TCF7 ISOFORMS IN ESC SELF-RENEWAL AND DIFFERENTIATION

ELUCIDATING THE ROLE OF *TCF7* ISOFORMS IN MOUSE EMBRYONIC STEM CELL SELF-RENEWAL AND DIFFERENTIATION

By SUJEIVAN MAHENDRAM, B.Sc.

A Thesis Submitted to the School of Graduate Studies in Partial Fulfillment of the Requirements for the Degree Master of Science

McMaster University MASTER OF SCIENCE (2013) Hamilton, Ontario

(Biochemistry & Biomedical Sciences)

TITLE: Elucidating the Role of *Tcf7* Isoforms in Mouse Embryonic Stem Cell Self-Renewal and Differentiation

AUTHOR: Sujeivan Mahendram, B.Sc. (Hon., McMaster University)

SUPERVISOR: Bradley W. Doble, Ph.D.

NUMBER OF PAGES: *xiii*, 76

ABSTRACT

Recent advances in gene targeting technology have significantly shaped modern-day mouse genetics, as they allow for the accurate analysis of gene function *in vivo*. By capitalizing on conventional methodologies that are based on homologous recombination, the advent of artificially engineered nucleases, like transcription activator-like effector nucleases (TALENs), enables precise genome editing without the need for conventional targeting vectors, which typically possess long "arms" of homology that are difficult to work with, even with recombineering strategies employing bacterial artificial chromosomes. Unlike traditional techniques, these novel nucleases can be engineered in less than a week and together with compact targeting vectors, can be used to easily manipulate almost any locus in the mouse genome.

The current selection of commercially available antibodies makes it difficult to assess the specific roles of protein isoforms during early development. The Tcf/Lef family of transcription factors comprise of key downstream effector proteins of the canonical Wnt/β-catenin signal transduction cascade. This pathway is implicated in the regulation of self-renewal and is dysregulated in a number of human diseases including cancers. Among the Tcf/Lef factors, Tcf3 has been heavily studied in mouse embryonic stem cells, due at least in part to the observation that its transcript levels are expressed at the highest levels compared to the others. Recently, it was proposed that a switch takes place between a repressive state mediated by Tcf3 to an activating β-catenin-Tcf1 complex in response to Wnt signals. Here, we use TALEN technology to introduce an epitope tag at the endogenous locus of *mTcf7*, the gene encoding the Tcf1 protein. By tagging the N-

terminus of full-length and N-terminally truncated dominant-negative variants of Tcf1, we establish a tool to better study a previously unappreciated role for Tcf1 in regulating embryonic stem cell self-renewal and differentiation. Furthermore, we also show that the tagged variants generated exhibit similar protein expression levels to those of wild-type controls, and display nuclear localization as expected.

ACKNOWLEDGEMENTS

First and foremost, I would like to express my deepest and sincere gratitude to my supervisor Dr. Bradley Doble without whom this opportunity would not have been possible. Thank you for your continued guidance and constructive feedback, but more importantly for providing me with a challenging yet rewarding project that helped develop my scientific ability.

A special thanks to my fellow lab members, both past and present: thank you for your encouragement, experimental contributions, and for having faith in my capabilities. I especially wish to thank Steven Moreira for working alongside me and helping to overcome obstacles through combined efforts; to Enio Polena for your technical assistance, and to Victor Gordon for assisting me with last-minute experiments. Thanks to my former lab members: Dr. Kevin Kelly for being my mentor and providing me with initial training, and to Deborah Ng for welcoming me with open arms.

Additionally, I would like to thank my committee members, Dr. Kristin Hope and Dr. Darren Bridgewater for taking precious time from your busy schedules in order to provide valuable input during scheduled meetings. Your advice, perspectives, and expertise have helped complete this project in a timely manner, as I could not have asked for a more knowledgeable duo.

I also wish to extend my thanks to all the members of the Stem Cell and Cancer Research Institute (SCC-RI). Thank you to all those with whom I shared meaningful conversations, whether it involved discussing life stories, experimental procedures, or sharing hours of

laughter. I especially want to thank Sonam Bhatia and Vickie Kwan for editing my reports and keeping me sane during my grad school experience. Thanks also to Andrew Allen for your technical expertise and for facilitating smooth operations at the SCC-RI. I am fortunate for having established friendships with you all, though some of which will undoubtedly last a lifetime.

A heartfelt thank you goes toward to my beloved parents for instilling in me the qualities that have made me the son I am today. No words can truly express how grateful I am for your patience, support, and understanding. Thanks also to the rest of my family and friends for believing in me, and for your genuine curiosity towards my work.

This thesis and the results presented herein could not have been attainable without the funding support received from the Natural Sciences and Engineering Research Council of Canada (NSERC) scholarship, and from the Department of Biochemistry and Biomedical Sciences at McMaster University.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

APC	Adenomatous Polyposis Coli
ARCUT	Artificial Restriction DNA Cutter
BAC	Bacterial Artificial Chromosome
BIO	6-bromoindirubin-3'-oxime
BMP	Bone Morphogenetic Protein
Cas	CRISPR-associated
ChIP	Chromatin Immunoprecipitation
CK1	Casein Kinase 1
CMV	Cytomegalovirus
Co-IP	Co-immunoprecipitation
CRD	Context-dependent Regulatory Domain
CRISPR	Clustered Regulatory Interspaced Short Palindromic Repeats
crRNA	CRISPR RNA
DKO	Double Knockout
DN	Dominant-Negative
DSB	Double-Strand Break
D-TA	Diptheria toxin-A
Dvl	Dishevelled
E-cadherin	Epithelial cadherin
E. coli	*Escherichia coli*
EB	Embryoid Body
ERK	Extracellular signal-Regulated Kinase
ESC	Embryonic Stem Cell
FISH	Fluorescence I*n Situ* Hybridization
FL	Full-Length
FRET	Fluorescence Resonance Energy Transfer
Fz	Frizzled
G. fascicularis	*Galaxea fascicularis*
GEM	Genetically Engineered Mouse
GFP	Green Fluorescent Protein
GSK-3	Glycogen Synthase Kinase-3
HA	Homology Arm
hESC	Human Embryonic Stem Cell

HIPK2	Homeodomain-Interacting Protein Kinase 2
HMG	High mobility group
HR	Homologous Recombination
Id	Inhibitor of Differentiation
iMEF	Inactivated Mouse Embryonic Fibroblast
Int-1	Integration 1
iPSC	Induced Pluripotent Stem Cell
LIF	Leukemia Inhibitory Factor
LRP5/6	Lipoprotein Receptor-related Protein 5/6
mAG	Monomeric Azami-Green
MAPK	Mitogen-Activated Protein Kinase
MEF	Mouse Embryonic Fibroblast
mESC	Mouse Embryonic Stem Cell
NES	Nuclear Export Signal
NHEJ	Non-Homologous End Joining
NLS	Nuclear Localization Signal
NPC	Nuclear Pore Complex
PAC	Phage Artificial Chromosome
pcPNA	Pseudo-complementary Peptide Nucleic Acid
PGK	Phosphoglycerate Kinase-1
qRT-PCR	Quantitative Real-time Reverse Transcription PCR
RIPA	Radio-immunoprecipitation Assay
RVD	Repeat-Variable Di-Residue
SDN	Site-Directed Nuclease
shRNA	Short Hairpin RNA
TALEN	Transcription Activator-Like Effector Nuclease
Tcf/Lef	T-Cell Factor/Lymphoid Enhancer Factor
TGFβ	Transforming Growth Factor β
TLE	Transducin-like Enhancer of Split
tracrRNA	Transactivating CRISPR RNA
Wg	Wingless
WRE	Wnt Responsive Element
YAC	Yeast Artificial Chromosome
ZFN	Zinc-Finger Nuclease

DECLARATION OF ACADEMIC ACHIEVEMENT

Sujeivan Mahendram created the endogenously epitope-tagged FL- and DN-Tcf1 cell lines, and carried out the vast majority of experiments and analyses in this thesis.

Bradley Doble provided funding support, as well as helped conceive and design experiments.

Victor Gordon assisted with a few western blot experiments.

CHAPTER 1
INTRODUCTION

1. Embryonic Stem Cells

The defining characteristics of embryonic stem cells (ESCs) include the ability to self-renew and to be *pluripotent*. Pluripotency enables these cells to differentiate into cell types of any of the three primary germ layers (i.e. mesoderm, endoderm, and ectoderm)[1,2]. In addition, ESCs can be maintained and propagated *in vitro* under defined conditions. Thus, the unique properties of ESCs give them the potential to contribute to regenerative therapies, while at the same time provide a suitable model system for studying development and cancer-related processes.

In terms of their regulation, the ability of ESCs to self-renew relies almost exclusively on a core transcriptional network, which consists of the transcription factors Oct4, Nanog, and Sox2[3,4]. These three factors establish an autoregulatory circuitry within pluripotent cells, with each factor having the capacity to induce the expression of the others by binding their promoters[5].

1.1 Mouse embryonic stem cells

1.1.1 Derivation

In the early 1980s, two independent groups derived the first mouse embryonic stem cell (mESC) lines from the inner cell mass of an early developmental stage known as the blastocyst[1,2]. Mouse blastocysts were grown on mitotically inactivated mouse embryonic fibroblasts (MEFs), called “feeders,” as this layer was shown to be essential for the isolation of mESCs. However, it was later established that the feeder layer was not

necessary for the prolonged maintenance of mESCs in culture, and that the soluble cytokine leukemia inhibitory factor (LIF) in addition to serum was sufficient to sustain these cells in their undifferentiated state[6,7]. The removal of LIF, combined with the growth of mESCs in suspension, results in the formation of embryoid bodies (EBs) and differentiation. EBs are spherical ESC-derived cellular aggregates, which when allowed to differentiate over time, resemble disorganized early mouse embryos possessing a variety of cells from each of the three primary germ layers[8,9].

1.1.2 Assessing validity and function

There are currently a number of assays that have been developed to assess the pluripotency and differentiation potential of mESCs. The predominant *in vitro* assay for testing differentiation is EB generation, as the expression of genes specific to each germ layer can be assayed by western blot or qRT-PCR analyses[9]. Alternatively, a more recent approach involves exposing mESCs to defined culture conditions and growth factors in order to induce differentiation into a particular cell type[10].

However, the gold standard test for measuring pluripotency lies not in *in vitro*, but *in vivo* assays. If truly pluripotent, once they are introduced into the developing blastocyst, mESCs can generate germ line transmissible chimeric organisms that contains germ cells derived entirely from the ESCs[11]. Thus, when chimeras are bred with a test mouse, the resulting offspring will be a hybrid. Alternatively, chimeras can also be achieved by tetraploid complementation using mESCs[12]. Here, a host embryo is isolated at the 2-cell stage and both cells are fused together so that subsequent daughter cells are tetraploid. Tetraploid embryos are incapable of generating an embryo proper on their own, but

mESCs introduced into tetraploid blastocysts will "complement" the tetraploid cells such that the mESCs will develop into the embryo proper and the tetraploid cells will give rise to most of the extraembryonic tissues. Pluripotency can also be demonstrated by the ability of a given cell line to give rise to endodermal, mesodermal, and ectodermal lineages through the formation of teratomas[13]. Teratomas are benign tumours comprised of cells from all three germ layers, which develop following injection of pluripotent cells into syngeneic or immunodeficient mice.

1.2 Signaling cascades regulating mESCs

1.2.1 LIF/Stat3

Leukemia inhibitory factor (LIF) is a soluble factor produced by feeder cells and is required to support the undifferentiated growth of mESCs[6,7]. This cytokine signals through the dimerization of cytokine receptors gp130 and LIF-R, which activate the Jak family of tyrosine kinases to phosphorylate both receptors[14]. This in turn signals the transcription factor Stat3 to be recruited to the cytokine receptor complex at which point, Jak phosphorylates it, and ultimately leads to its dimerization, nuclear translocation, and transcriptional activation of target genes[15-17]. While in the absence of LIF, Stat3 overexpression promotes self-renewal and pluripotency, inactivation of Stat3 function in the presence of LIF results in differentiation[17,18]. This suggests that Stat3 is an integral downstream mediator of LIF signaling for the maintenance of pluripotency.

1.2.2 BMP4/Smad

Interestingly, the ability of LIF to maintain mESC self-renewal and pluripotency is effective only when cells are grown in culture media containing serum. In the absence of serum however, LIF alone cannot maintain mESC pluripotency and cells undergo neural differentiation as a result, which is suggestive of another LIF-cooperating factor[19]. This additional factor was determined to be bone morphogenetic protein 4 (BMP4), as mESCs were shown to be able to sustain pluripotency in serum-free culture conditions with both BMP4 and LIF[20]. BMP4 is a member of the transforming growth factor β (TGFβ) family and has been shown to cause a block to neural differentiation[20]. In brief, the canonical BMP signaling pathway involves the phosphorylation of Smad1, 5, or 8 (i.e. R-Smads) by the type I receptor. Two phosphorylated R-Smads then form a complex with Smad4, and together, this complex translocates into the nucleus to induce transcriptional activation of target genes including the inhibitor of differentiation (Id) genes. Moreover, BMPs can also enhance pluripotency by acting as an inhibitor of MAPK/ERK (mitogen-activated protein kinase/extracellular signal-regulated kinase) pathways[21]. Upon withdrawal of LIF from serum-free culture conditions, mESCs expressing Id proteins readily differentiate, but fail to give rise to cell types of neural lineage[20]. This observation suggests that LIF and BMP4 act cooperatively in order to inhibit differentiation for the maintenance of mESC pluripotency.

1.2.3 Canonical Wnt/β-catenin

The canonical Wnt/β-catenin pathway is one of the fundamental regulators of normal embryonic development and tissue homeostasis[22]. It has also been strongly implicated in

the regulation and maintenance of pluripotent ESCs[23-28]. Additionally, mutations in Wnt signaling proteins have been observed in a number of human cancers. One of the first genes to be characterized as a component of the Wnt/β-catenin signaling pathway is Integration 1 (Int-1), which was shown to be activated upon viral integration of mouse mammary tumour virus[29]. The subsequent discovery and characterization of the Int-1 homolog in *Drosophila melanogaster*, Wingless (Wg), was found to control segment polarity in the developing larvae[30]. Thus, the modern name for the Int-1 gene is Wnt1, which stems from the synthesis of Int-1 and Wingless. Since then, many of the components of the signaling pathway were discovered. The specifics of Wnt/β-catenin signal transduction are discussed in greater detail below.

2. Canonical Wnt/β-catenin Signaling Pathway

The canonical Wnt/β-catenin cascade is initiated by the binding of a Wnt ligand, which are typically 40 kDa lipid-modified glycoproteins, to two cell surface receptors: Frizzled (Fz), a seven-pass transmembrane receptor, and its co-receptor low-density lipoprotein receptor-related protein 5 or 6 (LRP5/6). Within the human genome, there are 19 Wnt genes and 10 Fz proteins, the latter of which consists of a large extracellular cysteine-rich domain at the N-terminus to which Wnt ligands bind with great affinity[31]. In the absence of Wnt ligand, the ubiquitous kinase, glycogen synthase kinase-3 (GSK-3), is found in a destruction complex with casein kinase 1 (CK1), adenomatous polyposis coli (APC), Axin, and the effector molecule β-catenin (Figure 1). Initially phosphorylated by CK1, β-catenin is then N-terminally phosphorylated by GSK-3, which marks it for ubiquitin-dependent proteasomal degradation facilitated by the E3 ubiquitin ligase β-TrCP[32]. Wnt-

induced activation of the pathway leads to the stabilization of cytosolic β-catenin. More specifically, the initial binding of Wnt ligands to the heterodimeric receptor complex results in the phosphorylation of the LRP5/6 cytoplasmic tail. Facilitated by the scaffolding protein Dishevelled (Dvl), Axin is then recruited to the membrane where it utilizes the LRP5/6 phosphorylation event as a docking site[33]. This latter step is critical in deactivating the destruction complex, as GSK-3 and CK1 are subsequently tethered to the membrane as well. Deactivation of the destruction complex ultimately promotes the cytosolic accumulation and nuclear translocation of β-catenin, resulting in the transactivation of target genes in concert with the T-Cell Factor/Lymphoid Enhancer Factor (Tcf/Lef) family of transcription factors[34].

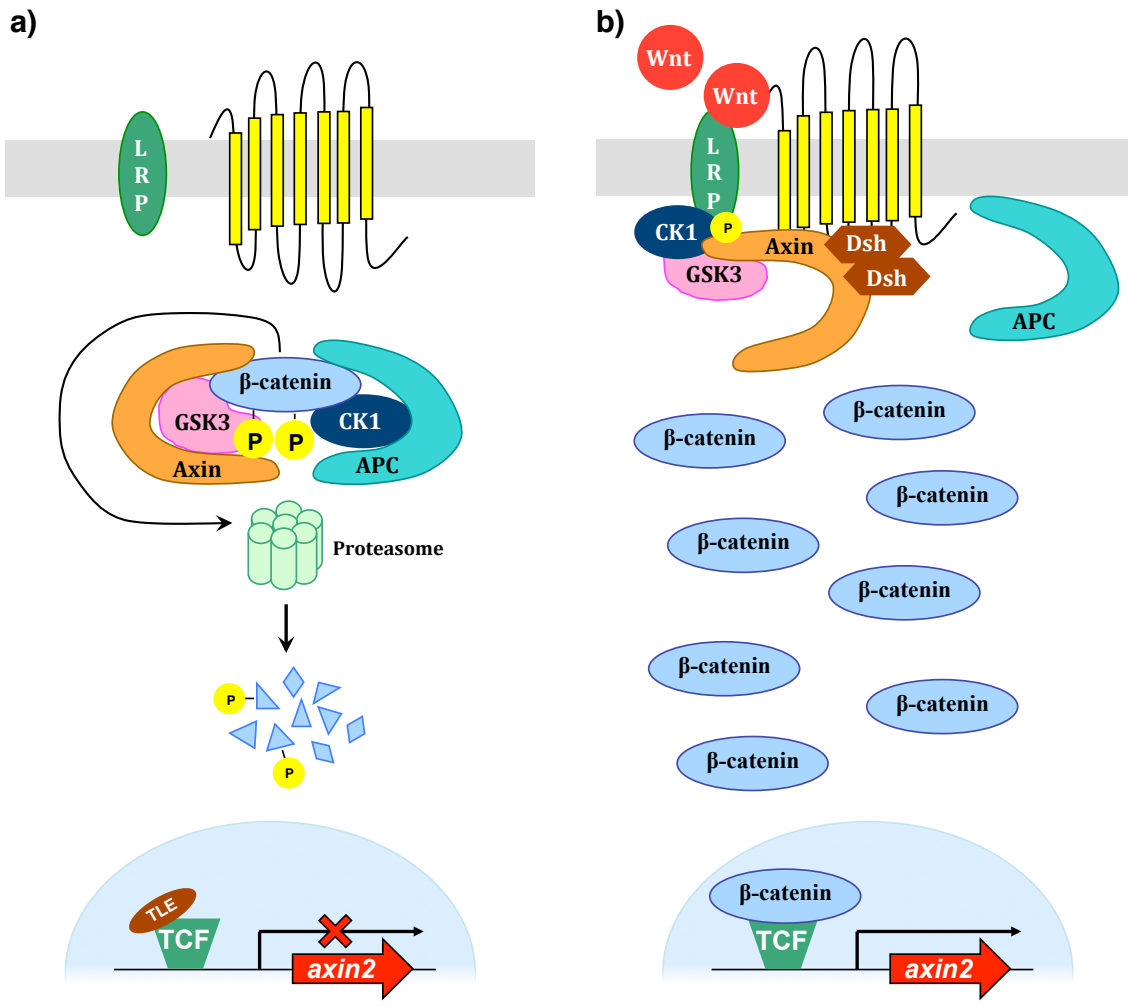

Figure 1: Canonical Wnt/β-catenin Signaling
(a) In the absence of Wnt ligand, β-catenin is found in a destruction complex with GSK-3, CK1, Axin, and APC. Once phosphorylated by GSK-3, β-catenin is ubiquitin-tagged for proteasomal degradation. Tcf/Lefs in the nucleus are bound to co-repressors and repress transcription of Wnt target genes. **(b)** In the presence of a Wnt stimulus, the destruction complex is disassembled, as a phosphorylation event on LRP5/6 creates a docking site for Axin. β-catenin is no longer degraded and its accumulation in the cytoplasm eventually leads to subsequent translocation into the nucleus where it transactivates Wnt target genes in concert with Tcf/Lefs.

