tr>
<tr>
<td></td>
<td>tTS –</td>
</tr>
<tr>
<td></td>
<td>- DNA binding inhibited by Dox</td>
</tr>
<tr>
<td></td>
<td>- Represses transcription</td>
</tr>
<tr>
<td></td>
<td>- Tet-On system</td>
</tr>
<tr>
<td>rtTA –</td>
<td>- DNA binding promoted by Dox</td>
</tr>
<tr>
<td></td>
<td>- Activates transcription</td>
</tr>
<tr>
<td></td>
<td>- Tet-On system</td>
</tr>
</tbody>
</table>

Table 1.2. Variants of the tetracycline inducible system.

The rtTA is a fusion protein between the tet operator-binding protein unit known as reverse tet-repressor (rTetR) derived from *E. coli* and the VP16 transactivator unit derived from the herpes simplex virus [114]. rtTA activates transgene transcription through rTetR-mediated TetO2 binding and VP16-mediated recruitment and interaction with transcriptional machinery [115-117]. Our TetR-KRAB repressor-based system and alternate rtTA-driven expression systems each comprise variations of Tet-On inducibility [114]. However, whereas the TetR-KRAB repressor system’s induced state is mediated by inhibition of repression, the rtTA protein instead actively promotes expression. Also in contrast to the TetR-KRAB, maximal expression in rtTA systems is not limited by the regulatory strength of the transgene’s promoter and is instead actively driven by the VP16 activator.
The Polyhistidine-FLAG tag is a dual high-affinity protein purification tag that is fused to the ORF of interest. The polyhistidine consists of six consecutive histidines whereas the FLAG tag has the amino acid sequence of DYKDDDDK. The tag is especially useful in instances where the antibodies for the proteins of interest are not readily available or show poor affinity against the proteins. In our schemes, the tag is either directly fused by PCR to the protein of interest’s C-terminal (Figures 2.5-2.6) or in its own Gateway® cassette (Figures 2.3-2.4).

The cHS4 insulator is a 1.2Kb DNA fragment initially identified at the 5’ end of the chicken β-globin gene [96] which insulates adjacent genes from the silencing effects of genomic chromatin or regulatory interference from neighboring promoters. More specifically, the insulator acts as a barrier to the spread of heterochromatin that is common in vertebrate genomes. In the piggyBac system, the insulator has been cloned into the transposon’s backbone immediately upstream of all other genetic cassettes (Figure A.3.). As the R4-Integrase system is a one-vector system, the cHS4 has been placed between the EF1α promoter controlling TetR-KRAB expression and the more upstream EF1α-TO2 or H1-TO7 promoters regulating cDNA or shRNA expression, respectively.

The ccdB-CmR cassette provides a bacterial positive/negative selection system for Gateway® cloning and does not comprise any of the vectors used for hESC transfection. The CcdB gene acts as a bacterial suicide gene causing gyrase-mediated linearization of the cell’s DNA [118] and allowing negative selection of DH5α E. coli. A CcdB-survival strain of E. coli (Life Technologies) is used when propagation of clones containing the CcdB gene is necessary in intermediate steps of the cloning scheme. The CmR cassette provides chloramphenicol resistance and is used for positive selection similar to other commonly used antibiotic selection cassettes.
The R4 Integrase *attB* recombination site located in the backbone of all R4-Integrase system constructs allows for site-specific genomic integration of the constructs. Consequently, hESC clones having an identical and well characterized genomic integration sites are produced.

The PiggyBac Terminal Repeats are located on the piggyBac transposon vector flanking the genetic cargo of interest (Figure A.3). These sites serve as recombination sites used by the piggyBac transposase for random integration of the transgenic cargo into TTAA sequences found throughout the hESC genome [119].

The T2A peptide is an 18 amino acid tag originally isolated from the *Thosea asigna* virus [120], and is used for polycistronic expression of multiple genes from a single promoter in mammalian cells. This is accomplished by fusing two or more coding sequences (CDS) of interest in frame with, and separated by, the T2A coding sequence, e.g. Gene 1 – T2A – Gene 2. Translation begins from the Gene 1 CDS and into the end of the T2A CDS, at which point the conformation of the T2A peptide causes the release of the Gene1 – T2A fusion peptide from the ribosome while still allowing for continued translation of Gene 2 [121]. This tag was recently used for polycistronic expression of the Yamanaka factors in piggyBac-based cell reprogramming [92].

IRES is an internal ribosome entry site (IRES) sequence of the encephalomyocarditis virus genome which permits polycistronic gene expression in mammalian cells. By promoting ribosome recruitment to its own sequence, IRES mediates initiation of translation at the location in which it appears on mRNA. Given this mechanism, constructs including IRES are assembled with specific design criteria. Firstly, transcriptional termination sequences such as the poly-A tail are omitted from all genes except the one that is most distal from the promoter. This allows the synthesis of a polycistronic mRNA. Secondly, IRES sequences are placed between all coding sequences of interest such that polycistronic translation is initiated on the polycistronic mRNA.
An intervening sequence (IVS) is included immediately upstream of each IRES as it has been shown to contribute to mRNA stabilization [124].

1.8 Project Objectives and Hypothesis

Given the strengths of the R4-Integrase and piggyBac genomic integration platforms and the wide availability of genetic elements for transgene regulation, I hypothesize that building overexpression and knockdown genetic vector systems based on R4-Integrase and piggyBac platforms will allow robust regulation of transgene expression in hESCs. To test the hypothesis, I have identified the following aims:

Aim 1. To create modular, inducible, protein-tagged and epigenetically insulated overexpression/shRNA-knockdown vector systems based on the R4-Integrase and piggyBac genomic integration platforms.

Aim 2. To test construct functionality by inducing transgenic fluorescence reporter and transcription factor overexpression and knockdown in stably transfected hESCs. In this respect, genetic elements are to be assembled into R4 or piggyBac platforms using a modular cloning scheme that facilitates their rearrangement or replacement. Functional testing of the constructs should identify the most robust genetic elements for optimized transgene expression in hESCs. Optimized vector systems should in turn aid in the elucidation of genetic pathways central to regulating pluripotency of hESCs.
Chapter 2

Generation of modular, inducible, protein-tagged, selectable and epigenetically insulated overexpression/shRNA-knockdown vector systems based on R4 Integrase and piggyBac platforms

2.1 Overview

The development and differentiation of embryonic stem cells is dependent on precise patterns of physical and temporal interactions between a multitude of factors which together dictate gene expression. An important area of focus in current embryonic stem cell developmental studies centers on deciphering the identity, function and corresponding interactions between genes maintaining pluripotency or driving differentiation [9]. Our lab has taken a systems biology approach to identifying and testing gene regulatory networks of hESC pluripotency by examining protein-DNA interactions as well as transcriptional and translational regulation. Central to this approach is the transgenic perturbation (overexpression and knockdown) of developmentally relevant genes to monitor the resulting changes in global gene expression patterns, protein-DNA interactions and the cell’s developmental state [28]. To achieve this, I have developed inducible vector systems to allow control of the timing, dose and identity of expression of transgenic transcription factors. Enhanced control of transgene expression level and timing is achieved through TetR-KRAB repression of the transgene of interest. Neomycin, Puromycin or Zeocin resistance cassettes allow drug selection of successfully transfected cells to more specifically focus on biological changes that may otherwise be masked by neighboring untransfected cells. The polyhistidine-FLAG dual protein tag allows immunoprecipitation of overexpressed proteins for mass spectrometry, chromatin immunoprecipitation sequencing (ChIP-Seq) and is particularly useful when antibodies for the gene of interest are unavailable or of poor quality. The genetic insulators flanking the gene of interest minimize epigenetic silencing
or interference arising from the vector’s insertion site. The Gateway® technology used to construct the system allows for high-throughput cloning of genes of interest as well as allowing customization of the vector system by easy addition or removal of genetic components such as fluorescent tagging. Lastly, all vectors function under two platforms (R4-Integrase & piggyBac transposase) designed to enhance vector integration into the genome.