2.1 Major components of pathway

2.1.1 Glycogen synthase kinase-3

A key member of the Wnt/β-catenin pathway and an important regulator of stem cell self-renewal is the multifunctional serine/threonine kinase GSK-3. The name GSK-3 is derived from its ability to phosphorylate glycogen synthase, an enzyme involved in glycogen synthesis[35]. Two variants of GSK-3 exist in mammals, namely GSK-3α and GSK-3β, each encoded by a separate gene. While the N- and C-termini of GSK-3α and GSK-3β differ quite substantially, they share almost identical sequence homology in their kinase domains and exhibit functional redundancy in Wnt signaling as observed in mESCs[36]. However, the phenotypes of GSK-3α and GSK-3β knockout mice differ. Although GSK-3α knockout mice are viable with few abnormalities, those that have GSK-3β knocked out undergo several developmental defects and die around birth[37,38].

A unique requirement for the enzymatic activity of GSK-3 is that substrates must be pre-phosphorylated by another kinase, like CK1. In addition, the priming phosphorylation event must occur four residues C-terminal to the GSK-3 target site, as depicted by the consensus sequence S/T-X-X-X-S-P/T-P, where the first serine/threonine (S/P) is the target site, followed by any three amino acids (X), and the last serine/threonine that is pre-phosphorylated by a priming kinase[39]. Several target proteins have multiple tandem target residues, such as β-catenin, which is phosphorylated by GSK-3 on three consecutive residues after a priming phosphorylation event executed by CK1-α.

2.1.2 β-catenin

β-catenin is the key effector molecule and a proto-oncogene of the Wnt pathway, as it is responsible for Wnt-induced signal transduction into the nucleus and the subsequent transcriptional induction of Wnt target genes. Apart from its involvement in signaling, β-catenin along with its homologues α-catenin and γ-catenin (plakoglobin), were identified to exhibit a structural role in cell adhesion by interacting with epithelial cadherin (E-cadherin) molecules[40]. The structure of β-catenin includes a central domain consisting of 12 Armadillo repeats, which facilitate interactions with its binding partners including Axin, APC, and transcription factors of the Tcf/Lef family[41]. The name given to these repeats stems from β-catenin's *Drosophila* orthologue Armadillo, which causes an armadillo-like phenotype of developing larvae, when mutated[42].

Following induction of the Wnt signaling pathway, cytoplasmic levels of β-catenin rise to a certain threshold after which it translocates into the nucleus to engage in Wnt target gene activation. Interestingly, β-catenin neither possesses a traditional nuclear localization signal (NLS) nor a nuclear export signal (NES) and as such, the precise mechanism by which it enters and exits the nucleus is poorly understood. Although its nuclear import and export has been shown to act independently of karyopherin/importin, recent studies have suggested that β-catenin directly interacts with nuclear pore complex (NPC) proteins[43-46]. Additionally, β-catenin also does not contain a DNA binding domain and consequently requires DNA binding partners like Tcf/Lefs for promoter docking[47].

2.1.3 T-Cell Factors and Lymphoid Enhancer Factor

Transcription factors of the Tcf/Lef family are a subset of the high mobility group (HMG) box protein family, with four members identified in vertebrates to date: Tcf1 (*Tcf7*), Tcf3 (*Tcf7l1*), Tcf4 (*Tcf7l2*), and Lef1 (*Lef1*). The names given to the Tcf/Lefs correspond to their discovery as regulators of lymphoid cells, including T-lymphocytes. Tcf1 was first identified in a screen seeking transcription factors that bound to a specific enhancer region of the T-cell gene, CD3-ε[48,49]. Lef1 was characterized following a screen for transcriptional regulators that bound to the T-cell receptor α enhancer[50]. Both Tcf3 and Tcf4 were later identified using hybridization screens with *Tcf7* cDNAs containing the HMG box[51].

Each of the Tcf/Lef family members possesses highly conserved regions, including an N-terminal β-catenin binding motif, a high mobility group (HMG) DNA binding domain, and a nuclear localization signal (NLS) (Figure 2). β-catenin is the most potent activator of Tcf/Lef-regulated transcription and was originally identified as the link between Tcf/Lefs and Wnt signaling through yeast two-hybrid screens[52-54]. Truncation mutants lacking the N-termini of Tcf/Lefs, referred to as dominant-negative mutants, do not bind β-catenin and fail to transcriptionally activate Wnt target genes. The HMG DNA binding domain consists of an HMG box as well as a basic tail comprised of basic amino acids. A notable characteristic of Tcf/Lefs is their ability to bend DNA when bound to their target consensus sequence, CCTTTGAAC in vertebrates, also known as the Wnt responsive element (WRE)[53]. This sequence is precisely located in the minor groove of DNA and when bound by a Tcf/Lef, the helix can bend between 90° and 127°[55]. This feature allows

for the interaction of distant transcriptional complexes. Additionally, the small basic tail functions to enhance Tcf/Lef binding affinity to the negatively charged DNA backbone as well as serving as a potent NLS by interacting with importins[55,56]. Domains of Tcf/Lefs that are dissimilar between family members include the context-dependent regulatory domain (CRD) and the C-terminal 'E' tail, the latter of which is translated from *Tcf*, and not *Lef*, genes[57]. The CRD contains an alternative exon in all vertebrate Tcf/Lefs, but also includes a repressive domain, which serves as a docking site for the repressor Groucho/TLE (transducin-like enhancer of split)[58]. The E tail contains a basic cysteine motif, termed the C-clamp, which provides additional binding specificity to target genes[53,59,60].

Figure 2: General schematic representation of Tcf/Lef gene structure
Despite their homologue-specific roles, Tcf/Lef genes all include regions of high similarity including the β-catenin binding motif, HMG box, and the nuclear localization signal (NLS). Relatively more variable regions between the transcription factors include the context-dependent regulatory domain and an E-tail found in the C-terminus of some Tcfs.

Furthermore, a number of different isoforms of Tcf/Lefs have been identified, and this diversity has been attributed to dual promoter usage and extensive alternative splicing, in addition to variations in the CRD and C-terminal tails (Figure 3)[61,62]. Alternative promoters create naturally occurring isoforms lacking the N-terminal β-catenin binding domain and are observed in all Tcf1 and Lef1, but not Tcf3 or Tcf4. On the other hand, alternative splicing can create variants missing other functional domains like the C-clamp, but all except one Tcf4 isoform retain the HMG DNA binding domain[63]. Full-

length isoforms of Tcf1 and Lef1 are often associated with activation of Wnt target genes, while Tcf3 and sometimes Tcf4 are linked to target gene repression[64-66]. Although the mechanism by which this occurs is not well understood, it is believed that having the amino acid motifs LVPQ and SxxSS renders the Tcf/Lef a transcriptional repressor of Wnt target genes[67,68]. The fact that the SxxSS motif is always present in Tcf3 and alternatively spliced in some Tcf4 isoforms supports the notion that Tcf3 and sometimes Tcf4 act to repress Wnt target genes. The activating and inhibitory roles of the Tcf/Lefs were further assessed in a recent study conducted by Ho and colleagues (2013). The group demonstrated that during the reprogramming of somatic cells into induced pluripotent stem cells (iPSCs), Wnt signaling represses the early stage but stimulates the latter stage of this process. In particular, Tcf3 and Tcf4 enhance early reprogramming events and are inhibited by Wnt signaling via Tcf1 and Lef1. In the later stage however, Tcf3 and Tcf4 are inhibitory, as depleting both genes enhanced iPSC reprogramming[69].

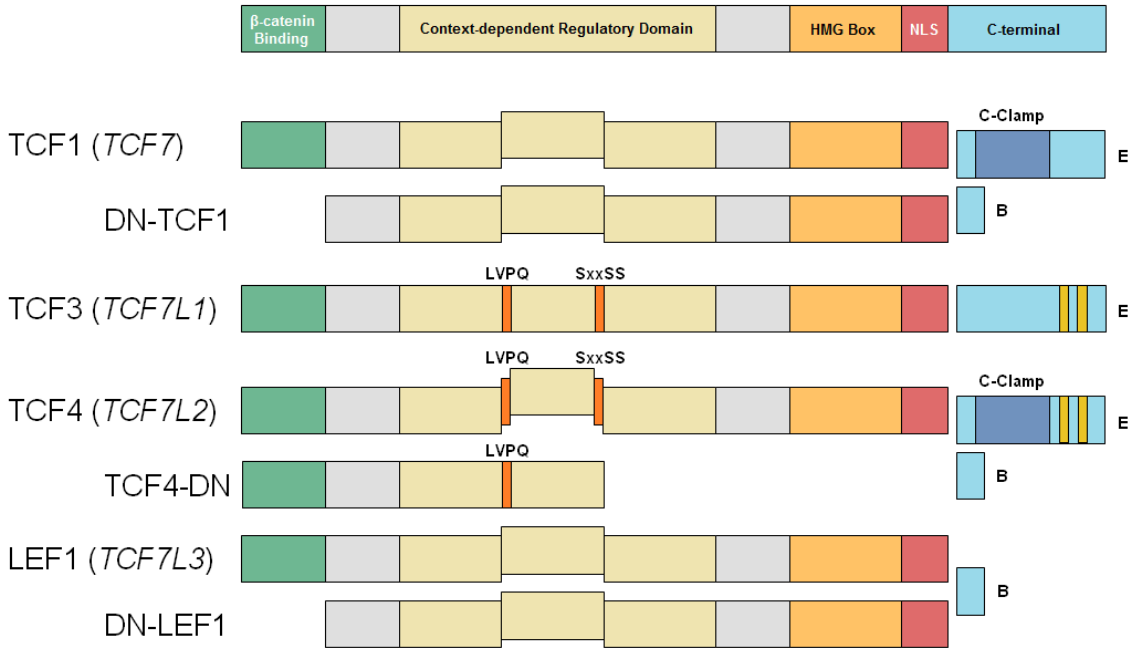

Figure 3: Schematic diagram of different Tcf/Lef isoforms in vertebrates
A number of isoforms exist in the Tcf/Lef family of transcription factors. The different variants are generated through the use of alternative splicing of the mRNA transcript or dual promoter usage, or a combination of both. LVPQ and SxxSS are repressive domains. B and E tails are designated with their corresponding letters. Figure adapted from Arce *et al*., 2006[57].

3. Activation of Wnt/β-catenin Signaling in mESCs

3.1 GSK-3 inhibition/ablation

GSK-3 inhibition by the small molecule inhibitor 6-bromoindirubin-3'-oxime (BIO) was shown to enhance pluripotency and the retention of the relevant markers in both mESCs and hESCs[23,27]. Similarly, treatment of mESCs with LiCl also yielded similar observations. However, these experiments were not carried out for more than 7 days or for multiple passages. The advent of a three inhibitor (i.e. "3i ") cocktail established naïve pluripotency in mESCs maintained in fully defined serum-free medium. Naïve pluripotency refers to ESCs having the ability to give rise to an entirely ESC-derived

mouse via tetraploid aggregation[70]. They can also be expanded clonally without genetic transformation. hESCs however, exist in a relatively lower state of pluripotency known as "primed" pluripotency and require different growth factors for their maintenance[70]. The 3i cocktail consists of two small molecule inhibitors of MAPK/ERK signaling, SU5402 and PD184352, as well as the highly specific GSK-3 inhibitor CHIR99021[71]. It is believed that the MAPK/ERK inhibitors serve to cause a block in FGF4-driven autoinductive signaling required for the differentiation of mESCs. The use of small molecule inhibitors however, does have the potential to cause unintended off-target effects.

Ablation of both GSK-3α and GSK-3β (double knockout; DKO) promotes self-renewal in the presence of LIF whereas in the absence of LIF it skews differentiation towards mesendodermal cell types, while blocking neuronal differentiation of mESCs[36,72,73]. DKO mESCs retain the expression of key markers associated with pluripotency and display extremely high levels of cytoplasmic and nuclear β-catenin[36]. Additionally, DKO cell morphology closely resembles that of mESCs treated with CHIR99021[74].

3.2 Role of β-catenin in mESCs

β-catenin's role in reinforcing pluripotency was demonstrated in a number of studies overexpressing the protein. Ogawa and colleagues (2006) used an inducible system to study the effects of wild-type and overexpressed stabilized β-catenin in mESCs[25]. This group, along with Takao and colleagues (2007), who generated the S33A stabilized β-catenin cell line, observed that stabilization resulted in the retention of pluripotency markers upon LIF withdrawal[75]. Additionally, Kelly and colleagues (2011) demonstrated

that mESCs stably expressing β-cateninS33A exhibited a cell morphology that was indistinguishable from GSK-3 DKO mESCs and exhibited a delayed loss of pluripotency[74]. A proposed mechanism by which stabilized β-catenin reinforces pluripotency is that it enhances the transcriptional activation of Oct-4 target genes, which was demonstrated by the observation that an Oct-responsive reporter and Oct-4 target genes were induced upon stimulation of the Wnt/β-catenin pathway by administering Wnt3a or CHIR99021[74,76].

Studies involving β-catenin knockout mice indicate embryonic lethality by d6.5, as embryos display severe defects in the formation of ectoderm and the anterior-posterior axis[77,78]. Of note, the protein expression levels of the β-catenin homologue γ-catenin (also known as Plakoglobin) have repeatedly been shown to be upregulated in a β-catenin-null background. However, γ-catenin is not further altered upon GSK-3 inhibition with CHIR99021 in cells lacking β-catenin. Furthermore, when β-catenin-deficient mESCs are differentiated as aggregates they fail to undergo mesendodermal and neuronal differentiation, revealing that γ-catenin levels are insufficient to compensate for the lack of β-catenin with regard to its role in mediating these differentiation events[77-80]. Intriguingly, ectopic expression of γ-catenin in β-catenin$^{-/-}$ mESCs was indeed responsive to GSK-3 inhibition, as transcript levels of prototypical Wnt target genes were upregulated as a result[79]. Also, in wild-type mESCs, overexpression of γ-catenin results in a similar phenotype as that observed in GSK-3 DKO mESCs or wild-type mESCs overexpressing stabilized β-catenin. That is the cells are refractory to differentiation and display a complete block in their ability to differentiate into neurectoderm. These

findings suggest that although γ-catenin cannot substitute for β-catenin, overexpression of γ-catenin mimics the effects of stabilized β-catenin on mESC self-renewal and differentiation. Importantly, Wray and colleagues (2011) demonstrated that β-catenin deficiency in mESCs eliminates its responsiveness to GSK-3 inhibition with regard to the maintenance of the pluripotent state[81]. This study underscored the importance of β-catenin in mediating the effects of CHIR99021 in mESCs maintained in the 3i cocktail.

4. Tcf/Lef Signaling in mESCs

The expression of both mRNA transcripts and protein expression of all four factors of the Tcf/Lef family have been detected in mESCs (Figure 4). Transcript levels of Tcf3 have been shown to be expressed at the highest levels, while western blot analysis revealed expression of the four proteins at readily detectable levels[74,82].

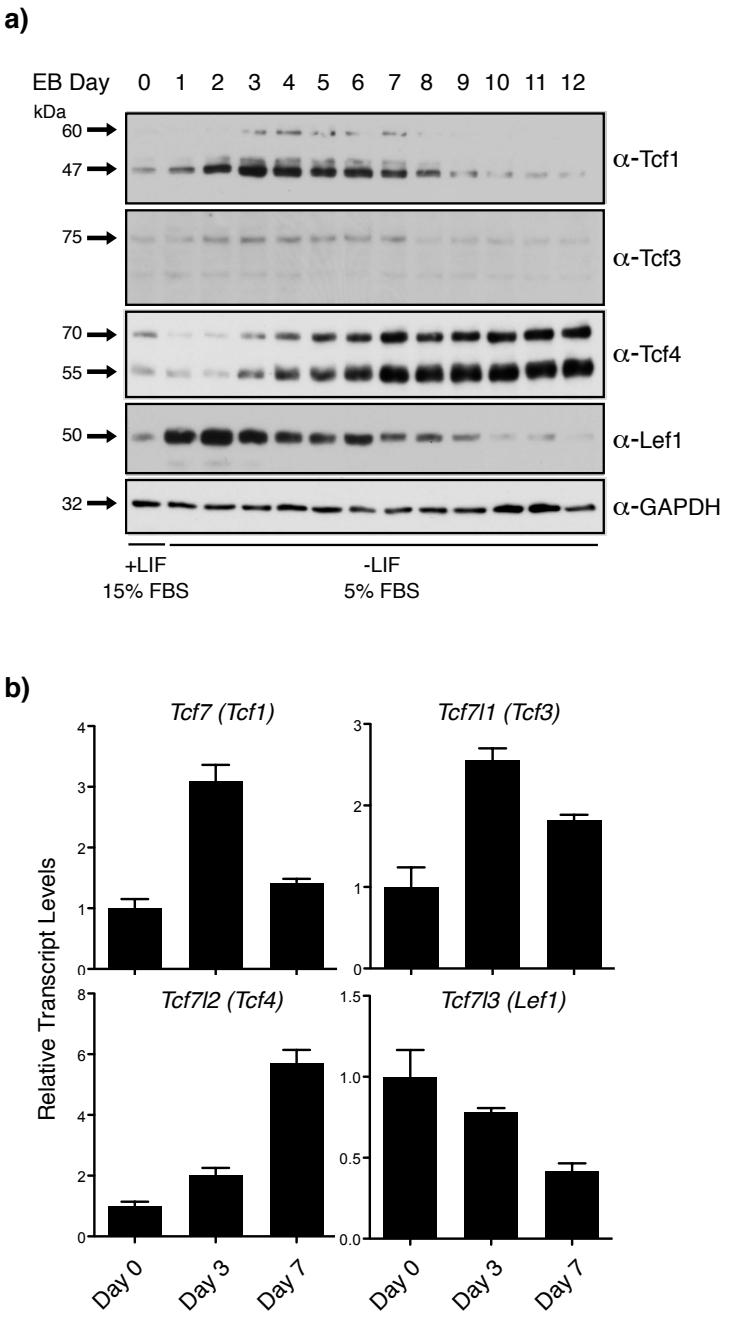

Figure 4: Tcf/Lef protein and mRNA expression in mESCs
a) Wild-type mESCs were cultured for 12 days in EB media (-LIF, 5% fetal bovine serum). Each day, cells were pelleted and whole cell lysates were prepared using RIPA lysis buffer. Protein samples were probed for Tcf1, Tcf3, Tcf4, and Lef1, using GAPDH as a loading control. **b)** mRNA expression levels of all Tcf/Lef factors were assayed at days 0, 3, and 7. Bars represent the mean of biological triplicates and error bars indicate SEM.

4.1 Tcf3-Tcf1 switch model

A recently emerging model of Wnt/β-catenin signaling proposed by Yi and colleagues (2011) by using bioinformatics studies suggests that there is a switch from a Tcf3-induced repressive state, to a Tcf1-induced state of transcriptional activation, in response to Wnt signals in mESCs. This theory is supported in part by the observation that mESC lines expressing Tcf3 lacking the N-terminal β-catenin binding domain display defects in the ability to self-renew in response to Wnt3a stimulation or chemically-induced GSK-3 inhibition[66,81]. Additionally, in *Xenopus*, Tcf3 is phosphorylated by homeodomain-interacting protein kinase 2 (HIPK2) in response to Wnt signals, leading to its dissociation from target genes in a β-catenin-dependent manner[83]. Interestingly, Tcf1 does not become phosphorylated and replaces the Tcf3-bound WRE during Wnt-induced activation of the pathway, suggesting a possible mechanism by which Tcf3 is derepressed in mESCs[84]. Tcf3 is the most abundantly expressed of all Tcf/Lefs in mESCs, at least at the transcript level, making it the factor that has been comparatively studied more in-depth in these cells. The relatively new Tcf3-Tcf1 switch model provides insight into a previously unappreciated role of Tcf1 in Wnt/β-catenin signaling and stem cell self-renewal and raises new questions regarding the mechanisms controlling the proposed Tcf3-Tcf1 switch in the context of mESCs.

4.2 Regulation of mESCs by Tcf1

Tcf1 plays important roles in the control of a number of developmental programs, and is required for early T-cell differentiation and the polarization of T helper 2 (T_H2) cells[85,86]. In addition, it has been shown to regulate ventroposterior development and dorsal

development in *Xenopus* and Zebrafish, respectively[85,87,88]. Unlike Tcf3 and Tcf4, Tcf1 (and Lef1) contains an alternative promoter that generates N-terminal truncated isoforms devoid of the β-catenin binding domain[61,89]. This helps to explain the observation of a number of bands ranging from 25 to 55 kDa in Western blot analyses of Tcf1 protein expression in human T-cell extracts[61]. N-terminal-truncated Tcf protein products are referred to as dominant-negative isoforms (e.g. DN-Tcf1) as they compete for WRE binding sites, and in doing so, prevent Wnt-induced activation of canonical Wnt/β-catenin signaling. In general, the usage of *Tcf7* alternative promoters is linked to changes in cell state, with full-length Tcf1 (FL-Tcf1) isoforms expressed abundantly in proliferating cells and dominant-negative variants expressed highly in differentiating and arrested cells[32,90,91]. These generalizations stem from a number of key observations. First, Tcf1-null ($Tcf1^{-/-}$) mice experience adenoma formation in the intestine and breast[92]. This can be explained by the observation that DN-Tcf1 is the predominantly expressed isoform in normal cells, where as FL-Tcf1 is upregulated in cancer. Interestingly, while DN-Tcf1 is detected in the normal colonic crypt, it is absent in proliferating colon cancer cells[90]. Moreover, genetic ablation of *Tcf7* in colon cancer cells is growth inhibitory[93]. Secondly, DN-Tcf1 is preferentially expressed in resting T cells whereas FL-Tcf1 is predominantly expressed during lymphocyte maturation and naïve $CD8^+$ T cell activation[91].

Two recent studies have also demonstrated that Tcf1-deficient mice develop thymic T-cell lymphoma, which is equivalent to human T-acute lymphoblastic leukemia[94,95]. These latter studies both note a significant upregulation of Lef1 mRNA and point toward a

dependency on Lef1 for the development of T-cell lymphomas. Furthermore, the expression of short hairpin RNAs (shRNA) against Tcf1 mRNA transcripts in mESCs appears to inhibit many of the same genes that are upregulated in $Tcf3^{-/-}$ mESCs[66]. Wnt3a-stimulated $Tcf3^{-/-}$ mESCs also display a defect in responsiveness following Tcf1 shRNA expression[66]. Such findings are suggestive of a tumour suppressive role of DN-Tcf1, and further support the Tcf3-Tcf1 switch model of Wnt/β-catenin signaling in mESCs. Nonetheless, it appears as though the system requires a balance of Tcf3 and Tcf1 activities in order to sustain prolonged self-renewal in mESCs.

5. Genome Editing Techniques

The holy grail of genetics can arguably be deemed as being able to fully characterize a given gene and determine its significance, if any, toward human inherited diseases. Recent advances in gene targeting technology have significantly shaped modern-day mouse genetics, as they allow for the accurate analysis of gene function *in vivo*.

5.1 Genetically engineered mouse models

The mouse has been used for several decades as a model organism to mimic human diseases. In addition to the relatively high similarity between the mouse and human genomes, the development of gene targeting technologies has allowed for the functional analysis of genes in the mouse that have been, until recently, impossible to carry out in other mammals[96,97]. With the completion of the human genome project in 2003, the demands for the development of genetically engineered mouse (GEM) models as a means to study functional genomics have become increasingly significant. Not only have GEM

models proven to be an invaluable tool to aid in the understanding of gene function *in vivo*, but they have also been capable of faithfully mimicking human diseases. The feasibility of performing such genetic alterations has rapidly evolved over the past three decades as genome-editing technologies have come to be more precise and less technically challenging. Of note, the number of gene targeting papers published each year has risen exponentially over the past 20 years (Figure 5).

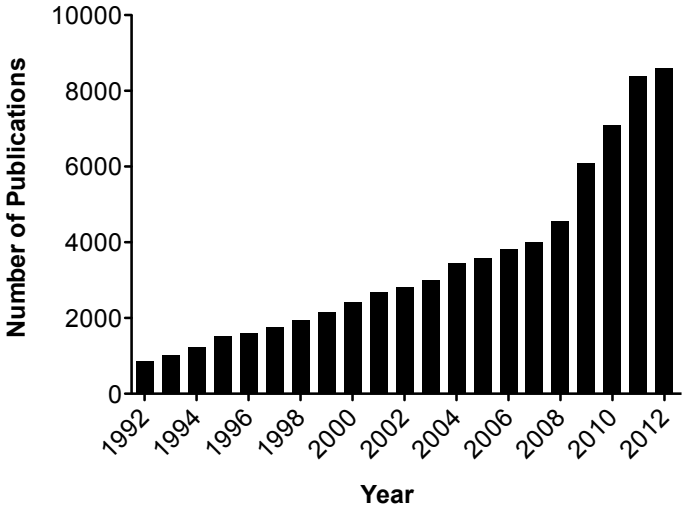

Figure 5: Number of gene targeting papers published in the past 20 years
The data collected was obtained using the PubMed database for the following search terms: gene targeting, homologous recombination, and ES cell technology.

5.2 Homologous recombination-mediated gene targeting

Gene targeting strategies commonly make use of homologous recombination (HR) events occurring between endogenous target gene loci and artificially introduced DNA to generate genome modifications in mice[98]. Using mESCs, HR-mediated genetic alterations can range from producing subtle gene disruptions like point mutations to causing complete gene ablations[99,100]. HR occurs naturally at a relatively low frequency

of approximately one event for every 10^5 to 10^7 cells and has been recognized to be one of two major DNA repair mechanisms in mammalian cells[101]. In response to DNA damage, particularly in the form of double-strand breaks (DSBs), the repair machinery of any given cell will attempt to repair the break by exchanging homologous sequences with the sister chromatid template or with those of artificially introduced DNA constructs[102,103]. Alternatively, the affected cell may resort to the second major repair mechanism, non-homologous end joining (NHEJ), which often results in imprecise rejoining of severed DNA strands ultimately leading to various gene disruptions[104,105]. A given gene of interest may be altered at the endogenous locus by designing a targeting vector so that "arms" of homology flank the desired gene modification. However, since HR-mediated gene targeting relies heavily on the low occurrence of DSBs, long arms of homology are required to increase the frequency of HR and to attain detectable targeting events[106]. Moreover, due to the complexity of mammalian genomes, targeting vectors are more likely to be inserted randomly into the mouse genome rather than in a specific manner as observed in yeast[107,108]. To circumvent this, negative selection can be employed by cloning a gene fragment for diphtheria toxin A (D-TA) with a poly-A signal (DT-ApA) outside of the homology arms[109]. Given that the A subunit inhibits protein synthesis, most random integrations will incorporate the cassette which will help select against these clones, whereas those that have successfully undergone HR will not have incorporated it[109]. Additionally, clones are typically screened further by Southern blotting in order to isolate those that are void of random integration and are correctly targeted[110].

5.3 Historical landmarks

Traditionally, conventional HR-based gene targeting approaches relied almost exclusively on the availability of conveniently positioned restriction enzyme sites to be able to construct targeting vectors (Figure 6a). However, this strategy suffers from two main limitations when generating the targeting construct. First, the likelihood of utilizing unique restriction enzyme cleavage sites surrounding the desired fragment of genomic DNA becomes almost impossible when fragment sizes are hundreds of kilobases long. In addition, traditional PCR amplification approaches are relatively error-prone, also limiting the length of desired fragments. Secondly, the cloning efficiency is significantly reduced when attempting to introduce large fragments of DNA into conventional vectors that can only hold a maximum of approximately 20 kb of foreign DNA. For this latter reason, other vector types like yeast artificial chromosomes (YACs), phage artificial chromosomes (PACs), and bacterial artificial chromosomes (BACs) were developed to accommodate large segments of DNA and to provide better feasibility in constructing large regions of homology. Of these, BACs are the most commonly used as they can accommodate up to 200 to 300 kb of foreign DNA and have proven to yield targeting efficiencies as high as 28% in mESCs[111,112]. Additionally, the replication of BACs is strictly controlled, as they are usually maintained in no more than two copies within host bacteria, thereby avoiding recombination events that may occur between homologous DNA fragments[113].

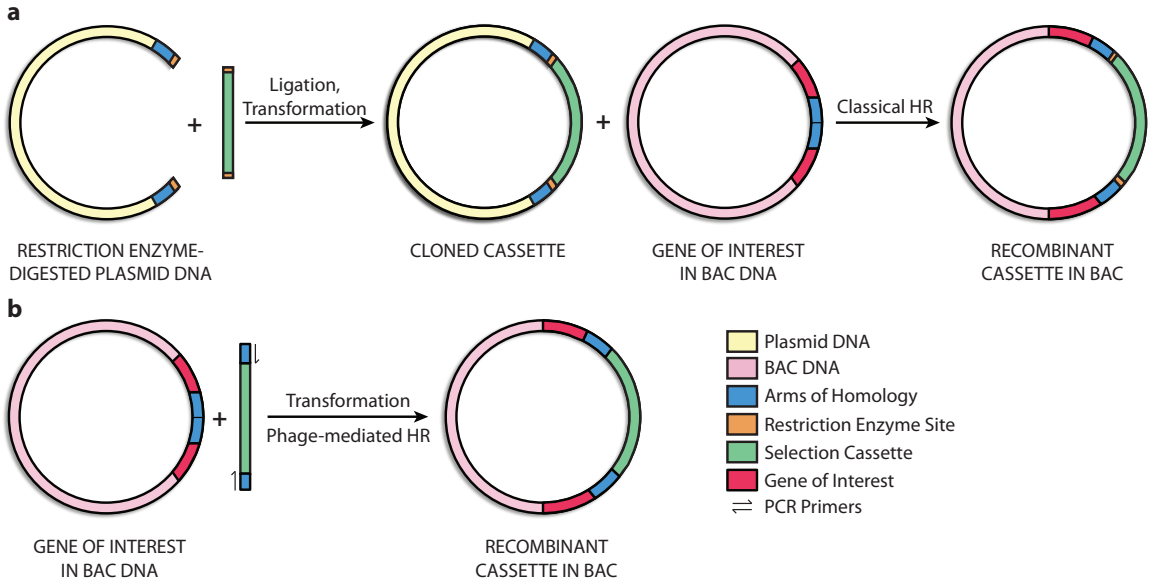

Figure 6: Conventional recombinant DNA technology versus phage-mediated recombineering
a) Conventional gene targeting relies on the use of unique restriction enzyme cleavage sites to introduce a selection cassette into a plasmid backbone that has already been digested with the same restriction enzymes. After ligation and transformation into competent cells, drug resistant clones are selected for and subsequently transformed into a BAC-containing bacterial strain to incorporate the cassette into the BAC by classical HR. **b)** Recombineering bypasses the use of an intermediate plasmid prior to BAC incorporation. The selection cassette is amplified to contain short regions of homology and transformed into bacteria containing a BAC and phage recombination functions. Adapted from Copeland *et al*., 2001[114].

Many of the issues associated with conventional targeting approaches have been somewhat alleviated by taking advantage of endogenous recombination machinery in *Escherichia coli* (*E. coli*)[115]. Recombinogenic engineering or recombineering, depends on the cooperation between one of the bacterial host's strand invasion proteins and an exonuclease (RecA and RecBCD in *E. coli*, respectively) along with linearized double-stranded DNA to initiate the recombination process[115]. Similarly, an alternative and comparatively more efficient method of recombineering makes use of phage-derived

protein pairs to aid in the recombination process of introducing desired gene modifications and selection cassettes in BAC DNA (Figure 6b)[115-117]. Unlike the previous method, this approach sometimes referred to as λ-mediated recombination, depends on the interaction of one protein pair encoding 5'-3' exonucleases and another pair coding for DNA annealing proteins[118]. The most significant advance of recombineering technology compared to conventional gene targeting is that only short arms of homology (<60 bp) are required to obtain efficient HR[119]. More importantly, recombineering bypasses the need to construct an intermediate targeting vector containing a selection cassette. The selectable gene can be amplified to produce a linear PCR product containing short regions of homology which can then be used to incorporate into a BAC vector[120]. Furthermore, this methodology presents a relatively more time saving and cost-effective way of manipulating BACs, which can subsequently be used as targeting vectors in mouse ES cells. However, due to the large size of BACs, conventional screening methods like Southern blotting or end-point PCR cannot be carried out to isolate correctly targeted clones. Therefore, either additional modifications to the BAC targeting vector are warranted, or more complicated screening techniques like fluorescence *in situ* hybridization (FISH) and qRT-PCR-based "loss-of-allele" assays are typically used[112,121].

5.4 Site-directed nucleases

In recent years, new technologies have been developed to overcome the drawbacks of recombineering and BAC-mediated gene targeting. One emerging approach to improve the poor HR efficiencies is the introduction of DSBs by an external source, which can

result up to a 1000- to 5000-fold increase in HR events compared to BAC transgenesis[122,123]. Such techniques are based on site-directed nucleases (SDNs), which generate DSBs at specific target gene loci to facilitate a higher percentage of HR events[124]. Initially, the homing endonuclease *Sce*I proved to be an efficient way of introducing DSBs in mouse ES cells, stimulating targeting efficiencies up to 50-fold[125]. However, due to the fact that *Sce*I requires a specific recognition sequence to be introduced at the target locus, other SDNs were developed as alternatives.

5.4.1 Zinc-finger Nucleases

Zinc-finger nucleases (ZFNs) are synthetic enzymes possessing DNA binding and DNA cleavage domains. Initial reports on ZFNs described these molecules as having a central zinc finger domain that provides binding specificity, which is fused to the catalytic domain of *Fok*I nuclease[126,127]. Zinc fingers are naturally found in many eukaryotic proteins and play an important role in facilitating nucleic acid-protein interactions. Each zinc finger recognizes a string of three nucleotides by binding into the major groove of the DNA helix (Figure 7a)[128]. Additionally, dimerization of the *Fok*I nuclease domain is essential to introduce DSBs at predetermined target sites[129]. Thus, a typical ZFN-mediated targeting experiment involves engineering a pair of ZFNs, each of which binds opposite strands of the DNA molecule to facilitate the heterodimerization and activation of the *Fok*I cleavage domain.

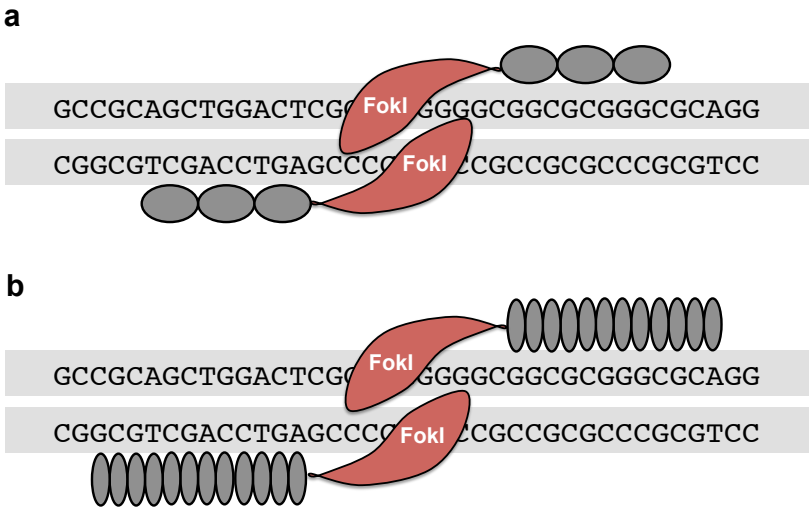

Figure 7: Schematic of the differences between ZFN and TALEN binding
a) Zinc finger motifs (horizontal ovals) recognize nucleotide sequences in triplets. **b)** TALEN monomers (vertical ovals) recognize individual nucleotides of the target site. *Fok*I nuclease cleaves following dimerization within the spacer region to create a DSB.

By appropriately designing ZFNs to recognize defined sequences of at least 18 bp, a unique target within the mammalian genome will likely be recognized. However, ZFN monomers are also prone to homodimerizing at unwanted locations in the host genome, leading to unanticipated gene disruptions. Thus, even though engineering ZFNs to recognize more than 18 bp will increase binding efficiency, the additional zinc finger motifs will increase the likelihood of binding off-target sites. In addition, some cell types display increased ZFN-induced cytotoxicity as a result of mis-targeting, which further limits the gene targeting potential of ZFNs. Therefore, while ZFN technology displays a number of advantages, it is important to develop novel ways of reducing the number of off-target DSBs and associated cytotoxic effects.

5.4.2 Transcription Activator-Like Effector Nucleases

Recently, a novel class of tools for genome editing using SDNs was developed, which makes use of transcription activator-like effector (TALE) proteins derived from the plant pathogenic bacteria *Xanthomonas*[130]. Naturally, TALEs infect plants by injecting effector proteins into the host cell that bind specific DNA target sequences to effectively suppress plant defense and support bacterial virulence[130]. The central domain of such TALEs consists of 33-35 tandem amino acid repeats, with the total number of repeats ranging from 12 to 27. Within each repeat, two adjacent amino acid residues at positions 12 and 13 collectively called a repeat-variable di-residue (RVD), serve to bind a single nucleotide, thereby giving a TALE its DNA-binding specificity (Figure 8)[131,132].

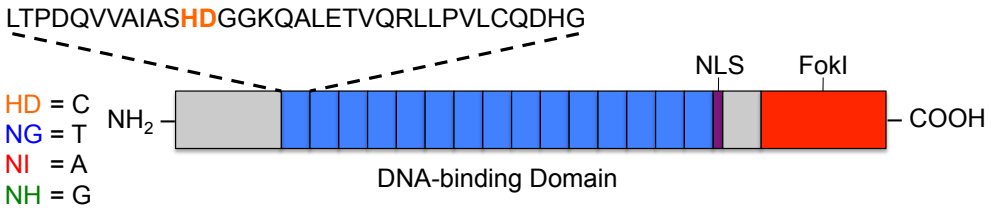

Figure 8: Schematic representation of a TALEN monomer
The monomer consists of a central domain containing 12-15 repeats (blue boxes) of 33-35 amino acids. The 12th and 13th amino acid of each repeat, called the RVD (e.g. HD RVD highlighted in orange), varies among repeats and gives the monomer its specificity. The four most common RVDs are HD, NG, NI, and NH, which bind to cytosine, thymine, adenine, and guanine, respectively.

By utilizing such target specificity, a number of groups have demonstrated that, similar to ZFNs, the fusion of TALEs to the catalytic domain of *Fok*I nuclease can be used to carry out targeted genome modifications *in vivo*[133-136]. Like ZFNs, two separate TALE nucleases (TALENs) must be constructed in order to induce a DSB in the target DNA, as *Fok*I nuclease must dimerize for it to be catalytically active (Figure 7b). The fact that a

TALEN monomer can be customized to recognize sequences at the individual nucleotide level, allows for a simpler design process compared to that required for ZFNs. Moreover, the technology can be used with HR-directed repair of DSBs by co-transfecting a targeting vector that includes arms of homology flanking the targeting site. This way, not only can TALENs be used to generate knockout cell lines via NHEJ, but they may also be used to target genes of interest through HR to generate desired modifications (i.e. fusions, deletions, knockouts, and knock-ins). Importantly, the use of TALENs significantly increases the efficiency of HR as much shorter arms of homology (<1 kb), compared to conventional approaches, can be employed in targeting vectors. Thus, clones generated via TALEN-mediated HR can be readily screened using standard PCR methods. Unlike ZFNs, custom TALEN monomers can be assembled by joining the tandem repeats using conventional cloning techniques[137]. However, due to the high degree of homology and number of repeats, such methodologies are difficult and impractical. For this reason, a novel high throughput approach called Golden Gate cloning was developed[138,139]. The Golden Gate method of cloning is a series of digestion-ligation reactions, which exploits Type IIS restriction enzymes to build an array of DNA fragments in a sequential manner. Using this cloning strategy, custom-made TALEN-encoding plasmids can be generated in less than one week. A drawback of the technology however, is that there seems to be no conclusive evidence explaining why some custom TALENs are ineffective. Some speculations include possible interactions between repeat domains that may negatively influence target DNA binding affinity.

Despite being first reported in 2010, TALEN technology has established itself as a better alternative to ZFNs. In fact, gene disruption studies in human cells have shown that even though TALENs and ZFNs induce similar HR frequencies, TALENs exhibited much lower nuclease-associated cytotoxic effects and minimal off-target cleavage[140]. However, since it is still a relatively new technique, other unknown off-target effects may have yet to be discovered.