2.2 Materials and Methods

2.2.1 PCR Fusion of Genetic Elements

While Multisite Gateway® cloning allows for the arrangement of any desired genetic elements in a specific order, fusion PCR [125] can also achieve this purpose when Gateway® is implausible or impractical. In the case of the constructs generated here, the number of cassettes to be cloned exceeded the number of available pDONR vectors and att recombination sites. Specifically, elements fused with this technique and the constructs in which they were used were:

- EF1α + dcGFP-Zeo → piggyBac Overexpressor
- TetR-KRAB + sv40 pA → piggyBac Overexpressor,
- ORF + Histidine-FLAG → piggyBac Overexpressor
- EF1α + Puromycin resistance → piggyBac Knockdown
- EF1α + Neomycin resistance → piggyBac Knockdown
- Polyhistidine-FLAG tag + sv40 pA → R4 Overexpressor
- cHS4x2 + EF1α → R4 Overexpressor
- TetR-KRAB + sv40 pA → R4 Overexpressor

Conventional PCR and fusion PCR differ only in the design of the primers and the number of reactions needed to generate the final product. Three reactions are carried out whereby reactions
1 and 2 generate products with mutually overlapping sequences which are in turn simultaneously used as templates and reamplified in reaction 3 to generate the fused product (Figure 2.1). Mutually overlapping products of reactions 1 & 2 are generated using primers that are 40bp long whereby 20bp are template-specific and 20bp comprise the sequence overhang specific to the second element to be fused.

**Reaction 1** uses a conventional Gateway® forward primer and a Fusion PCR reverse primer. The reverse primer must be designed such that its first 20bp are the reverse complement of the first 20bp of the template to be used in Reaction 2. The last 20bp of the primer are the reverse complement of the template to be used in Reaction 1.

**Reaction 2** uses a Fusion PCR forward primer and a conventional Gateway® reverse primer. The forward primer must be designed such that its first 20bp are identical to the last 20bp of the template to be used in Reaction 1. The last 20bp of the primer are identical to the first 20bp of the template to be used in Reaction 2.

**Reaction 3** uses the conventional PCR forward and reverse primers used in Reactions 1 & 2.
Figure 2.1. Three-step fusion PCR allows seamless fusion of two or more genetic cassettes. PCR amplification of two genetic cassettes (reactions 1 and 2) using primers which are partially homologous to one cassette and partially homologous to the adjacent cassette yields products which are in turn mixed and used as templates in PCR reaction 3. The third reaction yields a final product comprised of the two cassettes.

2.2.2 Gateway® Cloning

All components used in the vectors presented here were made Gateway-compatible by PCR amplification using plasmids as templates. Gateway® PCR primers were designed using Vector NTI Advance™ 11.0 (Life Technologies) using the appropriate attB sites. The cassettes used and their respective sources are as follows: **EF1α-TO2** and **sv40 pA** were amplified from pER4B-EF1α-TO-GFP-pA-(IlpCAG-TetR)x2 (Life Technologies). The **OCT4 ORF** was amplified from pMXs-hOCT4 (Plath, Addgene Plasmid 17964). **His-FLAG** was amplified from pJL214 (Stanford Lab). **cHS4** was amplified from pDONR201 R6-cHS4x2-R5 (Invitrogen).
TetR-KRAB was amplified from pLVCT-tTRKRAB (Addgene, Plasmid 11643). IVS-IRES-Neo and IVS-IRES-Puro were respectively amplified from pIRESNeo3 and pIRESNeo3 (Clontech). GFP-Zeo was amplified from pGreenZeo lenti-reporter (System Biosciences). H1 was amplified from pLVTHM (Addgene, Plasmid 12247). The T2A peptide from pGreenZeo lenti-reporter (System Biosciences) was directly synthesized as complementary oligonucleotides with BamHI and XhoI restriction sites at the 5’ and 3’ ends, respectively and ligated as follows: the oligonucleotides were mixed at a 1:1 ratio, denatured at 98°C for 5 min, followed by annealing to 25°C over 30 min. Agarose gel-purified PCR products were recombined with intermediary (pDONR) vectors containing attP sites catalyzed by a BP recombinase (Life Technologies) to yield BP products (entry vectors). Lastly, all resulting BP products were recombined through an LR recombinase-catalyzed reaction into destination (pDEST) vectors containing either an R4-attB site or piggyBac TRs, each of which mediates transgene genomic integration. All Life Technologies plasmids were obtained from Jon Chesnut (Invitrogen). Figure 2.1 illustrates the sequence of techniques used in the assembly of our vector systems.
Figure 2.2. Iterative workflow of Gateway cloning used in vector system assembly. Recombination product verification consists of a preliminary diagnostic restriction digest followed by nucleotide sequencing.

2.2.3 Sticky-End Ligation of shRNAs

OCT4 shRNA sequences were obtained from the Broad Institute’s RNAi Consortium Database [126] and chemically synthesized (Life Technologies) with MluI and Clal restriction sites at the 5’ and 3’ ends, respectively. Ligation was mediated by Quick T4 DNA Ligase (New England Biolabs). See Table B.1 for shRNA sequences.
2.2.4 PCR

All PCR reactions were carried out on a Mastercycler (Eppendorf) using the protocol based on Phusion High-Fidelity DNA Polymerase (New England Biolabs) recommendations whereby:

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Annealing temperature</td>
<td>+3°C of lowest primer T_m</td>
</tr>
<tr>
<td>Extension temperature</td>
<td>72°C</td>
</tr>
<tr>
<td>Denaturing temperature</td>
<td>98°C</td>
</tr>
<tr>
<td>Extension time</td>
<td>20s/1Kb</td>
</tr>
</tbody>
</table>

Oligonucleotide T_m was calculated using Vector NTI Advance 11.0 software (Invitrogen). For primers ≥40bp, e.g. Gateway® primers, the %GC T_m value was used whereas for primers <40bp, the thermodynamic T_m value was used, as recommended by the software.

Reagent quantities used in all PCR reactions:

<table>
<thead>
<tr>
<th>Reagent</th>
<th>Volume (µl)</th>
</tr>
</thead>
<tbody>
<tr>
<td>dH2O</td>
<td>31</td>
</tr>
<tr>
<td>HF Buffer (5X)</td>
<td>10</td>
</tr>
<tr>
<td>dNTP (10 mM)</td>
<td>1</td>
</tr>
<tr>
<td>Primer 1 (10 µM)</td>
<td>2.5</td>
</tr>
<tr>
<td>Primer 2 (10 µM)</td>
<td>2.5</td>
</tr>
<tr>
<td>Template (5 ng/µl)*</td>
<td>2.5</td>
</tr>
<tr>
<td>*Taq Polymerase (2U/µl)</td>
<td>0.5</td>
</tr>
</tbody>
</table>

*Template for Fusion PCR reaction 3 consisted of the gel-purified products from reactions 1&2

2.2.5 DNA Sequencing

All vector sequencing reactions were performed by The Centre for Applied Genomics (Toronto, Canada) or by the StemCore Facility (Ottawa, Canada) using universal primers provided by the facility or custom primers designed in our lab. Alignment of sequences was done with AlignX, a component of Vector NTI Advance 11.0.

2.2.6 Restriction Digest Analysis of Constructs

All digests were done with restriction enzymes from New England Biosciences according to manufacturer directions.
2.3 Results

All abbreviations of genetic elements depicted in (Figures 2.2-2.6) are listed on Figure A.7. All constructs were sequenced prior to use in transfection.

2.3.1 Assembly of Overexpression R4 Vector System

The final R4 Overexpression construct (Figure 2.2) is a modular platform into which new cDNAs are cloned to generate inducible and genetically stable constructs for each gene of interest. Generation of new overexpression constructs is a two-step process whereby the PCR product of the cDNA of interest is cloned into a pDONR vector followed by LR cloning with BPM at the ‘Point of Entry’. R4-overexpressor constructs coding for *DACH1*, *EVX1* or Crimson ORFs were constructed as well as all intermediates shown with the exception of BPM.