5.4 Future directions

5.4.1 Type II CRISPR/Cas System

Clustered Regulatory Interspaced Short Palindromic Repeats (CRISPR) is an established defense mechanism in bacteria using RNA-guided cleavage of pathogenic DNA[141]. In the type II CRISPR/Cas system, CRISPR-associated (Cas) genes and a non-coding guide RNA facilitate sequence-specific cleavage of foreign DNA[142,143]. Specifically, short fragments of foreign DNA are integrated into the CRISPR loci and transcribed in CRISPR RNA (crRNA). The crRNAs then anneal to transactivating crRNAs (tracrRNAs), which guide Cas9 proteins to a specific site in the foreign DNA to initiate cleavage. To date, the effectiveness of CRISPR/Cas-mediated genome editing has been demonstrated in human iPSCs, as well as in zebrafish and bacteria[143,144]. However, a thorough validation of the system has yet to be conducted, as unintended off-target effects have yet to be fully examined. Emerging data appears to suggest that the specificity of CRISPR-mediated genome cleavage is not robust, as off-target sites have been shown to be cleaved with frequencies of equal to or greater than those observed at the intended site[145].

5.4.2 Artificial Restriction DNA Cutters

Artificial Restriction DNA Cutters (ARCUT) is a chemistry-based system for manipulating target DNA[146]. The system consists of the Ce(IV)/EDTA complex and pseudo-complementary peptide nucleic acid (pcPNA). Two pcPNA strands first bind to their target sequence but are offset by several nucleotides to form the "double-duplex invasion complex." The remaining unpaired single-stranded portion of the double-stranded DNA serves as a target for Ce(IV)/EDTA, which selectively cleaves only single-stranded DNA[147]. While the major advantage of ARCUT is the freedom of cleavage site selectivity, the system is still in its infancy, as more exhaustive studies have yet to be carried out in order to successfully validate the system.

6. Project Rationale

The emergence of the Tcf3-Tcf1 switch model of canonical Wnt/β-catenin signaling has directed the scientific community to appreciate the potentially critical role of Tcf1 in mESC self-renewal and differentiation. Thus far, isoform-specific roles of Tcf1 in mESCs have not been delineated, due at least in part to the low sensitivity of available antibodies, and their cross-reactivity with other Tcf/Lefs. As a result, the experiments in this thesis will aim to genetically engineer an epitope tag at the N-terminus of both the full-length and dominant-negative isoforms of Tcf1 to facilitate studying the role of these isoforms in mESC biology. The implications of this design are significant, as endogenously tagging Tcf/Lefs provides a definitive way to study these factors, and eliminates any ambiguity in their detection. Not only will this model be important from a developmental perspective, but it will also provide a framework for understanding the

nuclear signaling events that arise from dysregulated Wnt signaling in numerous human cancers. Furthermore, our approach can also be used to fuse fluorescent proteins to endogenous Tcf1 isoforms, providing an unprecedented opportunity to study the regulation of their expression and subcellular localization in living cells.

Hypotheses

1. ESC fate is determined by the relative expression and cellular localization of all Tcf/Lef factors as well as their isoforms.
2. FL-Tcf1 is necessary for the maintenance of mESC self-renewal and pluripotency, and a switch from FL-Tcf1 to DN-Tcf1 expression occurs in response to cues for differentiation.

CHAPTER 2
MATERIALS AND METHODS

1. Cell culture

The growth and maintenance of mESCs (E14Tg2a) were carried out in DMEM (Thermo Scientific) supplemented with 15% FBS (Gibco), 1 × nonessential amino acids (Thermo Scientific), 1 × L-glutamine (Thermo Scientific), 1 × sodium pyruvate (Thermo Scientific), and 55 μM β-mercaptoethanol (Sigma). The media was filter-sterilized with a 0.22 μm filter (Millipore) after which 1000 U/mL ESGRO® LIF (Millipore) was added. Cells were plated on tissue culture-treated plates that were pre-coated with 0.1% gelatin, and grown in a humidified incubator at 37°C and 5% CO_2. Passaging of mESCs was carried out every other day with Accutase (Innovative Cell Technologies), and typically split at a 1:6 ratio.

2. Embryoid body assay

The medium used (EB media) for differentiating mESCs was identical to the above standard mESC media, but contained 5% FBS (rather than 15%) and lacked LIF. Cells were initially grown on inactivated mouse embryonic fibroblasts (iMEFs) before pre-plating to remove the feeders. Cells were then counted using the Countess® Automated Cell Counter (Invitrogen) and plated in EB media onto ultra-low attachment 6-well dishes (Corning). Upon lysis, EBs were collected and washed twice with 1 × PBS by centrifugation at 270 × g for 5 minutes.

3. Transfection and gene targeting

Transfection of cells was performed using Lipofectamine® LTX with Plus™ Reagent (Invitrogen) and counted using the Countess® Automated Cell Counter (Invitrogen). For targeting experiments, 4×10^6 mESCs were transfected with 5 μg of each TALEN vector and 1 μg of donor, and plated at clonal density. The cells were fed 24 hours following transfection, and neomycin-resistant clones were selected for with media containing G418 (250 μg/mL) 48 hours post-transfection. Six to eight days later, 60 colonies were isolated and replica-plated to carry out PCR-based screens for successful HR. Once clones were confirmed to have undergone successful HR, the neomycin cassette was excised by transfecting 2×10^6 cells with 2.5 μg of pCX-Cre-Puro and plating at clonal density. Transfected cells were pulsed with 2 μg/mL puromycin the day after transfection for 24 hours to select for those expressing the pCX-Cre-Puro, after which they were grown in standard mESC media for 6-8 days. Individual clones were picked and replica-plated to screen for their sensitivity to G418. Those cells that were no longer resistant to G418, due to the deletion of the neomycin cassette, were isolated and expanded. The plasmid encoding Cre recombinase, pCX-Cre-Puro, was provided as a gift from Dr. Jonathan Draper (McMaster University).

4. shRNA-mediated knockdown of Tcf1 expression

The shRNA-mir construct against exon 7 of *mTcf7* was purchased from Transomic Technologies™ (ID: RLGM-GU66089) and contained the following guide sequence: TATGACCTTGGCTCTCATCTCC. The construct was linearized using *Pvu*I and subsequently transfected into mESCs maintained under puromycin (2 μg/mL) selection.

Cells were then isolated, expanded, and characterized for effective knockout using western blot analysis. An shRNA-mir non-targeting construct (NegSH) was used as a negative control (Transomic Technologies™, ID: TRM1103).

5. Antibodies

The following primary antibodies were used for western blotting analysis and/or immunofluorescent staining: rabbit anti-Tcf1 (C46C9, Cell Signaling), rabbit anti-Tcf1 (C63D9, Cell Signaling), goat anti-Tcf3 (sc-8635, Santa Cruz), rabbit anti-Tcf4 (C48H11, Cell Signaling), rabbit anti-Lef1 (C12A5, Cell Signaling), mouse anti-3XFLAG (F1804, Sigma), rabbit anti-Nanog (A300-397A, Bethyl Laboratories), and mouse anti-GAPDH (ab8245, Abcam).

Horseradish peroxidase-conjugated secondary antibodies used for western blotting were obtained from Bio-Rad (goat anti-mouse and goat anti-rabbit; 170-6516 and 170-6515, respectively). Fluorochrome-conjugated secondary antibodies used for immunofluorescent staining were purchased from Invitrogen (goat anti-mouse Alexa Fluor 488 and goat anti-rabbit Alexa Fluor 647; A-11001 and A-21245, respectively).

6. Cell lysate preparation

Whole cell lysates were prepared by washing dishes of cells twice with 1 × PBS before lysing on ice with 1 × RIPA buffer [50 mM NaCl, 1% NP-40, 0.5% DOC, 0.1% SDS, 50 mM Tris pH 8.0, 1 mM EDTA, and Halt Protease Inhibitor Cocktail (Thermo Scientific)] for 10 minutes. Using a cell scraper, lysed cells were transferred to 1.5 mL microcentrifuge tubes and centrifuged at 16, 100 × g at 4°C for 10 minutes. The

supernatant was then collected and quantified using the Lowry method (DC Protein Assay, Bio-Rad). Samples were either normalized to 0.5 µg/µL or 1 µg/µL in 1 × LDS buffer (Thermo Scientific) with 5% TCEP Bond-Breaker solution (Thermo Scientific), and heated at 95°C for 5 minutes prior to electrophoresis.

7. Western blot analysis

Protein samples (10-15 µg per well) were separated using 10% Bis-Tris gels and transferred onto PVDF membranes using the Trans-Blot® Turbo™ Blotting System (Bio-Rad). Membranes were blocked with 5% skim milk/TBS for 30 minutes at room temperature and incubated in primary antibody (3% milk/TBS-T) at 4°C on a rotator overnight. Primary antibody dilutions were as follows: Tcf1 (1:1000), Tcf3 (1:1000), Tcf4 (1:1000), Lef1 (1:1000), 3XFLAG (1:2500), Nanog (1:5000), GAPDH ($1:10^6$). The next day, blots were washed 4-5 × with 3%/TBS-T, 10 minutes per wash, and incubated in the appropriate horseradish peroxidase-conjugated secondary antibody for 45 minutes at room temperature. All secondary antibodies used were diluted to 1:20 000 in 3%/TBS-T. Blots were then washed 5-6 × with 1 × TBS-T, 10 minutes per wash, prior to a 5 minute incubation in Luminata Forte Western HRP substrate (Millipore) for enhanced chemiluminescence detection using either standard film (CL-XPosure, Thermo Scientific) or the ChemiDoc™ MP Imaging System (Bio-Rad).

8. Immunofluorescent staining

All incubation and wash steps were carried out at room temperature unless specified otherwise. mESCs were prepared for immunofluorescent staining by seeding 30 000 cells

onto each well of a gelatin-coated 8-well μ-Slide (Ibidi). The next day, cells were washed twice with 1 × PBS for 5 minutes each, and fixed in freshly made 4% paraformaldehyde/PBS for 15 minutes. The fixed cells were then incubated with 0.2% Triton X-100 for 5 minutes, washed twice with 1 × PBS, and blocked in 2% BSA/PBS-T (0.1% Tween) for 30 minutes prior to incubating in primary antibody overnight at 4°C or 1.5 hours at room temperature. Primary antibody dilutions were as follows: Nanog (1:500), and 3XFLAG (1:500). Cells were then rinsed 3 × with 1 × PBS before incubating in the appropriate fluorochrome-conjugated secondary antibody (1:1000) for 1 hour. Finally, the cells were rinsed 3 × with 1 × PBS prior to adding 1-2 drops of ProLong® Gold antifade (containing DAPI) solution (Invitrogen).

9. Generation of targeting vectors

Genomic DNA from wild-type mESCs (E14Tg2a cell line) was PCR-amplified (Herculase II Fusion DNA Polymerase, Agilent Technologies) using primers flanking a coding region of *mTcf7*. Arms of homology flanking the 5' and 3' ends (5'HA and 3'HA, respectively) of this region were also generated using this method, and ranged from 1kb to 1.5kb in size. Incorporation of a 3XFLAG epitope tag at the N-terminus of both isoforms was engineered using the p3XFLAG-CMV10 plasmid vector (Sigma), and subsequently ligated into the pL452 targeting vector[148]. Likewise, targeting vectors for fluorescent-tagging the N-terminus of FL-Tcf1 and DN-Tcf1 with monomeric Azami-Green (mAG) were designed in a similar manner, except that the 3XFLAG was swapped with mAG from pCAG-mAG-Skp2, which was gifted by the J. Draper Lab (McMaster University). Details of vector construction are outlined below. All primers were designed

using IDT's online software (idtdna.com) or Vector NTI® (Invitrogen) and their sequences are outlined in Table 1 below.

9.1 N-terminal tagged 3XFLAG-Tcf1

All PCR products were gel purified (MinElute®, Qiagen), treated with blunting enzyme (Thermo Scientific), and cloned into pJET1.2/blunt (Thermo Scientific) prior to all ligation steps. The 3'HA was amplified using primers engineered to contain *Bam*HI and *Not*I (primers A, B, I, and J) restriction sites, and cloned into pL452 (pL452_3'HA). Next, primers flanking exons 1 and 2 for FL-Tcf1 (primers C and D), and exon 3 for DN-Tcf1 (primers K and L) were used to incorporate *Hind*III/*Kpn*I and *Hind*III/*Eco*RI sites, respectively. These regions were cloned directly downstream of the 3XFLAG sequence in p3XFLAG-CMV-10. Since a convenient and unique restriction enzyme site is not present directly upstream of the 3XFLAG sequence, the 5'HA was incorporated using the In-Fusion® HD Cloning Kit (Clontech) as per the manufacturer's protocol. In brief, the 5'HA was amplified using primers generated using the online In-Fusion® Primer Design software (Primers E, F, M, and N). This product, along with the linearized product of p3XFLAG-CMV-10 (using primers G, H, O, and P) were mixed in a single In-Fusion® reaction. Finally, the 5'HA-3XFLAG-coding-3'HA fragment was cloned into pL452_3'HA using *Kpn*I or *Kpn*I/*Eco*RI for FL-Tcf1 and DN-Tcf1, respectively. In the case for FL-Tcf1, the cloned fragment was screened for directionality using *Age*I. Both targeting vectors were also sequence-verified to confirm the absence of PCR-induced mutations (MOBIX Lab, McMaster University).

Table 1: PCR Primers Used to Generate Targeting Vectors

Primer		Sequence
FL-Tcf1	A	5'-GGATCCTTTTTAACCCCCCTCCGGAGCAGACA-3'
	B	5'-GCGGCCGCCATCCCAGAG-3'
	C	5'-GATTACAAGGATGACGAT-3'
	D	5'-GGTACCAAAAAAAAAAACCCTGCCAAGGGGTC-3'
	E	5'-CCGTCAGAATTAACCGGTACCGTCCTGTGGGGCCGGGAGTC-3'
	F	5'-GTCTTTGTAGTCCATGGTGCGCTGCGCCCGCCG-3'
	G	5'-ATGGACTACAAAGACCAT-3'
	H	5'-GGTTAATTCTGACGGTTC-3'
DN-Tcf1	I	5'-GGATCCCTGAAAATACTGGGATTCCAGAGTGT-3'
	J	5'-GCGGCCGCAGCTGATACC-3'
	K	5'-AAGCTTATGTACAAAGAGACTGTC-3'
	L	5'-GAATTCGAACCCCAGAACTTTTCAAAAGTCA-3'
	M	5'-CCGTCAGAATTAACCGGTACCAGAAGCCGAGGGAGCGCC-3'
	N	5'-GTCTTTGTAGTCCATGCCACTGGCACACTCCGGGGCC-3'
	O	5'-ATGGACTACAAAGACCAT-3'
	P	5'-GGTTAATTCTGACGGTTC-3'

10. TALEN pair design and construction

The TALEN vectors used to facilitate HR-directed repair were designed using the online tool, TAL Effector Nucleotide Targeter[149]. TALEN pairs were constructed using two different protocols, which are briefly outlined below.

10.1 Cermak et al., 2011

The construction of DN-Tcf1 TALEN pairs was carried out using this method, which involves two primary steps: 1) assembly of 1-10 repeat module plasmids (i.e. RVDs) into intermediary arrays, and 2) joining of the intermediary arrays into an acceptor backbone

vector containing a transgene coding for *Fok*I nuclease[150]. Both forward and reverse TALENs consisted of 20 RVDs separated by a 19 nucleotide spacer region. The forward TALEN contained the RVD array HD NG NN NI HD HD HD NI NN HD NG NN NG NN NN NG NG NG NG NG and bound to the sequence 5'-CTGACCCAGCTGTGGTTTTT-3', while that of the reverse TALEN was NN NG NI HD NI NG NN HD HD NI HD NG NN NN HD NI HD NI HD NG which corresponded to 5'-AGTGTGCCAGTGGCATGTAC-3'. Considering the forward TALEN as an example, RVDs 1-10 were cloned into the vector array plasmid, pFUS_A, and RVDs 11-19 were cloned into a second array plasmid, pFUS_B. A separate 20 µL digestion-ligation Golden Gate reaction using *Bsa*I was performed for each set of arrays. Next, the reactions were treated with Plasmid-Safe DNase (Epicentre) and 5 µL of each reaction was used to transform competent *Mach1* bacterial cells, which were plated onto LB agar containing 50 µg/mL spectinomycin with 20 mg/mL X-gal and 0.1 M IPTG for blue/white screening. The next day, white colonies were picked and inoculated, and clones were screened the following day using PCR. Positive pFUS_A and pFUS_B were then mixed into a single Golden Gate reaction using *Esp*3I, and contained the last repeat plasmid, pLR-NG, as well as the final acceptor vector pTAL3 encoding *Fok*I nuclease. Once completed, 5 µL of the reaction was used for transformation and were plated on LB agar containing 100 µg/µL carbenicillin. White colonies were inoculated the next day and plasmid mini-preps were isolated following overnight culture. A diagnostic digest was performed using *Stu*I, and constructs were further verified with sequencing.

10.2 Sanjana et al., 2012

The FL-Tcf1 TALENs were generated using this method, which involves dividing the TALEN recognition sites into hexamers initially, and further ligating these hexamers by carrying out a series of PCR amplification steps[151]. The forward and reverse TALENs consisted of 20 and 15 RVDs, respectively, and were separated by a 15 nucleotide spacer. The recognition sequences are as follows: forward TALEN, HD HD NH HD NH HD HD HD HD NH HD NI HD HD HD NH NH HD NH HD corresponding to 5'-CCGCGCCCCGCACCCGGCGC-3'; reverse TALEN, HD HD NI NH HD NG NH HD NH NH HD NI NG NH NH corresponding to 5'-CCATGCCGCAGCTGG-3'. Each RVD monomer was amplified using specific primers in order to add appropriate ligation adapters. TALEN recognition sequences were divided into hexamers, as monomers 1-6, 7-12, and 13-18 were added to 3 separate Golden Gate digestion-ligation reactions driven by *Bsm*BI to generate circularized hexameric tandem repeats. The reactions were then treated with Plasmid-Safe DNase to remove any linear DNA. Next, each tandem hexamer was PCR-amplified to generate a linearized product, which was subsequently purified and ligated into the appropriate final acceptor plasmid consisting of the last RVD and *Fok*I catalytic domain in a second Golden Gate reaction using *Bsa*I. The ligation product was then transformed into competent *Mach1* cells and plasmid DNA isolated from clones was screened by digestion with *Afe*I. Plasmids with appropriate *AfeI* digestion products were further validated by sequencing to confirm correct ligation and orientation of the hexameric repeats.

11. Quantitative RT-PCR

Total RNA was extracted from cells using the PureLink® RNA Mini Kit (Life Technologies), and 1 μg was reverse transcribed using qScript cDNA SuperMix (Quanta BioSciences) as per the manufacturers' protocols. The resulting cDNA was diluted to 10 ng/μL and 30 ng was used per PCR reaction using PerfeCTa® SYBR® Green FastMix® (Quanta BioSciences). Bio-Rad's CFX96 qPCR instrument and the accompanying CFX™ Manager 2.0 software were used to determine gene expression levels normalized to β-actin using the delta-delta Ct method. Primer sequences were either designed using the online software provided by IDT (idtdna.com) or obtained from a published source and are listed in Table 2 below.

Table 2: Quantitative RT-PCR Primers

Primer	Sequence	Reference
Actin (β)	FWD: 5'-TTGCTGACAGGATGCAGAAGGAGA-3'	[74]
	REV: 5'-ACTCCTGCTTGCTGATCCACATCT-3'	[74]
Axin2	FWD: 5'-AAAGAAACTGGCAAGTGTCCACGC-3	[74]
	REV: 5'-GGCAAATTCGTCACTCGCCTTCTT-3'	[74]
Lef1	FWD: 5'-AAGGCCAGAGAACACCCTGATGAA-3'	N/A
	REV: 5'-CTGCACCACGGGCACTTTATTTGA-3'	N/A
Nanog	FWD: 5'-AACCAAAGGATGAAGTGCAAGCGG-3'	[74]
	REV: 5'-TCCAAGTTGGGTTGGTCCAAGTCT-3'	[74]
Tcf1	FWD: 5'-CCAGTGTGCACCCTTCCTAT-3'	[66]
	REV: 5'-AGCCCCACAGAGAAACTGAA-3'	[66]
Tcf3	FWD: 5'-CCCCCTACTTTCCCAGCTAC-3'	[66]
	REV: 5'-CTTTGTGTTTCCCCCTTCCT-3'	[66]
Tcf4	FWD: 5'-AGCCATCAACCAGATTCTCG-3'	N/A
	REV: 5'-CTTCCCATAGTTATCCCGTGC-3'	N/A

12. Statistical analysis and imaging

All statistical analyses, including analysis of variance (ANOVA) and unpaired t-tests, were carried out using the software package Prism (GraphPad). Cells were visualized for fluorescence using the IX81® inverted microscope with DSU (Olympus) and accompanying MetaMorph software (Molecular Devices).

CHAPTER 3
RESULTS

1. TALEN-facilitated homologous recombination to epitope-tag Tcf1 isoforms

Two sets of TALEN pairs were constructed to facilitate HR of donor vectors to introduce the 3XFLAG epitope or monomeric Azami-Green (mAG) tag at the N-terminus of FL- and DN-Tcf1 isoforms. mAG is a monomeric green fluorescent protein variant derived from the stony coral *G. fascicularis*[152]. Unlike traditional green fluorescent protein (GFP) and its derivatives, mAG can be used in applications related to fluorescence resonance energy transfer (FRET)[152].

Given that the DN-Tcf1 TALEN backbone (i.e. pTAL3) was not optimized for expression in mammalian systems, the TALEN binding and *Fok*I catalytic domains were cloned into a modified pcDNA5-FRT (Invitrogen) mammalian expression vector. This essentially served to place the TALENs under the control of the potent mammalian CMV (cytomegalovirus) promoter. On the other hand, the TALEN pair for FL-Tcf1 did not need to be altered as the backbone already contained a CMV promoter and codon optimized sequence for expression in mammalian cells.

Both sets of TALEN pairs were strategically designed so that they would exclusively cut the endogenous *mTcf7* locus and not the accompanying targeting vector being co-transfected. In order to do so, the TALEN cut site needed to overlap the boundary either between the 3' end of the 5' homology arm and the 5' end of the coding sequence, or the 3' end of the coding region and 5' end of the 3' homology arm. This strategy was successfully implemented in the design of both FL- and DN-Tcf1. In both cases, the

TALEN recognition sequence was at the 5'HA-Exon1 boundary and 5'HA-Exon 3 boundary for FL- and DN-Tcf1, respectively (Figure 9).

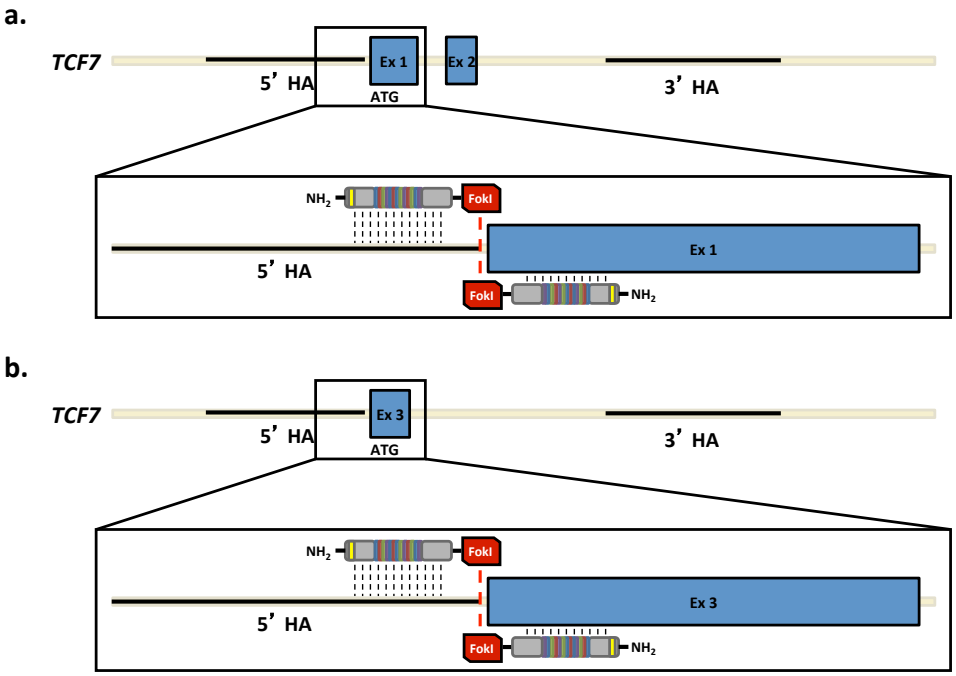

Figure 9: Schematic representation of TALEN pair binding sites at the endogenous *mTcf7* locus
TALEN-induced double-strand breaks upon *Fok*I dimerization (red dashed lines) were strategically designed to exclusively occur at the endogenous locus rather than the targeting vector. Inset provides a representation of TALEN binding (black dashed lines) in order to generate 3XFLAG-tagged full-length **(a)** and dominant-negative **(b)** isoforms of Tcf1.

In addition to the TALEN pair, the pL452 targeting vector was also co-transfected into wild-type mESCs. pL452 makes use of a neomycin cassette that is flanked by a single loxP site on either side, which is integrated into clones following transfection, either by HR or random integration. Treatment with Cre recombinase, transiently expressed by pCX-Cre-Puro, excised the neomycin cassette, which comprises the eukaryotic and prokaryotic promoters phosphoglycerate kinase-1 (PGK) and EM7, respectively and the

neomycin resistance gene coding sequence (NEO). As a result of the excision, a single loxP site was integrated into the *mTcf7* locus (Figure 10).

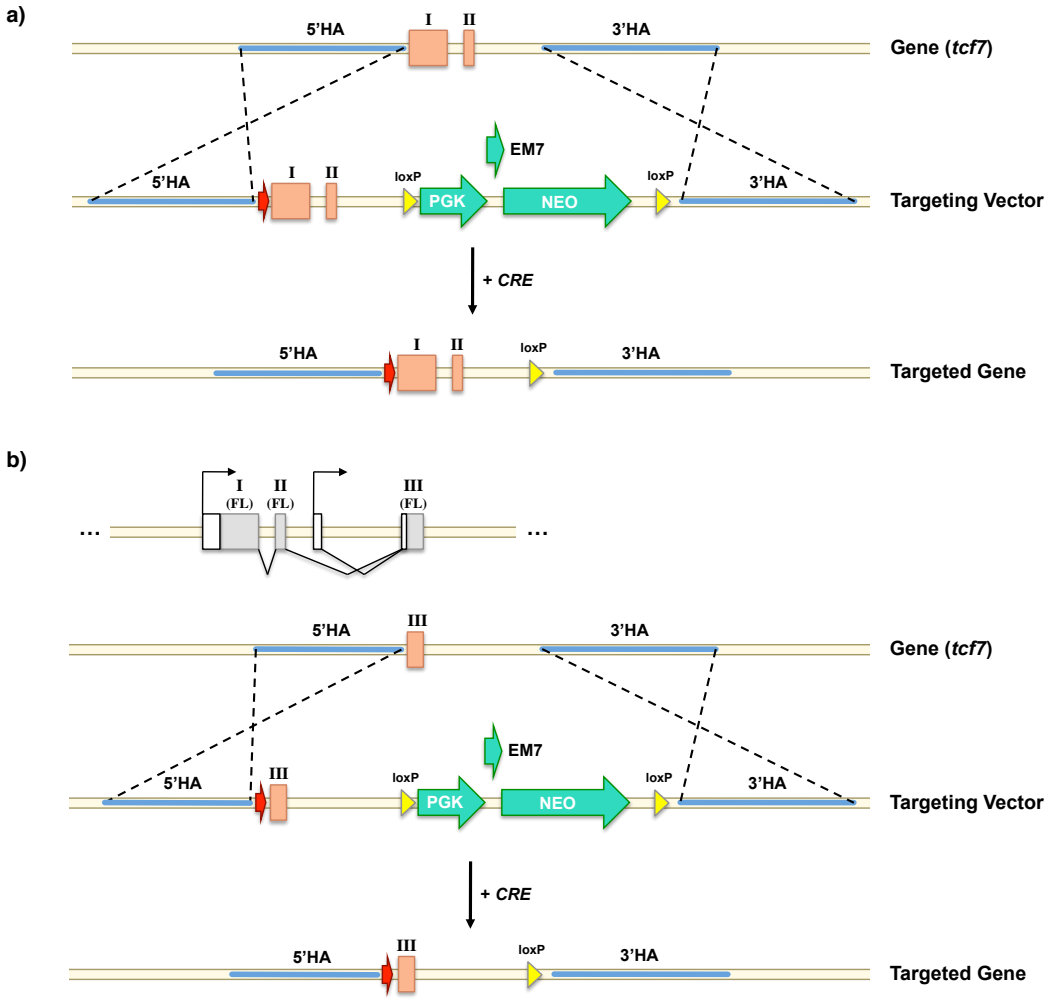

Figure 10: Schematic representation of the targeting design to incorporate an N-terminal 3XFLAG-tag at the *mTcf7* locus

Using TALEN technology to generate double-strand breaks, the above targeting vectors were co-transfected to carry out homologous recombination-directed repair (dotted lines) using arms of homology (HA), while incorporating a 3XFLAG epitope tag (red arrow) at the N-terminus of FL-Tcf1 **(a)** and DN-Tcf1 **(b)**. PGK and EM7 are eukaryotic and prokaryotic promoters, respectively. The NEO cassette was used to enrich for putative homologous recombinants by selection with the drug G418, which is toxic to wild-type mESCs lacking the cassette. Cre recombinase was then used to excise the NEO cassette leaving a single loxP site behind in the targeted locus.

Taking this into consideration, both arms of homology were designed such that the remaining loxP site following treatment with Cre recombinase would integrate into a non-conserved intronic region (Figure 11). Such regions of minimal to no conservation were analyzed by examining known transcription factor binding sites across a number of different species. Regions that are present in the mouse genome, and absent in most other organisms suggest that they are not evolutionarily conserved, and thus, were deemed nonessential to be able to serve as a suitable genomic location to incorporate the remaining loxP site.

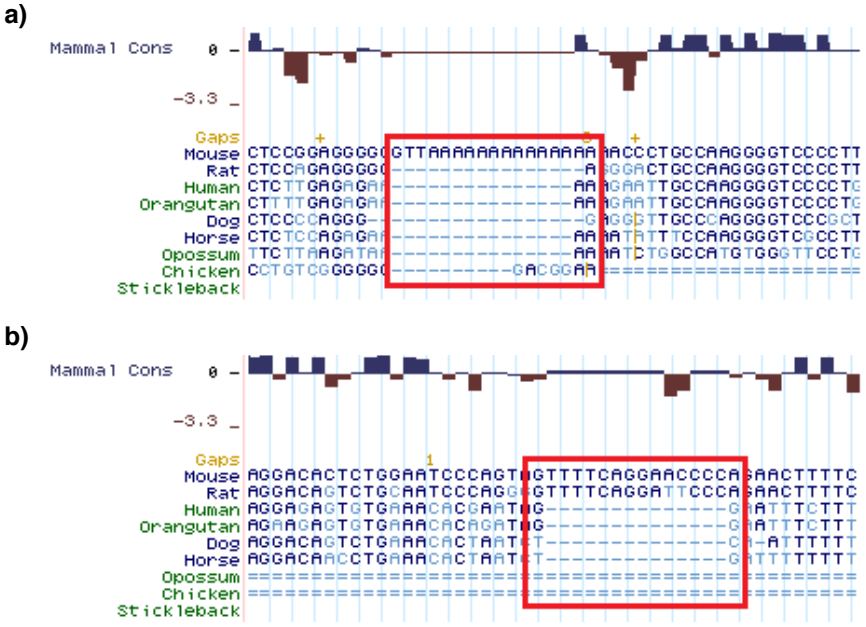

Figure 11: Non-conserved regions of *Tcf7*
Regions of low conservation across various species (outlined in red) were considered to be good candidate sites for loxP integration following Cre-mediated excision of the neomycin cassette after successful homologous recombination. **(a)** FL-Tcf1; **(b)** DN-Tcf1. Figure adapted from the UCSC Genome Bioinformatics Database[153].

2. Validation of proper homologous recombination and successful endogenous genome modification via PCR

The short arms of homology used in the TALEN-compatible targeting vectors permitted the rapid screening of mESC clones by PCR analyses, which allowed for the isolation of those that had undergone proper HR (Figures 12-15). Briefly, for the 3' arm of homology (i.e. 3'HA), a sense primer located within the neomycin cassette used in combination with an antisense primer located beyond the region of homology was used to discern between wild-type and targeted alleles. For the 5' end of the targeted gene (i.e. 5'HA), a sense primer located outside the region of homology together with an antisense primer located within the neomycin cassette was used to confirm proper recombination. While a working PCR-based screen was established for both homology arms in the DN targeting experiment, only the 3'HA screen was successfully implemented in the FL tagging experiment.

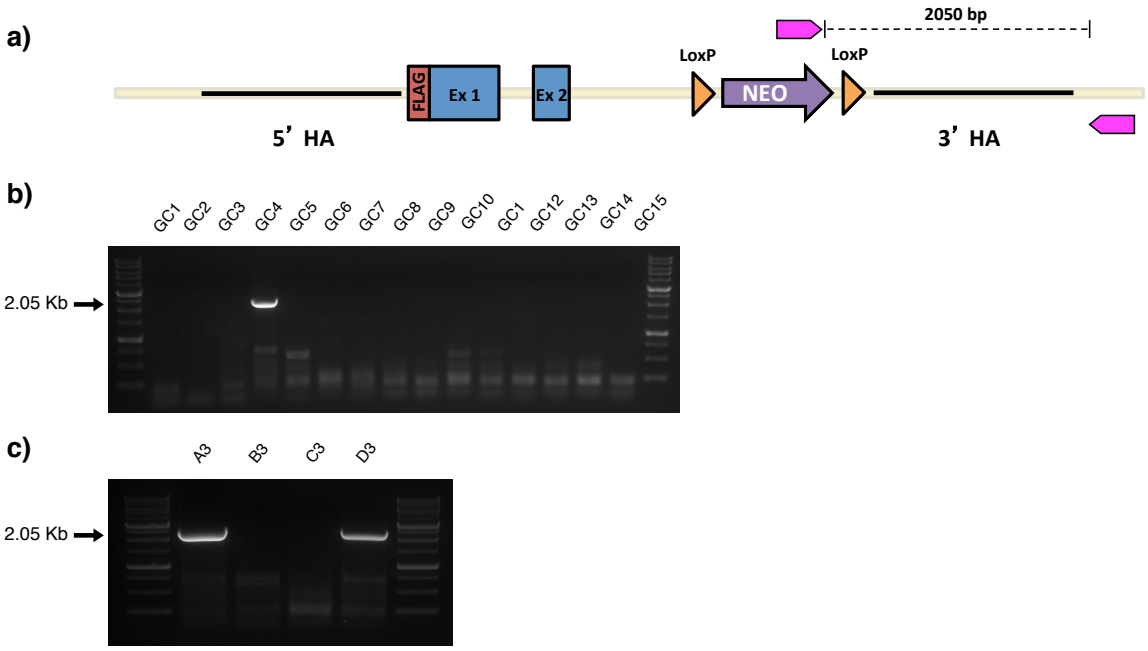

Figure 12: PCR-based screen for successfully targeted 3XFLAG-FL-Tcf1
a) Schematic of a working 3' homology arm PCR screen used to isolate targeted clones. The expected band size using primers for 3'HA incorporation (pink) is indicated. TALENs were co-transfected with the targeting vector. 60 neo-resistant colonies were picked and screened. **b)** Groups of 4 colonies were pooled and screened for the presence of the 3'HA. **(c)** The positive group (i.e. GC4) was subsequently screened for individual positive colonies. A working 5'HA screen was not established. The efficiency of targeting was approx. 3.3% (2/60 neo-resistant clones were positive for the 3'HA screen).

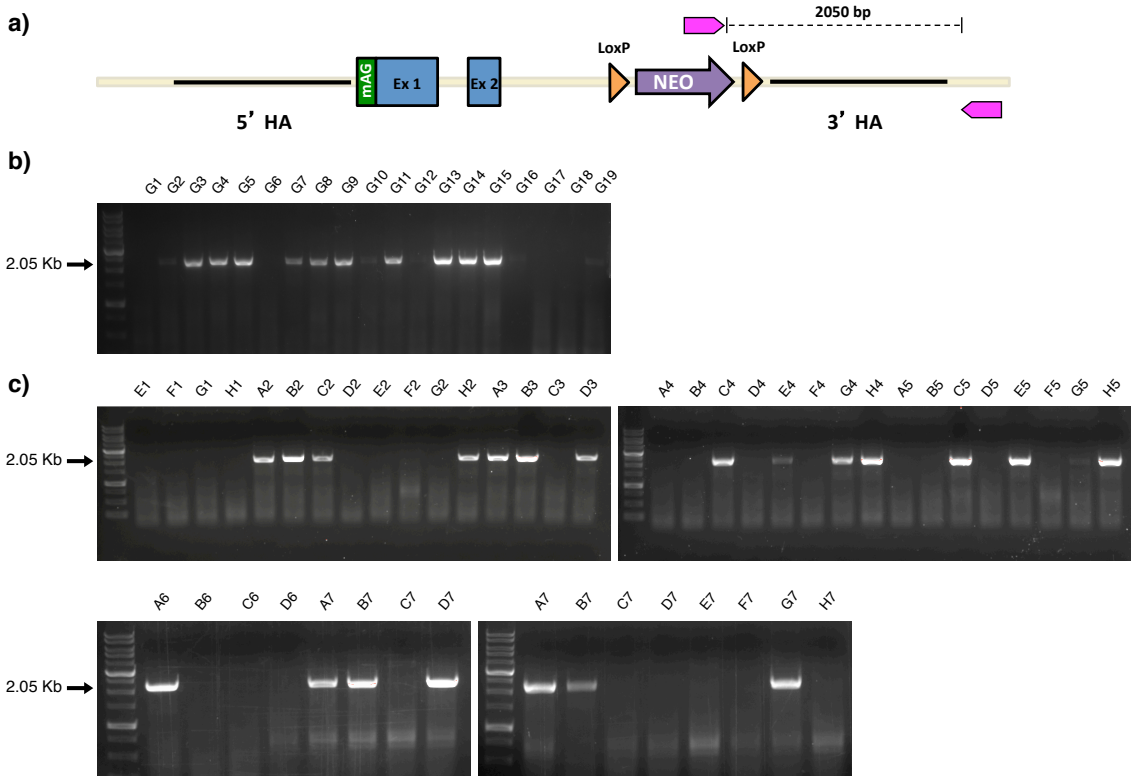

Figure 13: PCR-based screen for successfully targeted mAG-FL-Tcf1
a) Schematic of a working 3' homology arm PCR screen used to isolate targeted clones. The expected band size using primers for 3'HA incorporation (pink) is indicated. TALENs were co-transfected with the targeting vector. 80 neo-resistant colonies were picked and screened. **b)** Groups of 4 colonies were pooled and screened for the presence of the 3'HA. **(c)** The positive groups were further screened for individual positive colonies. A working 5'HA screen was not established. The efficiency of targeting was approx. 26.3% (21/80 neo-resistant clones were positive for the 3'HA screen).