![Diagram of R4 Overexpressor](image)

Figure 2.3. Gateway® recombination cloning scheme for the R4-Integrase overexpression construct assembly.
2.3.2 Assembly of shRNA Knockdown R4 Vector System

The design scheme of the knockdown R4 vector system (Figure 2.4) borrows most of its components from the overexpression counterpart. The most notable difference arises in the substitution of the EF1α promoter with the pol III promoter H1 which is required for transcription of shRNAs. Similar to the overexpression construct, a TetR-binding site (TetO7) is fused directly upstream of H1, allowing TetR-based inducibility of the system. The final construct (LR2-KD) is an inducible and genetically insulated plasmid. Generation of subsequent knockdown constructs begins at Point of Entry by a two-step process of ligating the gene of interest-specific 20-mer into BP1-KD followed by shuttling of the product into a R1-R2 Destination vector in a 2-way Gateway® reaction along with product BP8. R4 knockdown vectors were built encoding DACH1 shRNAs as well as all intermediates shown.
2.3.3 Genetic Insulation and Gateway compatibility of piggyBac destination vector

To increase the versatility of our piggyBac vector systems, two modifications (Figure 2.5) were made to the native piggyBac carrier plasmid (PB-MCSII) to create a modified version of the piggyBac which we have termed PB-MCSIIcHS4x2-DEST (Figure A.3). First, a pair of cHS4 genetic insulators were cloned upstream of the transgene cloning site. This alteration maximizes stable expression of the transgenes by reducing the effects of cellular epigenetic silencing often seen after transfection of hESCs [127]. Second, the piggyBac was made compatible with the Gateway® cloning system through blunt ligation of a cassette comprised of attR recombination sites flanking the ccdB and chloramphenicol-resistance gene. This second modification allows all
genetic cassettes of interest to be shuttled into the piggyBac backbone in a single Gateway® LR recombination reaction.

Figure 2.5. Dual modification of the native piggyBac plasmid. (Arrow 1) Genetic insulator sequences (cHS4) were cloned into the native piggyBac (PB-MCSII) carrier plasmid by BamHI and XhoI sticky end restriction digest and ligation. (Arrow 2) The Gateway®-compatibility cassette (att1-CmR-ccdB-attR2) was similarly cloned immediately downstream of the insulator sequences.

2.3.4 Assembly of Overexpression piggyBac Vector System

Given the 14.3 kb transgene size maximum of the piggyBac system [76], our overexpression construct was divided into a two-construct system (Figure 2.6) whereby cells are co-transfected with both plasmids. The first construct contains the EF1α-TO2 promoter, the ORF of interest, and an antibiotic (Neomycin or Puromycin) resistance cassette. This construct is 9.8 kb plus the open reading frame, allowing for incorporation of up to a 4.5 kb ORF of interest. The ORF of interest can be incorporated on its own or it can be PCR-fused to immune or fluorescence tags (e.g. polyhistidine-FLAG, yellow fluorescence protein) according to experimental requirements.
The second construct carries another EF1α driving simultaneous expression of GFP, Zeocin, and TetR-KRAB, whereby GFP/Zeocin serve as fluorescence/resistance markers for the construct and TetR-KRAB confers inducibility by repressing expression of the ORF of interest in the first construct. The final construct is a modular platform into which new cDNAs are cloned to generate inducible and genetically stable constructs for each gene of interest. Generation of new overexpression constructs is a two-step process whereby the PCR product of the cDNA of interest is cloned into the P5-P3r DONR vector followed by a 3-way LR cloning into the R1-R2 Destination Vector to generate the new LRi1-ORF final product. Overexpression constructs containing DACH1, EVX1, ZBTB8A, OCT4 and Crimson coding sequences as well as all intermediate vectors shown have been assembled.

![Gateway® recombination cloning scheme for the piggyBac overexpression construct assembly.](image-url)
2.3.5 Assembly of shRNA Knockdown piggyBac Vector System

The cloning scheme for the piggyBac shRNA knockdown system is depicted in Figure 2.7. The knockdown of gene expression in all our vectors systems is shRNA-mediated whereby a dsDNA oligonucleotide corresponding to the shRNA of the gene of interest is ligated downstream of a TO-H1 promoter (Figure A.7). This first cassette consists of the shRNA-specific dsDNA and the TO-H1 promoter driving shRNA expression. All shRNA-specific-dsDNAs are synthesized with MluI/ClaI restriction sites at the ends to allow ligation into the H1-TO7-containing vector. The second cassette contains the EF1α promoter driving simultaneous expression of GFP and a Puromycin resistance gene allowing selection and visualization of successfully transfected cells. The third cassette contains a pair of cHS4 genetic insulators and a second EF1α promoter to drive expression of the element in the fourth cassette. The fourth cassette contains the TetR-KRAB coding sequence for repression of TO7-H1 promoter activity in the first cassette. The four genetic cassettes comprise the final piggyBac knockdown two-vector system. Generation of subsequent knockdown constructs begins at Point of Entry by a two-step process of ligating the gene of interest-specific 20-mer into BP1-KD followed by shuttling of the product into a R1-R2 piggyBac Destination vector in a 2-way Gateway® reaction, along with the BP2-KD-Neo plasmid to generate the new LRi2-KD product. Knockdown vectors coding for DACH1, EVX1, ZBTB8A, OCT4 and Crimson shRNAs as well as all intermediate vectors shown have been assembled.
Figure 2.7. Gateway® recombination cloning scheme for the piggyBac knockdown construct assembly.
Chapter 3
Functional testing of vector systems in HEK 293T and hESCs

3.1 Overview
To functionally test the R4-Integrase and piggyBac vector systems described in Chapter 2, HEK 293T and H9 hESCs were transfected with vectors encoding fluorescence reporters or transcription factors. Stable R4-Integrase-based transfection of hESCs was unsuccessful under all conditions and vectors and as a result, the piggyBac was chosen as the primary platform for gene delivery in hESCs. Additional experiments aimed at troubleshooting the unsuccessful hESC transfections were performed on the more easily transfectable HEK 293T line. In the piggyBac system, we sought to individually test the functionality of the genetic components included in the vector. Firstly, the platform’s efficacy in generating stably transfected hESC lines was tested through a newly developed chemical transfection method and stringent drug selection regime. Secondly, successful immunoprecipitation of the FLAG tag was demonstrated in the overexpressor system. Thirdly, the system’s inducibility and leakiness were tested under a range of doxycycline concentrations and repressor protein quantities, respectively. Lastly, the first functional test in hESCs was performed in the form of an OCT4 knockdown assay.

3.2 Materials and Methods
3.2.1 The H9 Cell Line
The hESC H9 line is officially classified as WA09 with karyotype/blood type 46,XX/A+. The cell line was originally derived by Thomson et al. [4] and distributed by WiCell (USA). All experiments were done on cells ranging from passage numbers 35 to 65.
3.2.2 The H9 R4 Cell Line
The H9 R4 cell line was obtained from Jon Chesnut at Life Technologies. The line is a subclone of the H9 hESC line which has undergone PhiC31 genomic targeting at the 13q32.3 locus [78]. Cells were cultured as described in 3.2.5.1 Mouse embryonic fibroblast (MEF)-supported culture in 50µg/ml Hygromycin B (Life Technologies). Post-electroporation clone selection was done with 2µg/ml Zeocin (Life Technologies).

3.2.3 HEK 293T Cell Culture
The human embryonic kidney 293T cell line is a subclone of the HEK 293 line [128] and harbors the sv40 large T-antigen DNA sequence [129] which allows episomal replication of plasmids. Cells were cultured directly on tissue culture plates (Falcon) in monolayer format at 37°C in 5% CO₂. HEK 293T medium consisted of 88% DMEM, 10% FBS, 1% Glutamax and 1% Penicillin-Streptomycin. For passaging and expansion, single-cell dissociation was done with 0.25% Trypsin:EDTA. All medium and reagents obtained from Life Technologies.

3.2.4 The HEK 293 R4 Cell Line
The human embryonic kidney 293 R4 cell line was obtained from Michael Poderyckyi at Life Technologies. The line was derived by single-site PhiC31 genomic targeting [86] of HEK 293 cells [128] with a stably integrated vector containing the R4 \textit{attP} recombination site. Cells were cultured as described in 3.2.3 \textit{HEK 293T Cell Culture} with the addition of 50µg/ml Hygromycin B (Life Technologies). Post-transfection clone selection was done with 5µg/ml Blasticidin (Life Technologies).

3.2.5 hESC Cell Culture
All hESC culture was done at 37°C in 5% CO₂ and fresh medium replenished every 24 hours. All cell dissociation for normal passaging was done with 1mg/ml Collagenase Type IV (StemCell
Technologies). All tissue culture reagents obtained from Life Technologies unless otherwise noted.