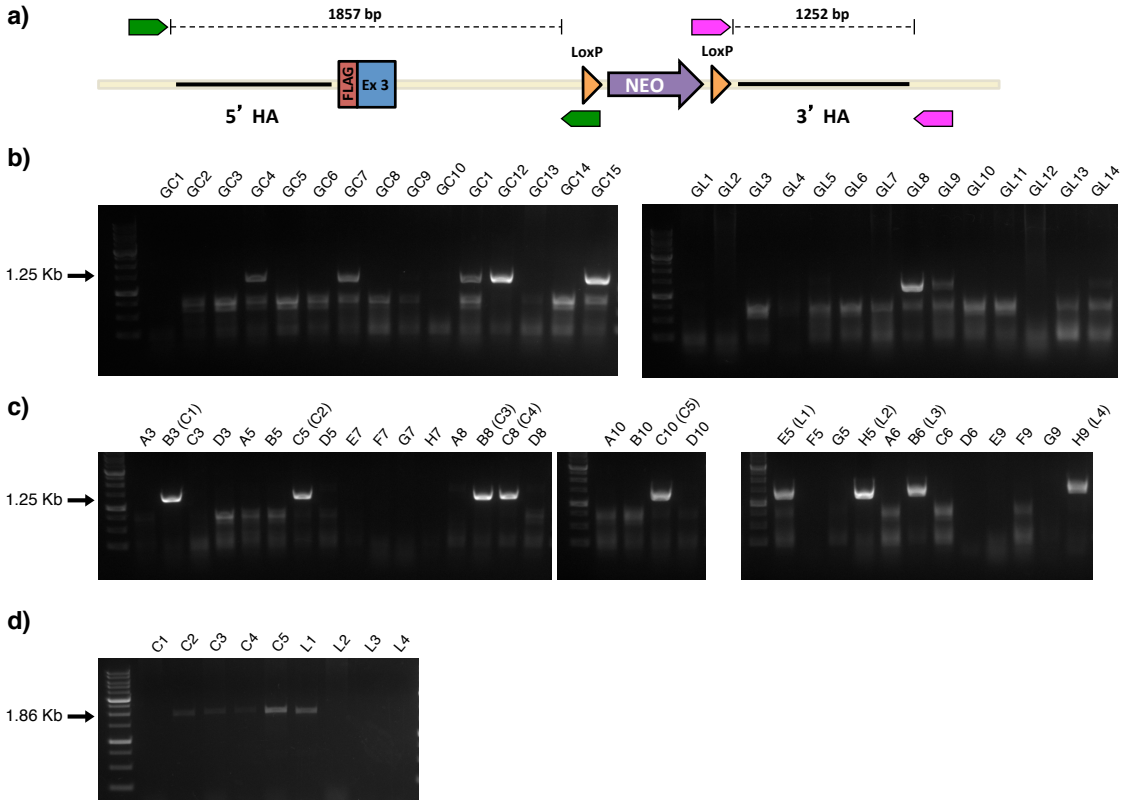

Figure 14: PCR-based screen for successfully targeted 3XFLAG-DN-Tcf1
a) Schematic of 5' and 3' homology arm (HA) PCR screens used to isolate targeted clones. Expected band sizes using primers for 5'HA incorporation (green) and 3'HA incorporation (pink) are indicated. TALENs were co-transfected with the targeting vector. Corresponding controls were transfected with the targeting vector and control plasmids lacking the TALEN portion. 60 neo-resistant colonies were picked and screened. **b)** Groups of 4 colonies were pooled and screened for the presence of the 3'HA. **c)** Positive 3'HA groups were subsequently screened as individual colonies. **d)** Isolated clones that were positive for 3'HA integration were further screened for the presence of the 5'HA. The efficiency of targeting was 8.3% (5/60 neo-resistant clones were positive in 5' and 3' HA screens).

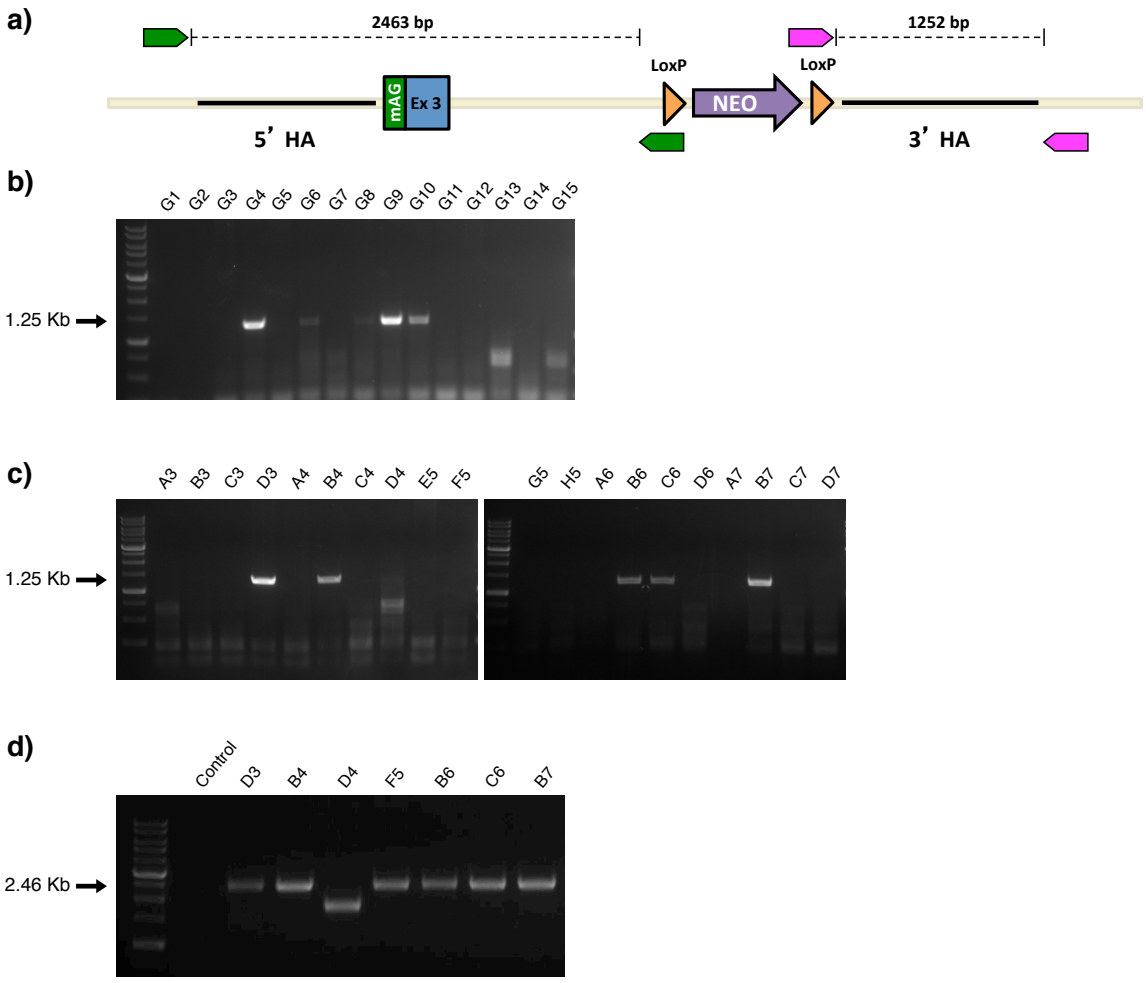

Figure 15: PCR-based screen for successfully targeted mAG-DN-Tcf1
a) Schematic of 5' and 3' homology arm (HA) PCR screens used to isolate targeted clones. Expected band sizes using primers for 5'HA incorporation (green) and 3'HA incorporation (pink) are indicated. TALENs were co-transfected with the targeting vector. Corresponding controls were transfected with the targeting vector and control plasmids lacking the TALEN portion. 80 neo-resistant colonies were picked and screened. **b)** Groups of 4 colonies were pooled and screened for the presence of the 3'HA. **c)** Positive 3'HA groups were subsequently screened as individual colonies. **d)** Isolated clones that were positive for 3'HA integration were further screened for the presence of the 5'HA. The efficiency of targeting was 10% (6/60 neo-resistant clones were positive in 5' and 3' HA screens).

3. Detection of endogenous protein expression of epitope-tagged Tcf1 isoforms

Once potentially targeted clones that were positive based on PCR analysis were isolated, they were further screened for the presence of 3XFLAG via western blotting. Promising clones that were positive for the 5'HA (DN-Tcf1 only), 3'HA, and 3XFLAG were subsequently sub-cloned following a transient transfection of a pCX-Cre-Puro in order to excise the loxP-flanked neomycin cassette. Colonies were picked and replica-plated in two separate 96-well plates with one containing standard mESC media and the other containing mESC media with 250 μg/mL G418. Clones that did not survive in G418 were indicative of having undergone successful excision of the neomycin cassette. The excision was extremely efficient, as more than 80% of clones underwent the removal of the neomycin cassette in both the FL- and DN-Tcf1 targeting experiments. Specifically, at least 25/30 3XFLAG-FL-Tcf1 clones underwent excision, while at least 24/30 clones were successfully excised in the DN-Tcf1 experiments.

Western blot analysis using an α-3XFLAG antibody of neomycin-excised clones revealed an additional band at approximately 65 kDa in FL-Tcf1 clones A3 and D3, which was not observed in any of the DN-Tcf1 clones (Figure 16). Furthermore, it appears as though a null mutant was unintentionally generated in DN-Tcf1 clone C5A, as it was missing the prominent 47kDa band and was not detected by the α-3XFLAG antibody.

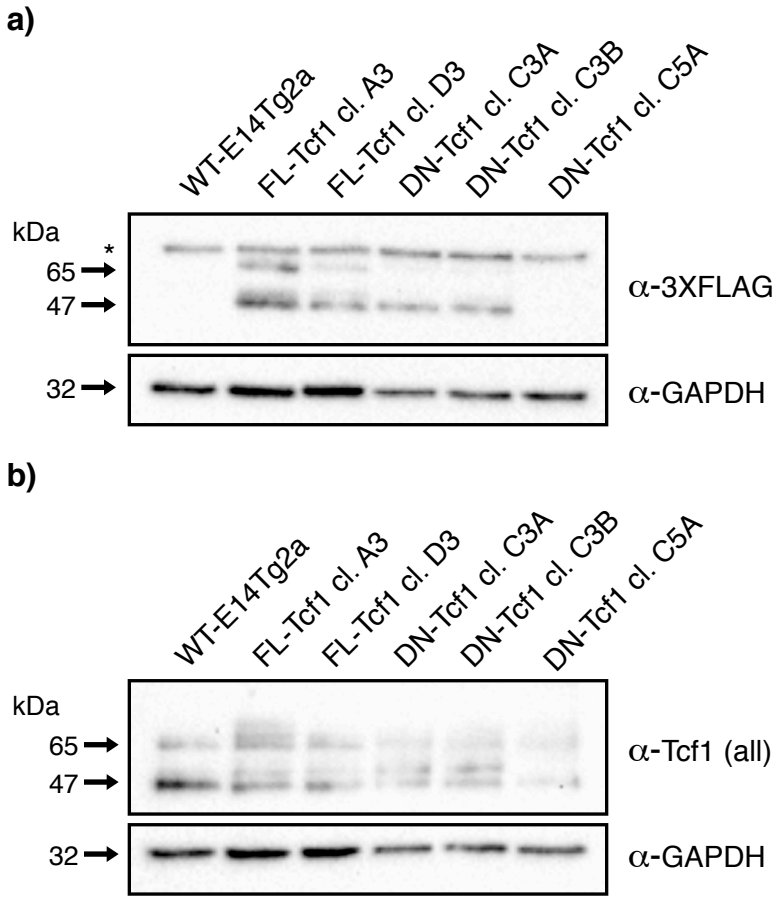

Figure 16: Protein expression levels of endogenous 3XFLAG-tagged Tcf1 isoforms
Both FL-Tcf1 and DN-Tcf1 protein expression was observed at readily detectable levels from whole cell lysates. Note that the DN-Tcf1 clones designated here may include the FL variant with an embedded 3XFLAG (FL-Tcfemb). The band sizes detected with the α-3XFLAG antibody **(a)** closely resembled those seen with the α-Tcf1 antibody **(b)**. An additional band at ~65 kDa was detected in the 3XFLAG-FL-Tcf1 clones. Clone C5A appears to be a null mutant. Non-specific binding is designated by an asterisk.

Similarly, whole cell extracts of mAG-FL-Tcf1 clones that were positive for the 3'HA were used to screen via western blot analysis to observe for the presence of a supershifted band compared to a wild-type control. Indeed, 4 out of 9 clones screened displayed the expected banding pattern (Figure 17). Specifically, an additional band that had migrated

at approximately 75 kDa was detected in clones H2, E4, H4, and C5 but not in the wild-type control.

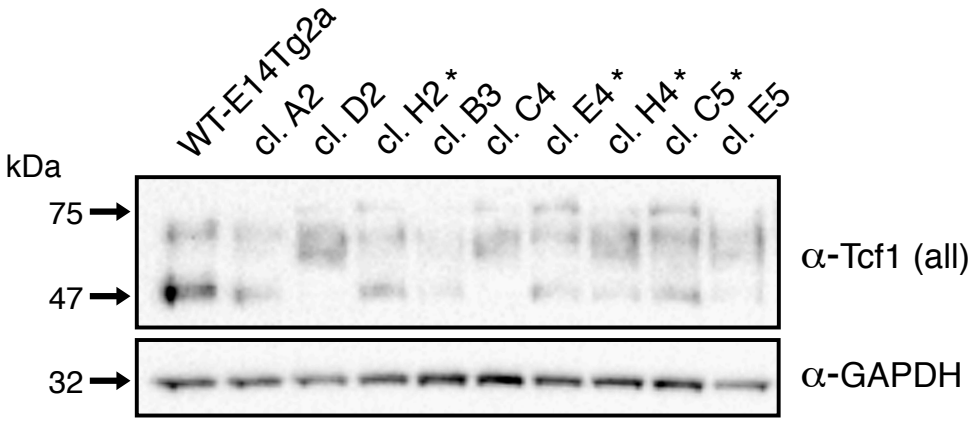

Figure 17: Screen for protein expression for successfully targeted mAG-FL-Tcf1
Several clones that were positive for the 3'HA screen via PCR were lysed and harvested for western blot analysis. A supershifted band at ~75 kDa indicates the presence of mAG tagged FL-Tcf1. Putative correctly targeted clones are designated by an asterisk.

4. Disruption of alternative promoter fails to detect 3XFLAG-DN-Tcf1

Due to the presence of alternative splicing and a second promoter in exon 3 of *mTcf7*, the targeted DN-Tcf1 clones resulted in the generation of a FL variant of Tcf1 with an embedded 3XFLAG tag (here on, referred to as FL-Tcf1emb) at the 5' end of exon 3 in addition to generating an N-terminally tagged DN-Tcf1 (Figure 18a). In an attempt to circumvent this and to eliminate the FL-Tcf1emb variant, the DN-Tcf1 targeting vector was modified to include an insertion mutation within the untranslated region of the DN variant (Figure 18b). The insertion of a nucleotide should cause a frame-shift mutation in the FL variant leading to its nonsense-mediated decay, while at the same time retaining the original and unaltered open reading frame driven by the alternative promoter, to give rise to clones exclusively expressing N-terminally tagged 3XFLAG-DN-Tcf1. However,

western blot analysis on isolated clones revealed an absence of bands following a 3-day EB assay (Figure 18c).

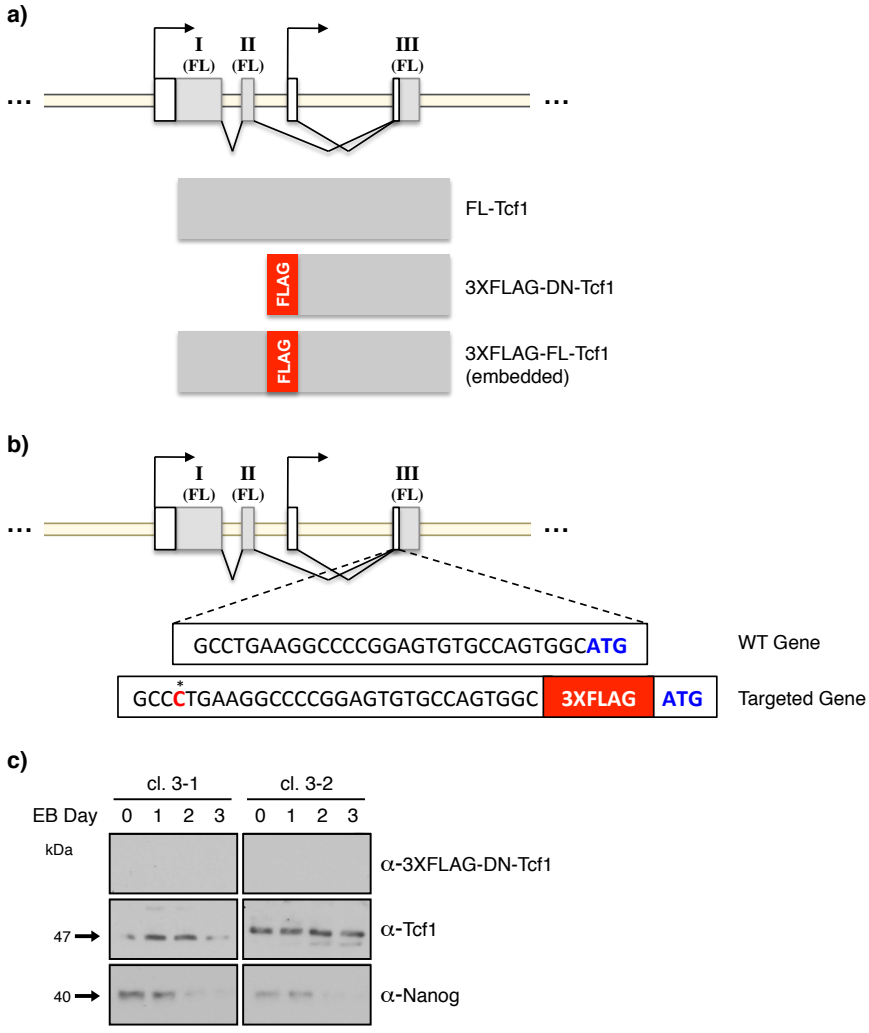

Figure 18: Unsuccessful detection of 3XFLAG-DN-Tcf1 following insertion mutation

a) Schematic representation of alternative splicing and dual promoter usage in *mTcf7* illustrates two possible 3XFLAG-tagged variants that may arise from targeting DN-Tcf1. **b)** A modified targeting vector containing an additional cytosine in the untranslated region of DN-Tcf1 was co-transfected with TALEN monomers in an attempt to disrupt the open reading frame of FL-Tcf1 and to retain that of DN-Tcf1. **c)** Western blot analysis on two isolated clones grown for 3 days in EB media (-LIF, 5% FBS) indicated that the modified strategy yielded no detectable 3XFLAG-DN-Tcf1 under the conditions tested.

5. 3XFLAG-FL-Tcf1 exhibits nuclear localization in differentiating mESCs

In order to determine the intracellular localization of FL-Tcf1 clones, immunofluorescence analysis was carried out for 3XFLAG. Full-length clones A3 and D3 were stained for 3XFLAG and Nanog following 2.5 days in EB media (i.e. -LIF, 5% FBS) (Figure 19). As expected, the presence of 3XFLAG-FL-Tcf1 began to appear on day 2.5 in the nuclear fraction of cells and displayed heterogeneity in its expression. Interestingly, in some cells there appeared to be an inverse relationship between FL-Tcf1 and Nanog expression. In particular, mESCs that display a relatively high level of Nanog expression, also exhibit a low expression of 3XFLAG-FL-Tcf1. Although this pattern did not hold for all cells, it calls into question the role of FL-Tcf1 in positively regulating Nanog levels, which would be expected from the Tcf3/Tcf1 switch model.

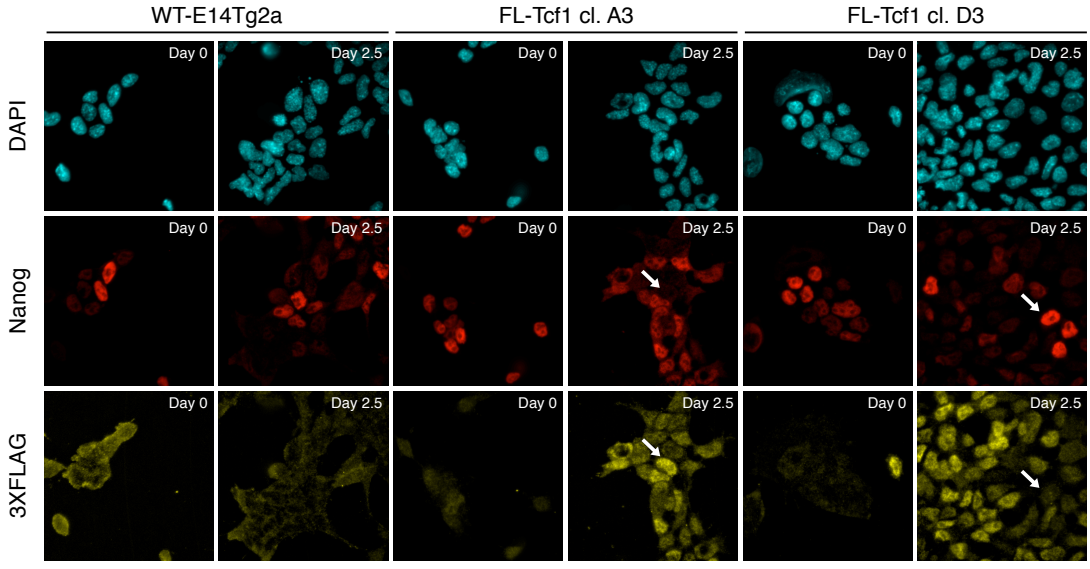

Figure 19: Immunofluorescence staining of 3XFLAG-FL-Tcf1 clones indicates nuclear localization and heterogeneity within cell populations.
Cells were maintained in EB media (-LIF, 5% FBS) for 2.5 days prior to staining with DAPI (nucleus), α-Nanog, and α-3XFLAG (FL-Tcf1). Cells stained at day 0 had been maintained in standard mESC conditions (LIF, 15% FBS) for at least 24 hours. Arrows indicate an inverse relationship between 3XFLAG-FL-Tcf1 and Nanog expression.

CHAPTER 4
DISCUSSION

1. Summary of findings

In this thesis, we establish a working tool to help assess the roles of FL and DN-Tcf1 in mESC self-renewal and differentiation. First, targeting vectors were successfully designed to act as a template for HR-directed repair in order to incorporate either a 3XFLAG or mAG tag at the N-terminus of FL- and DN-Tcf1 variants. In addition, TALEN pairs were strategically engineered so that they would only cause a DSB in the endogenous *mTcf7* locus while leaving the co-transfected donor vector uncleaved (Figures 9, 10). By capitalizing on the relatively short arms of homology (compared to conventional methods) at the 5' and 3' end of the targeting vector, correct HR and endogenous genome modification were validated using simple PCR-based screens (Figures 12-15).

Furthermore, the protein expression of 3XFLAG-FL-Tcf1 and 3XFLAG-DN-Tcf1 were observed at readily detectable levels and were somewhat comparable to wild-type controls based on the α-Tcf1 commercial antibody (Figure 16b). Interestingly, the presence of a ~65 kDa band was observed in the FL but not in the DN clones, suggesting that this additional band could actually be one of the FL-Tcf1 variants that does not co-migrate with a DN-Tcf1 isoform (Figure 16a). In addition, the higher band of the prominent doublet at ~47 kDa was also present in only the FL-Tcf1 clones. Based on these observations, it appears as though the major band at 47 kDa not only corresponds to FL-Tcf1, but to a DN variant that co-migrates with it. Najdi and colleagues (2009) made

a similar observation in normal colon and colon cancer cells, wherein it is possible that a FL-Tcf1 variant with a B tail has the same molecular weight (predicted to be 48 kDa) as a DN-Tcf1 isoform with an E tail[90]. However, it is important to recognize that the 3XFLAG-DN-Tcf1 may actually be a variant of FL-Tcf1 with an embedded 3XFLAG, which may contribute to the ambiguity of the identity of specific bands observed on a western blot.

Due to the possibility that the DN targeting strategy may have resulted in the unintentional generation of the FL-Tcf1emb variant, only the mAG experiment for the FL-Tcf1 strategy was carried forward to western blot analysis (Figure 17). Not only did the potentially positive clones for N-terminally-fused mAG protein display an additional band at ~75 kDa, but their overall protein expression for each Tcf1 band was approximately half as intense as the wild-type control sample. This observation suggests that the predicted positive clones are heterozygotes exhibiting successful targeting on one allele only.

The 3XFLAG-FL-Tcf1 clones A3 and D3 displayed nuclear subcellular localization in differentiating mESCs (Figure 19). In addition to the apparent heterogeneous expression observed, there appeared to be an inverse relationship in some cases between 3XFLAG-FL-Tcf1 expression compared to that of Nanog after 2.5 days in culture conditions that were conducive to differentiation. While it has been previously reported that *Nanog* gene transcription is repressed by Tcf3 in mESCs, this observation may be the first to suggest the existence of a regulatory mechanism between Nanog and FL-Tcf1 protein expression[82].

2. Future Directions

While a working 5'HA and 3'HA screen was established for the DN-Tcf1 targeting strategy, a 5'HA screen could not be implemented for the FL-Tcf1 targeting scheme. This is perhaps due to the observation that the region upstream of the 5'HA of FL-Tcf1 is highly GC-rich. In fact, the first 400 nucleotides upstream of the 5'HA contain a GC content of over 63%. GC-rich DNA sequences are extremely stable because of triple hydrogen bonding (as opposed to two hydrogen bonds in AT interactions) and stabilizing secondary structures. As a result, GC-rich DNA sequences are tightly associated as duplex DNA strands, which make it difficult for polymerases to amplify such regions of the genome. Incidentally, all targeting experiments with the FL-Tcf1 targeting constructs have given rise to bands on FLAG western blots with identical apparent molecular masses, which would not be expected if the fusion products being detected were merely artifacts of random targeting vector integration. To ensure that our clones are not reflective of random integration events, multiple clones that are positive for the 3'HA screen and 3XFLAG/mAG should be screened using Southern blotting with probes directed against the neomycin cassette. It is also likely possible to optimize a 5'HA screen with primers that are located further away from the GC-rich region combined with new PCR reagents that have been developed for difficult GC-rich templates.

Although the bands detected by the α-3XFLAG antibody shed some light on the identity of the Tcf1 isoforms that contain the first coding exon of the *mTcf7* locus, it may be helpful to re-target the endogenous *mTcf7* locus with a targeting vector that would incorporate a C-terminal epitope tag. For instance, either the end of the B tail or E tail

could be tagged with a 3XFLAG epitope. The results obtained from this strategy together with those observed during the course of this thesis might provide a more accurate prediction with regard to the identity of bands observed using a commercially available antibody.

As generating an insertion mutation upstream of the translational start site of DN-Tcf1 disrupted detection of any 3XFLAG-tagged isoform (Figure 18), an alternative strategy should be carried out. Perhaps the insertion mutation modified the alternative promoter in such a way that it led to the altering of mRNA hairpin structures which may be required for translation initiation. Alternatively, it is possible that there really is an absence of DN-Tcf1 expression in mESCs under these conditions. Recently, Kim and colleagues (2013) developed a reporter system (MACS Reporter System) to enrich for cells that have undergone engineered nuclease-induced mutations[154]. The group showed that cells that have taken up the reporter plasmid containing desired TALEN binding sites will express GFP only if successful cleavage occurs. This is due to the fact that the sequence for GFP, located downstream of the TALEN binding sites, is out-of-frame in the absence of engineered nucleases. If a DSB is induced by a set of TALENs, the cell will resort to NHEJ to repair the break. This repair process however, is highly inaccurate and imprecise as it will likely result in a frame-shift mutation that will place the GFP sequence in frame and thus allow the cell to fluoresce green. Therefore, using the established 3XFLAG-FL-Tcf1emb cell lines, a second TALEN pair specific for *mTcf7* exon 1 will induce a DSB to stimulate the cell to initiate NHEJ. In the absence of a targeting vector, this error-prone repair mechanism will likely result in misjoining of severed DNA strands and result in

premature termination of FL-Tcf1 translation. Since the promoter region of the DN variant will still be intact, it is expected that by selecting cells based on their expression of GFP arising from the TALEN-cleaved and an improperly repaired surrogate MACs reporter, the 3XFLAG-DN-Tcf1 can be isolated.

Once the exclusive DN-Tcf1 clones are isolated (i.e. without the embedded 3XFLAG variant), protein expression levels for 3XFLAG and Tcf1 should be assessed to observe for any differences compared to current observations as seen in Figure 16. These clones should also be stained for 3XFLAG as described in the immunofluorescence experiment (Figure 10) and comparisons should be made to wild-type and FL-Tcf1 clones.

Further characterization of epitope-tagged clones will also involve conducting knockdown experiments using shRNA. shRNAs against Tcf1 have previously been shown to be effective, as Yi and colleagues (2011) used an shRNA against exon 8 to demonstrate that endogenous Tcf1 stimulates Wnt target gene activation in a Tcf3-independent manner[66]. In addition, we show here that an shRNA against exon 7 reduces Tcf1 protein expression, as this was observed by the absence of all bands normally detected in wild-type mESCs with commercially available antibodies (Figure 20). Moreover, two different α-Tcf1 antibodies were used for detecting knockdown: 1) α-Tcf1 (all) was produced by immunizing rabbits with a peptide corresponding to a region surrounding Leu158 of human Tcf1, and should recognize all Tcf1 isoforms; 2) α-Tcf1 (-ΔN) should not detect DN variants, as it was generated by immunizing rabbits with a peptide corresponding to a region surrounding Pro95 of human Tcf1 which is found within the N-terminal β-catenin binding domain. However, since both commercially

available antibodies used for Tcf1 seem to be no different from one another, isoform-specific knockdown experiments against various exons should be carried out in order to further help decipher the identity of bands observed.

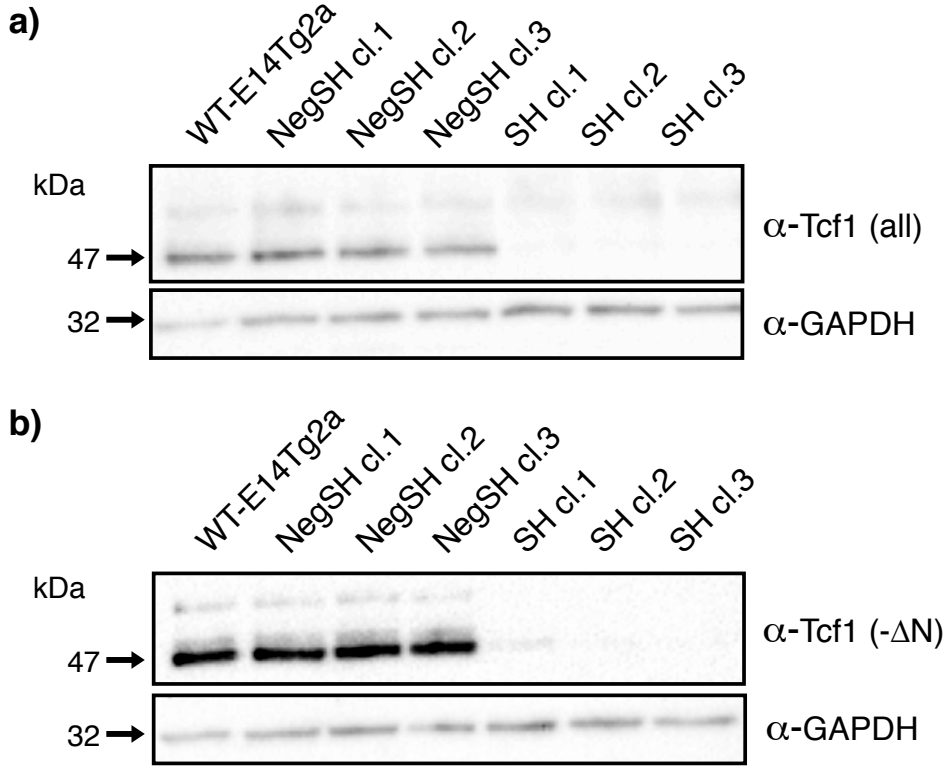

Figure 20: Stable knockdown of Tcf1 protein expression in wild-type mESCs.
Knockdown of Tcf1 was achieved by transfecting an SH construct against exon 7. Stable cell lines were generated by linearizing the construct, selecting in puromycin, and screening for knockdown via western blotting (SH cl. 1-3). A non-targeting control construct was also used in a separate transfection (NegSH cl. 1-3). Knockdown was detected using two different α-Tcf1 antibodies: **(a)** recognizes all isoforms **(b)** does not recognize dominant-negative variants.

It must also be verified that the 3XFLAG- and mAG-tagged Tcf1 isoforms are functionally equivalent to wild-type Tcf1 variants and do not cause obvious effects on the pluripotent nature of the mESCs. To test for alteration of mESC cell state following

genome modification, pilot studies should be conducted to evaluate for defects in pluripotency and differentiation. The transcript levels of several pluripotency and Wnt-target genes should be analyzed in both targeted cell lines. Using qRT-PCR, the expression of *Nanog*, *Oct4*, *Sox2*, *axin2*, *brachyury*, and *cdx1* should be compared to that of wild-type controls. In addition, protein expression levels of these markers should also be carried out using western blot analysis. The changes in both transcript and protein expression levels of the different Tcf/Lefs in the modified cell lines should then be compared to those observed in wild-type mESCs following an EB assay (Figure 4).

In order study the roles of FL and DN isoforms of Tcf1 in mESC biology, several assays may be carried out. Using the endogenously engineered 3XFLAG tag of both FL-Tcf1 and DN-Tcf1, chromatin-immunoprecipitation (ChIP) studies should be performed to determine the binding of these isoforms to promoters of Wnt target genes including *axin2*, *brachyury*, and *cdx1*, as well as to the promoters of those genes that partake in establishing pluripotency (i.e. *Nanog*, *Oct4*, and *Sox2*) in mESCs. Similarly, Tcf1-interacting proteins can be identified by taking advantage of a powerful new technique based on biotinylation. In brief, targeting vectors are generated to contain an N-terminal fusion of Tcf1 with the promiscuous *E. coli* biotin ligase BirA*. Facilitated by TALEN-mediated HR, cells expressing the BirA*-Tcf1 fusion protein are supplemented with biotin, which stimulates BirA* to biotinylate proximal and interacting proteins. The biotinylated proteins may then be purified using affinity purification columns following cell lysis, after which the identity of the purified proteins can be discerned by mass spectrometry. Both ChIP and Co-IP assays should be conducted using mESCs cultured in

different culture conditions: (1) standard mESC media (15% FBS, +LIF), (2) mESC media supplemented with CHIR99021, to allow for accumulation of β-catenin, and (3) EB media (-LIF, 5% FBS), to observe for differences between FL- and DN-Tcf1 isoforms during differentiation.

Lastly, the potentially positive mAG-FL-Tcf1 clones gleaned from Figure 17 must be transiently transfected with Cre recombinase in order to excise the existing neomycin cassette. Once the mAG-DN-Tcf1 have also been generated, both cell lines can be used to study the subcellular localizations of both Tcf1 isoforms during the course of early differentiation using live cell imaging approaches. Furthermore, the established cell lines may then be used to sort heterogeneous populations based on fluorescence intensity. It will be extremely informative to correlate the retention/loss of pluripotency markers with the expression of the specific isoforms of Tcf1.

3. Conclusions

The advent of TALEN technology not only has shown to be effective in generating mouse models of human disease, but it also shows promise as an emerging tool for gene correction studies in clinical trials. The aforementioned experiments and results presented in this thesis provide an extremely useful tool for studying the expression profiles of FL- and DN-Tcf1 isoforms over the course of early development and their role in maintaining pluripotency in mESCs. Upon further characterization of the established cell lines, the specifics behind the proposed Tcf3-Tcf1 switch mechanism may also be delineated. Not only will the methods described here be applicable to Tcf1, but will also increase our understanding of how the other Tcf/Lef transcription factors cooperate in regulating stem

cell self-renewal and differentiation. Due to the varying sensitivities of currently available antibodies, it is difficult to determine the relative expression profiles of the Tcf/Lefs. Thus, the model systems that have been generated along with similar epitope-tagging approaches for the other factors will allow for a more realistic cross-examination of their roles in ESC biology.

REFERENCES

1. Evans, M.J. & Kaufman, M.H. Establishment in culture of pluripotential cells from mouse embryos. *Nature* **292**, 154-6 (1981).
2. Martin, G.R. Isolation of a pluripotent cell line from early mouse embryos cultured in medium conditioned by teratocarcinoma stem cells. *Proc Natl Acad Sci U S A* **78**, 7634-8 (1981).
3. Rodda, D.J. *et al.* Transcriptional regulation of nanog by OCT4 and SOX2. *J Biol Chem* **280**, 24731-7 (2005).
4. Loh, Y.H. *et al.* The Oct4 and Nanog transcription network regulates pluripotency in mouse embryonic stem cells. *Nat Genet* **38**, 431-40 (2006).
5. Boyer, L.A. *et al.* Core transcriptional regulatory circuitry in human embryonic stem cells. *Cell* **122**, 947-56 (2005).
6. Williams, R.L. *et al.* Myeloid leukaemia inhibitory factor maintains the developmental potential of embryonic stem cells. *Nature* **336**, 684-7 (1988).
7. Smith, A.G. *et al.* Inhibition of pluripotential embryonic stem cell differentiation by purified polypeptides. *Nature* **336**, 688-90 (1988).
8. Desbaillets, I., Ziegler, U., Groscurth, P. & Gassmann, M. Embryoid bodies: an in vitro model of mouse embryogenesis. *Exp Physiol* **85**, 645-51 (2000).
9. Itskovitz-Eldor, J. *et al.* Differentiation of human embryonic stem cells into embryoid bodies compromising the three embryonic germ layers. *Mol Med* **6**, 88-95 (2000).
10. Trounson, A. The production and directed differentiation of human embryonic stem cells. *Endocr Rev* **27**, 208-19 (2006).
11. Okita, K., Ichisaka, T. & Yamanaka, S. Generation of germline-competent induced pluripotent stem cells. *Nature* **448**, 313-7 (2007).
12. Nagy, A., Rossant, J., Nagy, R., Abramow-Newerly, W. & Roder, J.C. Derivation of completely cell culture-derived mice from early-passage embryonic stem cells. *Proc Natl Acad Sci U S A* **90**, 8424-8 (1993).
13. Brivanlou, A.H. *et al.* Stem cells. Setting standards for human embryonic stem cells. *Science* **300**, 913-6 (2003).
14. Boulton, T.G., Stahl, N. & Yancopoulos, G.D. Ciliary neurotrophic factor/leukemia inhibitory factor/interleukin 6/oncostatin M family of cytokines induces tyrosine phosphorylation of a common set of proteins overlapping those induced by other cytokines and growth factors. *J Biol Chem* **269**, 11648-55 (1994).
15. Boeuf, H., Hauss, C., Graeve, F.D., Baran, N. & Kedinger, C. Leukemia inhibitory factor-dependent transcriptional activation in embryonic stem cells. *J Cell Biol* **138**, 1207-17 (1997).
16. Burdon, T., Chambers, I., Stracey, C., Niwa, H. & Smith, A. Signaling mechanisms regulating self-renewal and differentiation of pluripotent embryonic stem cells. *Cells Tissues Organs* **165**, 131-43 (1999).

17. Matsuda, T. *et al.* STAT3 activation is sufficient to maintain an undifferentiated state of mouse embryonic stem cells. *EMBO J* **18**, 4261-9 (1999).
18. Niwa, H., Burdon, T., Chambers, I. & Smith, A. Self-renewal of pluripotent embryonic stem cells is mediated via activation of STAT3. *Genes Dev* **12**, 2048-60 (1998).
19. Wilson, S.I. & Edlund, T. Neural induction: toward a unifying mechanism. *Nat Neurosci* **4 Suppl**, 1161-8 (2001).
20. Ying, Q.L., Nichols, J., Chambers, I. & Smith, A. BMP induction of Id proteins suppresses differentiation and sustains embryonic stem cell self-renewal in collaboration with STAT3. *Cell* **115**, 281-92 (2003).
21. Qi, X. *et al.* BMP4 supports self-renewal of embryonic stem cells by inhibiting mitogen-activated protein kinase pathways. *Proc Natl Acad Sci U S A* **101**, 6027-32 (2004).
22. Haegebarth, A. & Clevers, H. Wnt signaling, lgr5, and stem cells in the intestine and skin. *Am J Pathol* **174**, 715-21 (2009).
23. Sato, N., Meijer, L., Skaltsounis, L., Greengard, P. & Brivanlou, A.H. Maintenance of pluripotency in human and mouse embryonic stem cells through activation of Wnt signaling by a pharmacological GSK-3-specific inhibitor. *Nat Med* **10**, 55-63 (2004).
24. Hao, J., Li, T.G., Qi, X., Zhao, D.F. & Zhao, G.Q. WNT/beta-catenin pathway up-regulates Stat3 and converges on LIF to prevent differentiation of mouse embryonic stem cells. *Dev Biol* **290**, 81-91 (2006).
25. Ogawa, K., Nishinakamura, R., Iwamatsu, Y., Shimosato, D. & Niwa, H. Synergistic action of Wnt and LIF in maintaining pluripotency of mouse ES cells. *Biochem Biophys Res Commun* **343**, 159-66 (2006).
26. Nichols, J. & Smith, A. The origin and identity of embryonic stem cells. *Development* **138**, 3-8 (2011).
27. Anton, R., Kestler, H.A. & Kuhl, M. Beta-catenin signaling contributes to stemness and regulates early differentiation in murine embryonic stem cells. *FEBS Lett* **581**, 5247-54 (2007).
28. ten Berge, D. *et al.* Embryonic stem cells require Wnt proteins to prevent differentiation to epiblast stem cells. *Nat Cell Biol* **13**, 1070-5 (2011).
29. Nusse, R. & Varmus, H.E. Many tumors induced by the mouse mammary tumor virus contain a provirus integrated in the same region of the host genome. *Cell* **31**, 99-109 (1982).
30. Rijsewijk, F. *et al.* The Drosophila homolog of the mouse mammary oncogene int-1 is identical to the segment polarity gene wingless. *Cell* **50**, 649-57 (1987).
31. Bhanot, P. *et al.* A new member of the frizzled family from Drosophila functions as a Wingless receptor. *Nature* **382**, 225-30 (1996).
32. Clevers, H. Wnt/beta-catenin signaling in development and disease. *Cell* **127**, 469-80 (2006).
33. Tamai, K. *et al.* A mechanism for Wnt coreceptor activation. *Mol Cell* **13**, 149-56 (2004).