3.2.5.1 Mouse embryonic fibroblast (MEF)-supported hESC culture
Cells were cultured on mitotically inactivated MEFs in hESC medium: 77% DMEM/F12, 20% Knockout Serum Replacement, 1% Glutamax, 1% MEM Non-essential Amino Acids, 1% Penicillin-Streptomycin, 0.1% β-mercaptoethanol and 10ng/ml bFGF.

3.2.5.2 Feeder-free hESC culture with conditioned medium (CM)
hESCs grown on MEF-supported culture were passaged onto Matrigel (BD) coated plates and fed using CM. Conditioned medium consists of hESC medium containing MEF-secreted growth factors and was prepared by culturing MEFs in hESC medium for 24 hours and re-collecting. CM was supplemented with additional 20ng/ml bFGF prior to use on hESC culture.

3.2.5.3 Feeder-free hESC culture with chemically defined E8 medium
hESCs grown on MEF-supported culture were passaged onto Matrigel (BD) coated plates and cultured with the chemically defined E8 medium developed by Chen et al [130] consisting of: DMEM/F12, 64mg/L Ascorbic Acid-2-Phosphate Magnesium (Sigma), 14µg/L Sodium Selenium, 100µg/L bFGF, 19.4 mg/L Insulin (Roche), 10.7mg/L Transferrin (Sigma), 2µg/L TGFβ-1, 543mg/L Sodium Bicarbonate (Sigma), 1mg/L Gentamicin.

3.2.6 Cell Transfection
3.2.6.1 hESC Electroporation
One hour prior to electroporation, cells were treated with E8 medium containing 10µM Y-27632 (Tocris Bioscience) but lacking Gentamicin, followed by 5 minutes of single-cell dissociation using 0.05% Trypsin:EDTA. All hESC electroporations were done with Neon™ Transfection System (Life Technologies) according to manufacturer’s manual with the following variables remaining constant:
- **10-cm dish electroporations**: 1 x 10^6 cells/dish + 10 µg total DNA/dish using the 100 µl Neon™ kit with settings of 950V, 30ms, 2 pulses.

- **12-well plate electroporations**: 1.5 x 10^5 cells/well + 3 µg total DNA/well using the 10 µl Neon™ kit with settings of 850V, 30ms, 1 pulse.

- **PiggyBac-based electroporations**: a 2:1 mass ratio of piggyBac transposase plasmid to transgene-containing plasmid was used. Following electroporation, cells were re-seeded on Matrigel-coated plates and cultured in E8 medium containing 10 µM Y-27632 and lacking Gentamicin for 24 hours. Subsequently, the cells were cultured in normal E8 medium for three to seven more days prior to commencing antibiotic selection of clones.

- **R4 Integrase-based electroporations**: a 2:1 mass ratio of pJTI™-R4 Integrase plasmid to transgene-containing plasmid was used. Following electroporation, cells were re-seeded and cultured in MEF-supported culture with medium containing 10 µM Y-27632 and lacking Penicillin-Streptomycin for 24 hours. Subsequently, the cells were cultured in normal MEF-supported medium for 4 more days prior to commencing antibiotic selection of clones.

3.2.6.2 **Chemical Stemfect™ Transfection of H9 hESCs**

While the Stemfect™ transfection reagent is manufactured for transfection of attached cells, a protocol for transfection of cells in suspension was developed in our lab as follows: One hour prior to electroporation, cells were treated with E8 medium containing 10 µM Y-27632 (Tocris Bioscience) but lacking Gentamicin, followed by 5 minutes of single-cell dissociation using 0.05% Trypsin:EDTA. Trypsin was neutralized by resuspending the cell suspension in MEF medium. The cells were spun at 200g and the MEF medium decanted. The cell pellet was resuspended in E8 medium (1.3 ml per transfected well at a density of 4.9 x 10^5 cells/well) containing 10 µM Y-27632 (Tocris Bioscience) but lacking Gentamicin. The cell suspension was
seeded on matrigel-coated plates immediately prior to transfection. DNA (8.4µg total DNA/well for OCT4 knockdown transfections, 6.7µg total DNA/well for GFP-Zeo transfections) was mixed with 25µl of Stemfect™ buffer while 0.67µl of Stemfect™ polymer was mixed with 25µl of Stemfect™ buffer. The two mixtures were in turn mixed together, vortexed for 10 seconds, incubated for at least 10 minutes and added to the freshly seeded cells in suspension. Sixteen hours later, the polymer-containing medium was replaced with E8 medium containing 10µM Y-27632 (Tocris Bioscience) but lacking Gentamicin. The cells were fed with standard E8 medium 8 hours later and every 24 hours thereafter. A 2:1 mass ratio of piggyBac transposase plasmid to transgene-containing plasmid was used for all transfections. All parameters listed above are per well of a 12-well plate (BD).

3.2.6.3 Chemical Transfection of HEK 293T & HEK293 R4 Cells
Transfection was done on adherent monolayer culture at 50-70% confluence using 1µg/µl Polyethylenimine (PEI) of linear MW ~25,000 (Polysciences) according to the ExGen 500 in vitro Transfection Reagent Protocol (Fermentas).

3.2.7 Cell Harvest for fluorescence analysis
Single cell dissociation was done with .25% Trypsin:EDTA (Life Technologies) for 3 min at 37°C, followed by neutralization with 5% FBS in PBS + 0.1% sodium azide. Cells were then centrifuged and resuspended in 5% FBS in PBS + 0.1% sodium azide prior to flow cytometry.

3.2.7.1 Cell Surface (SSEA-3 and TRA-1-60) Immunostaining
Cells were harvested as described in 3.2.7 Harvest for fluorescence analysis followed by incubation at 4°C with either primary antibody (mouse monoclonal αTRA-1-60 - Millipore MAB4360, used at 1:10) followed by secondary antibody (Donkey 555 anti mouse – Life Technologies A31570, used at 1:100) or with a conjugated primary antibody (rat monoclonal IgM SSEA-3-647 - BD 561145, used at 1:10) for 20 minutes each.
3.2.7.2 OCT4 Immunostaining for flow Cytometry
Cells were harvested as described in 3.2.7 *Harvest for fluorescence analysis* and fixed by 15 minute incubation in 4% Paraformaldehyde followed by permeabilization with 0.3% Triton X-100 in PBS for 15 min at room temperature. Permeabilized cells were stained at 4°C for 30 minutes in the dark with primary antibody mouse monoclonal αOCT4 antibody (BD 611203, used at 1:500) followed by secondary antibody IgG1- 488 Goat Anti-Mouse (Life Technologies – A21121, used at 1:100).

3.2.8 OCT4 Immunostaining for *in situ* immunofluorescence imaging
Cells were fixed by incubation in 4% Paraformaldehyde for 15 minutes, blocked for 30 minutes at room temperature with 5% FBS (Life Technologies) in PBS (Life Technologies) followed by permeabilization with 0.3% Triton X-100 in PBS for 1 hour at 4°C. Permeabilized cells were co-stained with primary mouse monoclonal αOCT4 antibody (BD 611203, used at 1:500) for 1 hour at room temperature followed by 1 hour co-staining with secondary antibody (anti-mouse IgG1-488 – Life Technologies A21121, used at 1:100) and Hoechst 33342 (Life Technologies, used at 1:1000). Immunofluorescence images were taken with a Cellomics high-content cell imager (ThermoFisher).

3.2.9 Flow Cytometry Analysis
Flow cytometry data were acquired using a Beckman Coulter CyAn™ ADP analyzer. All flow cytometry data were analyzed with Summit v.4.3 (Dako Colorado).

3.2.10 Immunoprecipitation (IP)
Cells were washed with PBS/- (Life Technologies), treated with NP-40 lysis buffer (200µl/well of a 6-well plate), collected by scraping from the plate and incubated at 4°C for 30 minutes. The lysate was centrifuged for 12000 rpm for 5 minutes and the supernatant was collected and quantified for protein concentration. IP was performed using ANTI-FLAG® M2 Magnetic Beads
(SIGMA) as per the product manufacturer’s guide [131], using the 0.1M glycine HCl, pH 3.0 elution method. NP-40 lysis buffer was composed of: 1X complete Mini EDTA-free Protease Inhibitor Cocktail Tablets (Roche), 1mM phenylmethylsulfonyl fluoride, 10% NP40, 5M NaCl, 1M Tris, in PBS/- and adjusted to pH 8.0.

3.2.11 Immunoblotting

Cells were lysed for whole cell protein isolation as described in 3.2.10 Immunoprecipitation. Prior to loading, 40µg of protein was resuspended and boiled in loading buffer (1x LDS Buffer - Life Technologies, 1x Reducing Agent – Life Technologies, in NP-40 Lysis buffer). Samples were run on a Novex 4-12% Bis-Tris Gel (Life Technologies) in MOPS running Buffer (Life Technologies). The blot was incubated with primary antibodies αOCT4 (BD – 611203, used at 1:500) or αFLAG (Sigma – F1804, used at 1:5000) overnight at 4°C followed by secondary antibody probing with 680 rabbit αmouse IgG (Life Technologies – A21065, used at 1:1000) for 1 hour at room temperature. All blots were imaged on a ODDYSEY® CLx Infrared Imaging System (Li-COR).

3.3 Results

3.3.1 R4 – Integrase Vector Systems

3.3.1.1 Testing of R4 Overexpression construct in HEK 293 R4 cells

Transfection of H9 R4 hESCs with the overexpression constructs failed to generate stable and selectable clones; as such, I sought to systematically decipher the source of this shortcoming. Apart from the lack of amenability of hESCs to electroporation and the single-site genomic integration nature of the R4 platform, at least three factors were hypothesized to account for the lack of transfection. First, if mutations exist in the R4-Integrase coding sequence in the helper plasmid, these would render the protein unable to mediate genomic integration of the
overexpressor plasmid. Second, mutations in the R4-\textit{attB} recombination site of the overexpressor plasmid’s (pER4B-DEST, Figure A.2) backbone would prevent integration even in the presence of a functional Integrase. Third, mutations in the cell’s genomic \textit{attP} integration site would similarly hinder integration of the overexpressor plasmid. The first factor was ruled out by sequencing the R4-Integrase coding sequence, which was found to contain no mutations (data not shown). The second factor was ruled out by transfecting HEK 293 R4 cells with a fluorescence reporter plasmid constructed using the same backbone as that used in the unsuccessful H9 R4 retargeting. Like H9 R4 hESCs, the HEK 293 R4 cells contain a single genomic integration site that allows recombination with the R4-\textit{attB}-containing vectors. HEK cells were chosen for this test given their higher transfection efficiencies and ease with which they can be cultured. The chosen selection regime of 5µg/ml Blasticidin for nine days yielded a 52% crimson positive population as measured by flow cytometry (Figures. 3.1 and B.5). In contrast, the cells transfected without R4-Integrase were fluorescent prior to selection but died entirely by the time of harvest (Figure B.1). An untransfected/unselected population was included as a negative fluorescence control for setting the flow cytometer gates and yielded negligible levels of fluorescence (4%). Mutations in the H9 R4 genomic \textit{attP} site have yet to be ruled out and thus stand as a potential explanation for the lack of stable clones. Alternatively, the inherently low hESC transfection propensity and the low integration efficiencies expected for a single-site integration platform most likely account for the failure to generate clones in this particular cell line.
Figure 3.1. R4-Integrase retargeting of HEK 293 cells generates selectable reporter overexpression clones. Phase contrast (left) and fluorescence (right) images of HEK 293 R4 cells, transfected with the R4 overexpression construct coding for E2-crimson fluorescent protein and selected with Blasticidin (5µg/ml) for nine days. Percentage values of Crimson-positive are as measured by flow cytometry.

3.3.2 PiggyBac Vector Systems

3.3.2.1 Inducibility of piggyBac Vector System

To test the two primary parameters modulating the piggyBac overexpression system’s functional activity, i.e. doxycycline concentration and TetR-KRAB repressor amounts, HEK 293T cells were co-transfected with fixed amounts (0.27µg/well) of LRi1-Crim-Puro fluorescence reporter
plasmid (Figure A.4) along with varying amounts (0.81 – 2.7µg/well) of TetR-KRAB repressor plasmid (Figure A.5). Concurrently, reporter expression in each transfected population was induced with doxycycline concentrations ranging from 0 to 10µg/ml. HEK 293T cells were chosen given their high amenability to transfection and ease of culturing. As expected, increasing the doxycycline concentration in this Tet-On system from 0 to 10µg/ml resulted in substantial increases in reporter expression in all transfection groups, ranging from a 4-fold fluorescence increase in the 1:3 reporter to repressor group to an 11-fold increase in the 1:10 group (Figure 3.2). See Figure B.8 for representative primary flow cytometry data. Inversely, increasing the transfected amount of the TetR-KRAB repressor from 1:3 to 1:10 caused a decrease in reporter expression at all doses of doxycycline ranging from an 81% decrease at 10µg/ml doxycycline group to a 96% decrease in the 0.1µg/ml doxycycline group. Interestingly, a 1:3 transgene to repressor ratio allows the greatest range of absolute fluorescence with the doxycycline concentrations tested and would be the primary choice in future inducible assays using the HEK 293T cell line. The piggyBac overexpression system allows for doxycycline-based inducibility as well as a method of reducing residual expression by increasing the ratio of TetR-KRAB repressor to transgene-of-interest transfected.
Figure 3.2. Inducibility in the overexpression system is dependent on doxycycline dose and the Crimson-Puromycin to TetR-KRAB-Zeocin repressor plasmid ratio. HEK 293T cells co-transfected with fluorescence reporter (Crim-P) and expression repressor (TetR-KRAB) at varying ratios of DNA mass as shown in legend. ‘Blank’ represents untransfected cells. All samples except ‘Blank’ were selected with Zeocin (0.5mg/ml) for 12 days post transfection and fluorescence measured by flow cytometry. Data are shown as mean ±SD (n=3) except the ‘1 to 5’ group at 0.5µg/ml Dox, and the ‘Blank’ group at 2µg/ml Dox, where n = 2.

3.3.2.2 High-efficiency chemical transfection of hESCs

A transfection and selection regime was established to stably transfect hESCs with vectors of interest developed in Chapter 2, while minimizing the impact on the cell’s developmental fate (Figure 3.3). Using Stemfect™ transfection reagent with an adapted methodology (outlined in 3.2.6.2 Chemical Stemfect™ Transfection of H9 hESCs), H9 hESCs were transfected with the LRi2-GFP-Zeo plasmid (Figure A.5) to assess the transfection efficiency. In comparison to the
control population of untransfected/non-selected cells, the transfected/selected population showed high percentages (90 - 93%) of stably transfected cells following Zeocin selection. As a marker of hESC pluripotency [132], SSEA-3 expression was measured by immunostaining and found to be highly expressed in the transfected population (86 - 99%) as well as the total population (84 - 99%) (Figure 3.3a). While flow cytometry revealed transfection efficiency and pluripotency at the population level, microscopy revealed variation in intensity of fluorescence between colonies (Figure 3.3b). See Figure B.9 for primary flow cytometry data. The newly developed transfection/selection method allows for the culture and isolation of a highly-transfected hESC population, with minimal impact on the cell’s pluripotency.
Figure 3.3. Chemical transfection and drug selection of hESCs yields a highly transfected and pluripotent cell population. (A) Flow cytometry analysis of SSEA3-immunostained H9 cells stably transfected with the piggyBac-based GFP reporter, LRi2-GFP-Zeo followed by seven days of non-selected culture and six days of Zeocin selection (1µg/ml) compared to an untransfected/non-selected population. For H9 Untransfected, n = 1. For GFP-Zeocin, n=2. (B) LRi2-GFP-Zeo-transfected cells in culture immediately prior to flow cytometry harvest at 10x magnification.
3.3.2.3 FLAG-tag immunoprecipitation of OCT4 overexpressor

To examine OCT4 overexpression and FLAG-tag immunoprecipitation capacity, HEK 293T cells were transfected with the OCT4 overexpression construct (Figure A.6). Whole cell lysate western blots of transiently transfected cells yield a band of the expected 49 kDa size (46 kDa OCT4 + 3 kDa HisFLAG). Immunoprecipitation of the cell lysate using αFLAG-conjugated beads yields the same band. The small upwards shift observed in the IP bands is attributed to the acidic composition of the IP elution buffer (0.1M Glycine HCl, pH 3.0) in comparison to the lysis buffer (described in 3.2.10 Immunoprecipitation) used in whole cell lysate samples. The shift disappears when elution is done with SDS-PAGE sample buffer (not shown). Whole cell lysates from LRi2-GFP-Zeo-transfected cells served as a negative control and yielded no bands. Immunoblotting (Figure 3.4) demonstrates successful OCT4 transgenic overexpression in HEK 293T cells mediated by the LRi1-Oct4-Puro vector while anti-FLAG immunoprecipitation demonstrates the tag’s functionality for purifying the transgenic protein from the cell lysate (Figure A.6).

Figure 3.4. Immunoblotting of whole cell and FLAG-immunoprecipitated lysates from transiently transfected HEK 293T cells. Lanes 1 and 4: LRi1-Oct4-Puro + LRi2-GFP-Zeo with doxycycline (1µg/ml) induction whole cell lysates. Lanes 2 and 5: LRi1-Oct4-Puro immunoprecipitated with αFLAG antibody-conjugated beads. Lanes 3 and 6: LRi2-GFP-Zeo-transfected whole cell lysate. All lysates collected three days post-transfection and blots probed with αOCT4 or αFLAG antibodies as indicated. PageRuler Protein ladder marker sizes denoted on the side.
3.3.2.4 Constitutive OCT4 shRNA Knockdown in hESCs

hESCs transfected with GFP-tagged knockdown vectors targeting OCT4 or scrambled sequences (Figure A.7) were pulse-selected with 10µg/ml or 40µg/ml G418 whereby selection began 3 days post transfection at 80% culture confluency for 24 hours followed by 48 hours of no selection and lastly 4 days of selection. Pulsed selection, instead of continuous selection, was carried out to allow survival recovery of the population given that mass death (>90%) was unexpectedly observed within the first 24h hours of selection. Cells were harvested and immunostained against OCT4 for flow cytometry (Figure 3.5B). Two scrambled shRNA vectors were used whereby ScrOct4 was non-fluorescent and used for setting the gates for flow cytometry analysis, and Scr1864 which was GFP-tagged. As ScrOct4 is not GFP-tagged, the GFP values (2.75% and 2.90%) measured in these samples represent background fluorescence arising from the methodology used to set the flow cytometry fluorescence gate. The two drug treatments yield similar proportions of GFP-positive cells (3-19% in the 10µg/ml population vs. 11-16% in the 40µg/ml population). In contrast, some divergence is seen in the proportion of OCT4-positive cells among GFP-positive (transfected) cells whereby the 40µg/ml samples show stronger knockdown (16-45%) than the 10µg/ml (36-58%). Unexpectedly, the GFP-tagged scrambled shRNA (Scr1864-GFP) appears to have a measurable knockdown effect as seen in the 60% OCT4-positive cells in the 40µg/ml population. This effect appears even more pronounced in the 10µg/ml population, where only 12% of cells are OCT4-positive. Finally, although additional wells were transfected with a GFP-tagged version of ScrOct4, these wells did not survive the experiment. Although the results require repetition and an earlier onset of selection, knockdown activity of the sh1 and sh2 samples remains consistent between the two sample sets (10µg/ml and 40µg/ml) which preliminarily suggests shRNA expression is successful in the piggyBac knockdown system.
Figure 3.5. Drug selection optimization of piggyBac OCT4-knockdown vector transfections in hESCs. H9 cells were stably transfected with piggyBac GFP-tagged vectors encoding constitutively expressing shRNAs targeting OCT4 (Oct4-sh1-GFP, Oct4-sh2-GFP) or scrambled sequences (ScrOct4, Scr1864-GFP) and pulsed with G418 for one day followed by two days of no selection and pulsed again for 4 days. (A) Cell images immediately prior to flow cytometry harvest. (B) Flow cytometry data from cells immunostained for OCT4. n = 1.
Chapter 4

Discussion and Conclusion

4.1 Discussion

Genetic manipulation of stem cells is essential to studying key aspects of their molecular biology. Precise control of gene-of-interest expression is a powerful approach in establishing the function of genes and their regulatory networks. Characterized regulatory networks can in turn be monitored for changes in gene expression in response to exogenous stimuli such as small molecule or drug exposure. Identifying stimuli-responsive networks is of importance when screening molecules aimed at attenuating the disease phenotypes that arise from network deregulation.

To achieve controlled transgene expression, a robust transgenic system is highly beneficial. For this project, I have designed, built and begun to test the functionality of inducible knockdown and overexpression vector systems for future use in transgenic perturbation of stem cell gene expression. The Gateway® recombineering system allows increased flexibility in building multi-cassette constructs when compared to restriction-enzyme based cloning. While conventional cloning is limited to short restriction sites which must be absent from the coding sequence of the gene of interest, the longer 30bp Gateway® attB recombination sites do not appear in human coding sequences and can thus be freely used to clone any gene. Given the modular nature of the vector systems we have generated tens of intermediate vectors which can be readily incorporated, replaced or arranged in order to build additional constructs or to shuttle in new genes of interest in a high-throughput manner. Lastly, the technology readily lends itself to simultaneous cloning of up to four cassettes and could be used to test the effect of multi-gene knockout or overexpressor vectors on hESC fate.
While R4-Integrase retargeting of HEK 293 cells was successful, hESC retargeting failed to generate stable clones. The successful generation of stably transfected R4 HEK 293s is indicative of a functioning overexpressor plasmid R4 recombination site as well as a functioning Integrase-coding helper plasmid (Figure 3.1). Ultimately, techniques such as Southern blotting or Splinkerette PCR [133] would be necessary to confirm the location of the integration site of the transfected plasmid. The relatively low (52%) percentage of fluorescent retargeted cells could be attributed to genomic silencing of the Crimson cassette but not the drug resistance gene, as has been previously reported [66]. More likely, drug selection may have been incomplete at the time of harvest. If this is the case, extending selection time and/or increasing drug concentrations should increase the percentage of fluorescence-positive cells. The lack of survival of the Integrase-free control cells under the chosen selection regime demonstrates the requirement of Integrase for transgene integration. Genomic sequencing of the hESC attP site has not been carried out by us and would be necessary to rule out mutations that are potentially impeding integration. More likely, the R4 platform’s inherently low integration efficiency (Table 1.1) in combination with the low transfection amenability of hESCs ultimately yields impractically low numbers of stable clones. These drawbacks consequently render this platform unsuitable for high-throughput transgenesis. Given the single site integration nature of the Integrase platform as well as the previous characterization of the attP 13q32.3 genomic integration site as transcriptionally active in the H9 R4 hESC line [78], the R4-Integrase platform was originally chosen by our group to minimize differences between clones stemming from genomic integration location. Nonetheless, the drastically low observed transfection efficiencies quickly outweighed any benefits brought by its single-site genomic integration specificity. Although use of this platform has been successfully applied in multiple cell types [134-135], phiC31/R4-Integrase
systems in hESCs have not been adopted in the literature outside of Invitrogen-authored studies [78, 136-138].

An optimal Tet-inducible system maximizes transgene expression upon induction (+Doxycycline) while minimizing residual expression in the uninduced (-Doxycycline) state. To test the system’s range of inducibility, I measured its responsiveness to increasing doses of doxycycline (Figure 3.2). In theory, greater doxycycline doses more thoroughly perturb TetR-KRAB’s interaction with its TetO2 operator binding site and in turn allow EF1α promoter-driven expression of the reporter. This effect was observed in our system. The doxycycline dose-dependent response of the piggyBac inducible system provides the first level of control over transgene overexpression. This quantitative aspect becomes particularly important in cases when hESC fate is dependent on the levels of expression of the transgene; such is the case for OCT4 [139]. In addition to increasing doxycycline concentration, overexpression levels could be maximized either by increasing the absolute amounts of transgene DNA used for transfection or by using transactivator-based inducible systems such as the reverse-tetracycline transactivator (rtTA). Although no side-by-side analyses of rtTA and TetR-KRAB have so far been carried out, it is expected that the rtTA’s VP16 activator domain-driven transcription should achieve higher expression levels compared to a repressor-based system which is solely dependent on the strength of the transgenic promoter.

To minimize residual transgene expression in the uninduced state, increased quantities of TetR-KRAB repressor can be used to more extensively occupy TetO2 DNA sites and therefore reduce leakiness of expression. This effect is also observed in our system irrespective of doxycycline concentration (Figure 3.2). Indeed, higher repressor concentrations reduce basal expression in the system while similarly suppressing maximal levels of expression in the induced state. While the needs of each experiment will dictate repressor and doxycycline concentrations, the current data
suggest that a 1:3 transgene to repressor ratio allows the greatest range of doxycycline-based inducibility and would thus be used as a starting point for experiments in hESCs. Arguments have been previously presented touting activator-based systems as the most effective in minimizing residual expression [140]. Specifically, it is argued that while repressors must outcompete the endogenous transcriptional machinery in binding DNA to successfully block transcription from full promoters, activators are in used in conjunction with minimal promoters which display minimal to no residual expression when uninduced. This notion may at least partially account for the residual activity observed in our repressor-based system even under the 1:10 transgene to repressor ratio. Additionally, while the EF1α promoter employed in our repressor-based system drives among the most robust and stable gene expression in hESCs [111], it may also contribute to the leakiness observed in the uninduced state.

Yet another approach exists to both maximize induced expression and minimize residual expression. The strategy employs the simultaneous use of TetR-KRAB and rtTA (or mutation-optimized rtTA variants such as rtTA2-M2 [141-143]). This approach is particularly effective when combined with minimal promoters which have undergone rational design to eliminate residual expression. So far, only CMV-based minimal promoters such as Pstitial (Clontech) and P_{tet-T6} [144] have undergone such optimizations. Given the well-established silencing of CMV in hESCs [111], optimization of an EF1α minimal promoter would generate the most optimal repressor-activator mixed systems for hESCs. Although my current two-vector system is compatible with the incorporation of a third (rtTA) vector, this addition would incur additional challenges. First, optimization of triple drug selection would be required. Second, a decrease in transfection efficiency would be expected in a three-vector system in comparison to a two-vector system. Thirdly, consideration should be given to reports of rtTA toxicity on mESCs [145-146]. In its current form, our inducible system provides three parameters (doxycycline concentration,
transgene to repressor ratio, total DNA transfected) allowing modulation of transgene expression. While the inducibility data are presented for an overexpression system, it is expected that H1 promoter-driven shRNA expression can be similarly controlled in the knockdown system.

Effective hESC transfection and selection regimes are necessary to generate pure populations of transgenic lines in a high-throughput manner. A highly transfected cell population allows a more accurate detection of changes in cell biology, which may otherwise be masked by neighboring untransfected cells. While hESC transfection methods continue to improve, no single approach has achieved widespread adoption in the literature. We have previously isolated populations of highly transfected hESCs using the Neon™ electroporation system (Figure B.2), yet its labor-intensive and cost-prohibitive nature makes it undesirable for high-throughput projects. Although Stemfect™ transfection reagent is marketed for transfecting attached hESCs in monolayer; in our hands this protocol yielded transfected cells only in cells along the periphery of colonies (Figure B.3). In contrast, our modified protocol to transfect cells in suspension yielded a more abundant and uniform transfection distribution within nascent colonies. The transfected cells also retained high percentages of GFP expression over long-term (25 days) culture under selection, indicating their potential to be expanded in culture (data not shown). The observed differences between colonies in reporter fluorescence intensity (Figure 3.3B) are consistent with the semi-random nature of the piggyBac system’s integration mechanism as well as the variability between cells in the vector’s integration copy number. As a result, a range of expression levels or silencing of the transgene is expected within a transfected population depending on the plasmid’s integration site and/or copy number. Such variation may be theoretically overcome by a clonal selection approach whereby individual colonies derived from single cells are isolated manually or sorted by FACS. This approach is not without challenges related to the inherent sensitivity of hESCs in
culture. We have previously isolated small colonies shortly after single-cell transfection to create clonal sub-populations and found sub-optimal rates of survival after colony picking. Although increasing the number of sub-populations maintained in culture ultimately allowed the isolation of several clones, the approach is labor intensive. A polyclonal approach analyzing the entire transfected population currently appears more technically feasible. While polyclonal analysis relies on average population expression levels, drug selection may be used to set a minimum threshold of levels of expression in the cells that are to be analyzed. For example, higher drug concentrations could be used to exclusively select cells with high levels of transgene expression. FLAG-tag immunoprecipitation and immunoblotting of cells transfected with the OCT4-overexpressor yielded OCT4 bands of the expected size. In addition to confirming successful transfection, immunoblotting analysis will be useful to qualitatively (by band size) and quantitatively (by band intensity) corroborate transgene expression level data gathered by flow cytometry. Furthermore, the FLAG tag provides a standardized system to immunoprecipitate proteins for use in promoter occupancy ChIP-Seq analyses of candidate transcription factors in the quest to decipher pluripotency regulatory networks.

Preliminary OCT4 knockdown results show consistency in relative knockdown activity of the shRNAs tested whereby sh2 results in greater knockdown than sh1 under both G418 concentrations used for selection (Figure 3.5). Nonetheless, the fact that the GFP-positive percentage values (indicative of transfected cells) remained similar under both selection regimes (9-19% GFP-positive, with a 2% outlier in the Oct4-sh1-GFP sample) is indicative of incomplete selection. Supporting this view is the observation that similar GFP-positive values (9-16%) were previously measured when cells were harvested three days post-transfection under no selection (data not shown). Furthermore, the similarity in GFP-positive values between the 10 and 40µg/ml G418 treatments suggests that the drug concentration is unlikely to be the variable
currently requiring optimization. Instead, the timing of onset of drug selection should be optimized in the next transfection round as this variable is more likely to account for the low proportion of transfected (GFP +ve) cells observed. Onset of selection in the current experiment was begun at 80-90% culture confluency and excessive death (>90%) was observed after 24 hours. Such a sharp drop in survival is likely as a result of overconfluence, as similarly sharp decreases have been observed in cultures without any drug selection (not shown). Given these observations, the goal now becomes determining the optimal timing for onset of selection based on the confluency of the cell culture (likely 50-60%) rather than the currently fixed time point (3 days) after transfection. The G418 selection concentrations for the selection of transfected hESC range from 25-100µg/ml [63-66, 147] and are likely to be used for final selection.

Reduced OCT4 levels in the total (transfected and untransfected) population are observed and may be arising as a result of the possibility that knockdown activity in transfected cells may trigger a reduction of OCT4 expression in neighboring untransfected cells. However, this possibility is unlikely given that the effect was not observed in the 10µg/ml population. Although the OCT4 knockdown assay still requires replication, it is apparent that the current selection regime has not yielded a purely transfected population. A number of approaches exist for G418 selection on hESCs: selection at a constant concentration [64-65], initial selection at a low concentration followed by an increase in concentration at a midpoint in the experiment [66, 147], and lastly, pulsed selection [63] such as that applied in the current experiment. The GFP-tagged scrambled shRNA (Scr1864-GFP) appears to show some knockdown activity whereby only 60% of GFP-positive cells show OCT4 expression in the 40µg/ml population and an even smaller (12%) proportion is observed in the 10µg/ml population. Although Scr1864 may possess off-target silencing effects, experiment replication would be necessary to assess this possibility. Nonetheless, OCT4 knockdown effects can still be observed both in the total population and the
transfected GFP-positive sub-population. In the latter, the shRNAs chosen are consistent in the pattern of knockdown which is markedly different from the scramble control under both selection schemes.

4.2 Future work
Immediately following drug resistance optimization of the knockdown system, it will be crucial to correlate knockdown levels of OCT4 with changes in expression of known pluripotency and fate-specific markers. For example, upregulation of Gata6 (endoderm), Sox1 (ectoderm) and Gata2 (mesoderm) would be expected upon knockdown of OCT4; such an observation would demonstrate a biologically functional impact of the vector system and further validate its utility. In my repressor-based system, maximal levels of induced-state transgene expression are to a large extent dependent on the EF1α promoter’s efficacy in recruiting transcriptional machinery. As the availability of transcriptional machinery is expected to vary between cell lines of different developmental and proliferative states, it will be necessary to establish whether the levels of inducibility achieved in HEK 293T cells can be translated to hESCs. Indeed, absolute levels of transgene expression will be of importance when assessing the effects of transcription factor overexpression/knockdown on hESC developmental status. Furthermore, time-course assays measuring changes in reporter expression upon induction will establish the kinetics of the inducible system and in particular the necessary time to reach maximum induction. If desired expression levels cannot be reached in hESCs, consideration should be given to the incorporation of the rtTA transactivator to augment maximum levels of expression. Ultimately, combinations of doxycycline, transcription factor, TetR-KRAB and rtTA amounts must be established and used according to whether high or low levels of transgene expression are desired.
Lastly, genomic sequencing of the H9 R4 line’s *attP* site should reveal or rule out the presence of mutations in this site.

### 4.3 Conclusion

I have generated a piggyBac transposon-based gene delivery vector system with the capacity to modulate transgene overexpression and knockdown activity in hESCs. I have developed a new cost-effective chemical transfection protocol based on Stemfect® reagent to isolate pure populations of transfected hESCs. Functional overexpression of transgenes is demonstrated by successful immunoprecipitation of OCT4-FLAG fusion proteins constitutively expressed by the piggyBac overexpressor system. Furthermore, the overexpression is responsive to doxycycline inducibility mediated by the TetR-KRAB repressor. While I have preliminarily shown piggyBac-based knockdown of OCT4 levels in transfected hESCs, the drug selection scheme for this system requires further optimization. The R4-Integrase platform did not generate stable hESC clones with any of the vectors or electroporation and selection conditions attempted. In contrast, stably transfected R4 HEK 293 cells were successfully generated. The R4 platform’s dependence on single-site genomic integration in combination with the inherently low amenability of hESC to transfection may explain the inability to obtain stably transfected cells. In addition, the possibility of a mutation preventing recombination at the genomic R4 *attP* site of the hESC line has not been ruled out. Finally, it is worth highlighting that the modular nature of the vector cloning schemes in combination with the Gateway® cloning technology allows for flexible adaptation of the systems. For example, the addition of a transactivator (rtTA) vector to the current piggyBac systems stands as a potential option to further maximize transgene overexpression and knockdown levels. Further improvements in vector composition and
transfection methods will enhance exogenous control of gene levels in hESCs and in turn aid in deciphering the genetic pathways regulating development in these cells.
References


Appendix A. Supplementary Data for Chapter 2

Figure A.1. LR1 Crimson plasmid map.

Figure A.2. pER4B-DEST plasmid map.
Figure A.3. PB-MCSIIcHS4x2-DEST plasmid map.
Figure A.4. LRi1-Crimson-Puro plasmid map.

Figure A.5. LRi2-GFP-Zeo plasmid map.

Figure A.6. LRi1-Oct4-Puro plasmid map.
Figure A.7. LRi2-KD-GFP-Neo plasmid map.

Figure A.8. Genetic elements used in cloning schemes in figures 2.2-2.6.
Appendix B. Supplementary Data for Chapter 3

Figure B.1. Phase contrast (left) and fluorescence (right) images of HEK 293 R4 cells 24 hours after transfection with the R4 overexpression construct (Figure A.1) coding for E2-crimson with or without Integrase.
Figure B.2. Neon™ electroporation of hESCs yields a highly transfected and pluripotent cell population. Flow cytometry analysis of SSEA3-immunostained H9 cells stably transfected with a piggyBac-based GFP reporter, LRi2-GFP-Zeo, followed by five days of non-selected culture and six days of Zeocin selection (1µg/ml) compared to an untransfected/non-selected population.
Figure B.3. Stemfect™ transfection of attached and suspended cells yields peripheral versus uniform distribution of transfection, respectively. H9 hESCs 24 hours post transfection with a piggyBac-based GFP reporter LRi2-GFP-Zeo using standard (top) or modified (bottom) transfection protocols.
B.1 Validation of OCT4 Antibody

Prior to use in knockdown or overexpression assays, αOCT4 antibody specificity was validated by immunofluorescence of monolayer hESC and HEK 293T cultures (Figure B.4) or whole cell lysate immunoblotting (Figure B.5). Strong endogenous OCT4 expression is expected in normally cultured H9s [148].

B.1.1 Immunofluorescence Antibody Validation

αOCT4 immunostaining of hESCs (Figure B.4A-B) reveals a staining pattern predominantly localized to the cell nucleus. Conversely, HEK 293T cells (Figures. B.4E-F) do not show positive staining for either protein as is expected for non-pluripotent cells [104]. Secondary-only antibody staining of hESCs (Figure B.4C-D) and HEK 293T (Figure B.4G-H) cells are negative staining controls illustrating background fluorescence and non-specific binding.

---

Figure B.4. OCT4 antibody shows specificity by nuclear colocalized staining in hESCs and absence of positive signal in HEK 293T cells. OCT4 immunofluorescence staining of H9 hESCs and HEK 293T cells to validate specificity of the antibodies subsequently used in flow cytometry knockdown assays. H9 hESCs (A-B) and HEK 293T cells (E-F) were co-stained with Hoechst, anti-OCT4 and secondary 488nm antibody. Negative controls of secondary antibody-only staining of H9 (C-D) and HEK 293T (G-H) cells are also shown.
B.1.2 Immunoblotting Antibody Validation

Apart from cellular localization, αOCT4 antibody specificity was verified by observing the expected band size (46kDa) [149] in a Western blot assay. As a well-known marker of pluripotency, OCT4 expression was detected in H9 hESC whole cell lysate but not in HEK 293T lysate.

**Figure B.5. Immunoblotting assay validates OCT4 antibody specificity.** Mouse IgG1 monoclonal α-Oct4 (BD) detects the expected endogenous expression of OCT4 in H9 whole cell lysate while detecting no expression in HEK 293T whole cell lysate. PageRuler Protein ladder marker sizes denoted on the side.
Figure B.6. Mean values of chemical transfection and drug selection of hESCs. Values shown for LRi2-GFP-Zeo correspond to the mean ±SD of Figure 3.3A. Flow cytometry analysis of SSEA3-immunostained H9 cells stably transfected with the piggyBac-based GFP reporter, LRi2-GFP-Zeo followed by seven days of non-selected culture and six days of Zeocin selection (1µg/ml) compared to an untransfected/non-selected population. For H9 Untransfected, n = 1. For GFP-Zeocin, n=2 where data are shown as mean ±SD.
Figure B.7. Flow cytometry data for R4-Integrase retargeting of HEK 293 cells. Graphs correspond to samples of Figure 3.1 images. Cells were transfected with the R4 overexpression construct coding for E2-crimson fluorescent protein and selected with Blasticidin (5µg/ml) for nine days.
Figure B.8. Flow cytometry graphs for overexpression inducibility assay. Representative data from each 'reporter to repressor' group. Data are shown in the uninduced (0µg/ml Dox) and maximally induced (10µg/ml Dox) conditions whereby the x-axis measures the intensity of Crimson fluorescence and the y-axis shows the frequency of a cell with a particular fluorescence.
Figure B.9. Flow cytometry graphs for Stemfect™ chemical transfection assay. Data are shown whereby y-axis-FITC represents GFP and x-axis-APC represents SSEA-3.

<table>
<thead>
<tr>
<th>Name</th>
<th>Target Sequence</th>
<th>Sequence Source</th>
</tr>
</thead>
<tbody>
<tr>
<td>Oct4 sh1</td>
<td>AGCGATCAAGCAGCGACTATG</td>
<td>Broad Institute, Clone TRCN0000235523</td>
</tr>
<tr>
<td>Oct4 sh2</td>
<td>GAGGATCACCCTGGGATATAC</td>
<td>Broad Institute, Clone TRCN0000235523</td>
</tr>
<tr>
<td>Scramble 1864</td>
<td>CCTAAGGTTAAGTCGCCCTCG</td>
<td>David Sabatini, Addgene Plasmid 1864</td>
</tr>
<tr>
<td>Scramble Oct4</td>
<td>GGCTGGAATTACGCGACAACT</td>
<td>InvivoGen shRNA Wizard, Scrambled Oct4 sequence</td>
</tr>
</tbody>
</table>

Table B.1. shRNA target sequences cloned into vector BP1-KD-GFP for knockdown experiments.