34. Angers, S. & Moon, R.T. Proximal events in Wnt signal transduction. *Nat Rev Mol Cell Biol* **10**, 468-77 (2009).
35. Woodgett, J.R., Davison, M.T. & Cohen, P. The calmodulin-dependent glycogen synthase kinase from rabbit skeletal muscle. Purification, subunit structure and substrate specificity. *Eur J Biochem* **136**, 481-7 (1983).
36. Doble, B.W., Patel, S., Wood, G.A., Kockeritz, L.K. & Woodgett, J.R. Functional redundancy of GSK-3alpha and GSK-3beta in Wnt/beta-catenin signaling shown by using an allelic series of embryonic stem cell lines. *Dev Cell* **12**, 957-71 (2007).
37. Hoeflich, K.P. *et al.* Requirement for glycogen synthase kinase-3beta in cell survival and NF-kappaB activation. *Nature* **406**, 86-90 (2000).
38. Kerkela, R. *et al.* Deletion of GSK-3beta in mice leads to hypertrophic cardiomyopathy secondary to cardiomyoblast hyperproliferation. *J Clin Invest* **118**, 3609-18 (2008).
39. Fiol, C.J., Mahrenholz, A.M., Wang, Y., Roeske, R.W. & Roach, P.J. Formation of protein kinase recognition sites by covalent modification of the substrate. Molecular mechanism for the synergistic action of casein kinase II and glycogen synthase kinase 3. *J Biol Chem* **262**, 14042-8 (1987).
40. Nelson, W.J. Regulation of cell-cell adhesion by the cadherin-catenin complex. *Biochem Soc Trans* **36**, 149-55 (2008).
41. Nelson, W.J. & Nusse, R. Convergence of Wnt, beta-catenin, and cadherin pathways. *Science* **303**, 1483-7 (2004).
42. Riggleman, B., Schedl, P. & Wieschaus, E. Spatial expression of the Drosophila segment polarity gene armadillo is posttranscriptionally regulated by wingless. *Cell* **63**, 549-60 (1990).
43. Fagotto, F., Gluck, U. & Gumbiner, B.M. Nuclear localization signal-independent and importin/karyopherin-independent nuclear import of beta-catenin. *Curr Biol* **8**, 181-90 (1998).
44. Shitashige, M. *et al.* Regulation of Wnt signaling by the nuclear pore complex. *Gastroenterology* **134**, 1961-71, 1971 e1-4 (2008).
45. Sharma, M., Jamieson, C., Johnson, M., Molloy, M.P. & Henderson, B.R. Specific armadillo repeat sequences facilitate beta-catenin nuclear transport in live cells via direct binding to nucleoporins Nup62, Nup153, and RanBP2/Nup358. *J Biol Chem* **287**, 819-31 (2012).
46. Yokoya, F., Imamoto, N., Tachibana, T. & Yoneda, Y. beta-catenin can be transported into the nucleus in a Ran-unassisted manner. *Mol Biol Cell* **10**, 1119-31 (1999).
47. Huber, A.H., Nelson, W.J. & Weis, W.I. Three-dimensional structure of the armadillo repeat region of beta-catenin. *Cell* **90**, 871-82 (1997).
48. Oosterwegel, M. *et al.* Cloning of murine TCF-1, a T cell-specific transcription factor interacting with functional motifs in the CD3-epsilon and T cell receptor alpha enhancers. *J Exp Med* **173**, 1133-42 (1991).

49. van de Wetering, M., Oosterwegel, M., Dooijes, D. & Clevers, H. Identification and cloning of TCF-1, a T lymphocyte-specific transcription factor containing a sequence-specific HMG box. *EMBO J* **10**, 123-32 (1991).
50. Travis, A., Amsterdam, A., Belanger, C. & Grosschedl, R. LEF-1, a gene encoding a lymphoid-specific protein with an HMG domain, regulates T-cell receptor alpha enhancer function [corrected]. *Genes Dev* **5**, 880-94 (1991).
51. Korinek, V. *et al.* Two members of the Tcf family implicated in Wnt/beta-catenin signaling during embryogenesis in the mouse. *Mol Cell Biol* **18**, 1248-56 (1998).
52. Brannon, M., Gomperts, M., Sumoy, L., Moon, R.T. & Kimelman, D. A beta-catenin/XTcf-3 complex binds to the siamois promoter to regulate dorsal axis specification in Xenopus. *Genes Dev* **11**, 2359-70 (1997).
53. van de Wetering, M. *et al.* Armadillo coactivates transcription driven by the product of the Drosophila segment polarity gene dTCF. *Cell* **88**, 789-99 (1997).
54. Brunner, E., Peter, O., Schweizer, L. & Basler, K. pangolin encodes a Lef-1 homologue that acts downstream of Armadillo to transduce the Wingless signal in Drosophila. *Nature* **385**, 829-33 (1997).
55. Giese, K., Kingsley, C., Kirshner, J.R. & Grosschedl, R. Assembly and function of a TCR alpha enhancer complex is dependent on LEF-1-induced DNA bending and multiple protein-protein interactions. *Genes Dev* **9**, 995-1008 (1995).
56. Prieve, M.G., Guttridge, K.L., Munguia, J. & Waterman, M.L. Differential importin-alpha recognition and nuclear transport by nuclear localization signals within the high-mobility-group DNA binding domains of lymphoid enhancer factor 1 and T-cell factor 1. *Mol Cell Biol* **18**, 4819-32 (1998).
57. Arce, L., Yokoyama, N.N. & Waterman, M.L. Diversity of LEF/TCF action in development and disease. *Oncogene* **25**, 7492-504 (2006).
58. Daniels, D.L. & Weis, W.I. Beta-catenin directly displaces Groucho/TLE repressors from Tcf/Lef in Wnt-mediated transcription activation. *Nat Struct Mol Biol* **12**, 364-71 (2005).
59. Atcha, F.A., Munguia, J.E., Li, T.W., Hovanes, K. & Waterman, M.L. A new beta-catenin-dependent activation domain in T cell factor. *J Biol Chem* **278**, 16169-75 (2003).
60. Atcha, F.A. *et al.* A unique DNA binding domain converts T-cell factors into strong Wnt effectors. *Mol Cell Biol* **27**, 8352-63 (2007).
61. van de Wetering, M., Castrop, J., Korinek, V. & Clevers, H. Extensive alternative splicing and dual promoter usage generate Tcf-1 protein isoforms with differential transcription control properties. *Mol Cell Biol* **16**, 745-52 (1996).
62. Weise, A. *et al.* Alternative splicing of Tcf7l2 transcripts generates protein variants with differential promoter-binding and transcriptional activation properties at Wnt/beta-catenin targets. *Nucleic Acids Res* **38**, 1964-81 (2010).
63. Kennell, J.A., O'Leary, E.E., Gummow, B.M., Hammer, G.D. & MacDougald, O.A. T-cell factor 4N (TCF-4N), a novel isoform of mouse TCF-4, synergizes with beta-catenin to coactivate C/EBPalpha and steroidogenic factor 1 transcription factors. *Mol Cell Biol* **23**, 5366-75 (2003).

64. Merrill, B.J. *et al.* Tcf3: a transcriptional regulator of axis induction in the early embryo. *Development* **131**, 263-74 (2004).
65. Reya, T. *et al.* Wnt signaling regulates B lymphocyte proliferation through a LEF-1 dependent mechanism. *Immunity* **13**, 15-24 (2000).
66. Yi, F. *et al.* Opposing effects of Tcf3 and Tcf1 control Wnt stimulation of embryonic stem cell self-renewal. *Nat Cell Biol* **13**, 762-70 (2011).
67. Liu, F., van den Broek, O., Destree, O. & Hoppler, S. Distinct roles for Xenopus Tcf/Lef genes in mediating specific responses to Wnt/beta-catenin signalling in mesoderm development. *Development* **132**, 5375-85 (2005).
68. Tsedensodnom, O. *et al.* Identification of T-cell factor-4 isoforms that contribute to the malignant phenotype of hepatocellular carcinoma cells. *Exp Cell Res* **317**, 920-31 (2011).
69. Ho, R., Papp, B., Hoffman, J.A., Merrill, B.J. & Plath, K. Stage-specific regulation of reprogramming to induced pluripotent stem cells by Wnt signaling and T cell factor proteins. *Cell Rep* **3**, 2113-26 (2013).
70. Nichols, J. & Smith, A. Naive and primed pluripotent states. *Cell Stem Cell* **4**, 487-92 (2009).
71. Ying, Q.L. *et al.* The ground state of embryonic stem cell self-renewal. *Nature* **453**, 519-23 (2008).
72. Aubert, J., Dunstan, H., Chambers, I. & Smith, A. Functional gene screening in embryonic stem cells implicates Wnt antagonism in neural differentiation. *Nat Biotechnol* **20**, 1240-5 (2002).
73. Woodgett, J.R. Molecular cloning and expression of glycogen synthase kinase-3/factor A. *EMBO J* **9**, 2431-8 (1990).
74. Kelly, K.F. *et al.* beta-catenin enhances Oct-4 activity and reinforces pluripotency through a TCF-independent mechanism. *Cell Stem Cell* **8**, 214-27 (2011).
75. Takao, Y., Yokota, T. & Koide, H. Beta-catenin up-regulates Nanog expression through interaction with Oct-3/4 in embryonic stem cells. *Biochem Biophys Res Commun* **353**, 699-705 (2007).
76. Faunes, F. *et al.* A membrane-associated beta-catenin/Oct4 complex correlates with ground-state pluripotency in mouse embryonic stem cells. *Development* **140**, 1171-83 (2013).
77. Haegel, H. *et al.* Lack of beta-catenin affects mouse development at gastrulation. *Development* **121**, 3529-37 (1995).
78. Huelsken, J. *et al.* Requirement for beta-catenin in anterior-posterior axis formation in mice. *J Cell Biol* **148**, 567-78 (2000).
79. Mahendram, S. *et al.* Ectopic gamma-catenin expression partially mimics the effects of stabilized beta-catenin on embryonic stem cell differentiation. *PLoS One* **8**, e65320 (2013).
80. Lyashenko, N. *et al.* Differential requirement for the dual functions of beta-catenin in embryonic stem cell self-renewal and germ layer formation. *Nat Cell Biol* **13**, 753-61 (2011).

81. Wray, J. *et al.* Inhibition of glycogen synthase kinase-3 alleviates Tcf3 repression of the pluripotency network and increases embryonic stem cell resistance to differentiation. *Nat Cell Biol* **13**, 838-45 (2011).
82. Pereira, L., Yi, F. & Merrill, B.J. Repression of Nanog gene transcription by Tcf3 limits embryonic stem cell self-renewal. *Mol Cell Biol* **26**, 7479-91 (2006).
83. Hikasa, H. *et al.* Regulation of TCF3 by Wnt-dependent phosphorylation during vertebrate axis specification. *Dev Cell* **19**, 521-32 (2010).
84. Hikasa, H. & Sokol, S.Y. Phosphorylation of TCF proteins by homeodomain-interacting protein kinase 2. *J Biol Chem* **286**, 12093-100 (2011).
85. Verbeek, S. *et al.* An HMG-box-containing T-cell factor required for thymocyte differentiation. *Nature* **374**, 70-4 (1995).
86. Maier, E. *et al.* Inhibition of suppressive T cell factor 1 (TCF-1) isoforms in naive CD4+ T cells is mediated by IL-4/STAT6 signaling. *J Biol Chem* **286**, 919-28 (2011).
87. Standley, H.J., Destree, O., Kofron, M., Wylie, C. & Heasman, J. Maternal XTcf1 and XTcf4 have distinct roles in regulating Wnt target genes. *Dev Biol* **289**, 318-28 (2006).
88. Bonner, J. *et al.* Proliferation and patterning are mediated independently in the dorsal spinal cord downstream of canonical Wnt signaling. *Dev Biol* **313**, 398-407 (2008).
89. Hovanes, K. *et al.* Beta-catenin-sensitive isoforms of lymphoid enhancer factor-1 are selectively expressed in colon cancer. *Nat Genet* **28**, 53-7 (2001).
90. Najdi, R. *et al.* A Wnt kinase network alters nuclear localization of TCF-1 in colon cancer. *Oncogene* **28**, 4133-46 (2009).
91. Willinger, T. *et al.* Human naive CD8 T cells down-regulate expression of the WNT pathway transcription factors lymphoid enhancer binding factor 1 and transcription factor 7 (T cell factor-1) following antigen encounter in vitro and in vivo. *J Immunol* **176**, 1439-46 (2006).
92. Roose, J. *et al.* Synergy between tumor suppressor APC and the beta-catenin-Tcf4 target Tcf1. *Science* **285**, 1923-6 (1999).
93. Tang, W. *et al.* A genome-wide RNAi screen for Wnt/beta-catenin pathway components identifies unexpected roles for TCF transcription factors in cancer. *Proc Natl Acad Sci U S A* **105**, 9697-702 (2008).
94. Yu, S. *et al.* The TCF-1 and LEF-1 transcription factors have cooperative and opposing roles in T cell development and malignancy. *Immunity* **37**, 813-26 (2012).
95. Tiemessen, M.M. *et al.* The nuclear effector of Wnt-signaling, Tcf1, functions as a T-cell-specific tumor suppressor for development of lymphomas. *PLoS Biol* **10**, e1001430 (2012).
96. Paigen, K. One hundred years of mouse genetics: an intellectual history. I. The classical period (1902-1980). *Genetics* **163**, 1-7 (2003).
97. Paigen, K. One hundred years of mouse genetics: an intellectual history. II. The molecular revolution (1981-2002). *Genetics* **163**, 1227-35 (2003).

98. Smithies, O., Gregg, R.G., Boggs, S.S., Koralewski, M.A. & Kucherlapati, R.S. Insertion of DNA sequences into the human chromosomal beta-globin locus by homologous recombination. *Nature* **317**, 230-4 (1985).
99. Doetschman, T. *et al.* Targetted correction of a mutant HPRT gene in mouse embryonic stem cells. *Nature* **330**, 576-8 (1987).
100. Thomas, K.R. & Capecchi, M.R. Site-directed mutagenesis by gene targeting in mouse embryo-derived stem cells. *Cell* **51**, 503-12 (1987).
101. Vasquez, K.M., Marburger, K., Intody, Z. & Wilson, J.H. Manipulating the mammalian genome by homologous recombination. *Proc Natl Acad Sci U S A* **98**, 8403-10 (2001).
102. Johnson, R.D. & Jasin, M. Double-strand-break-induced homologous recombination in mammalian cells. *Biochem Soc Trans* **29**, 196-201 (2001).
103. Wyman, C., Ristic, D. & Kanaar, R. Homologous recombination-mediated double-strand break repair. *DNA Repair (Amst)* **3**, 827-33 (2004).
104. Santiago, Y. *et al.* Targeted gene knockout in mammalian cells by using engineered zinc-finger nucleases. *Proc Natl Acad Sci U S A* **105**, 5809-14 (2008).
105. Katada, H. & Komiyama, M. Artificial restriction DNA cutters as new tools for gene manipulation. *Chembiochem* **10**, 1279-88 (2009).
106. Hasty, P., Rivera-Perez, J. & Bradley, A. The length of homology required for gene targeting in embryonic stem cells. *Mol Cell Biol* **11**, 5586-91 (1991).
107. Jeggo, P.A., Geuting, V. & Lobrich, M. The role of homologous recombination in radiation-induced double-strand break repair. *Radiother Oncol* **101**, 7-12 (2011).
108. Shrivastav, M., De Haro, L.P. & Nickoloff, J.A. Regulation of DNA double-strand break repair pathway choice. *Cell Res* **18**, 134-47 (2008).
109. Yanagawa, Y. *et al.* Enrichment and efficient screening of ES cells containing a targeted mutation: the use of DT-A gene with the polyadenylation signal as a negative selection maker. *Transgenic Res* **8**, 215-21 (1999).
110. Cheah, S.S. & Behringer, R.R. Contemporary gene targeting strategies for the novice. *Mol Biotechnol* **19**, 297-304 (2001).
111. Shizuya, H. *et al.* Cloning and stable maintenance of 300-kilobase-pair fragments of human DNA in Escherichia coli using an F-factor-based vector. *Proc Natl Acad Sci U S A* **89**, 8794-7 (1992).
112. Yang, Y. & Seed, B. Site-specific gene targeting in mouse embryonic stem cells with intact bacterial artificial chromosomes. *Nat Biotechnol* **21**, 447-51 (2003).
113. Mori, H., Kondo, A., Ohshima, A., Ogura, T. & Hiraga, S. Structure and function of the F plasmid genes essential for partitioning. *J Mol Biol* **192**, 1-15 (1986).
114. Copeland, N.G., Jenkins, N.A. & Court, D.L. Recombineering: a powerful new tool for mouse functional genomics. *Nat Rev Genet* **2**, 769-79 (2001).
115. Murphy, K.C. Use of bacteriophage lambda recombination functions to promote gene replacement in Escherichia coli. *J Bacteriol* **180**, 2063-71 (1998).
116. Yu, D. *et al.* An efficient recombination system for chromosome engineering in Escherichia coli. *Proc Natl Acad Sci U S A* **97**, 5978-83 (2000).

117. Muyrers, J.P., Zhang, Y., Testa, G. & Stewart, A.F. Rapid modification of bacterial artificial chromosomes by ET-recombination. *Nucleic Acids Res* **27**, 1555-7 (1999).
118. Kolodner, R., Hall, S.D. & Luisi-DeLuca, C. Homologous pairing proteins encoded by the Escherichia coli recE and recT genes. *Mol Microbiol* **11**, 23-30 (1994).
119. Zhang, Y., Buchholz, F., Muyrers, J.P. & Stewart, A.F. A new logic for DNA engineering using recombination in Escherichia coli. *Nat Genet* **20**, 123-8 (1998).
120. Narayanan, K., Williamson, R., Zhang, Y., Stewart, A.F. & Ioannou, P.A. Efficient and precise engineering of a 200 kb beta-globin human/bacterial artificial chromosome in E. coli DH10B using an inducible homologous recombination system. *Gene Ther* **6**, 442-7 (1999).
121. Frendewey, D. *et al.* The loss-of-allele assay for ES cell screening and mouse genotyping. *Methods Enzymol* **476**, 295-307 (2010).
122. Jasin, M., Moynahan, M.E. & Richardson, C. Targeted transgenesis. *Proc Natl Acad Sci U S A* **93**, 8804-8 (1996).
123. Edgell, D.R. Selfish DNA: homing endonucleases find a home. *Curr Biol* **19**, R115-7 (2009).
124. Smih, F., Rouet, P., Romanienko, P.J. & Jasin, M. Double-strand breaks at the target locus stimulate gene targeting in embryonic stem cells. *Nucleic Acids Res* **23**, 5012-9 (1995).
125. Jasin, M. Genetic manipulation of genomes with rare-cutting endonucleases. *Trends Genet* **12**, 224-8 (1996).
126. Kim, Y.G., Cha, J. & Chandrasegaran, S. Hybrid restriction enzymes: zinc finger fusions to Fok I cleavage domain. *Proc Natl Acad Sci U S A* **93**, 1156-60 (1996).
127. Carroll, D. Zinc-finger nucleases: a panoramic view. *Curr Gene Ther* **11**, 2-10 (2011).
128. Pavletich, N.P. & Pabo, C.O. Zinc finger-DNA recognition: crystal structure of a Zif268-DNA complex at 2.1 A. *Science* **252**, 809-17 (1991).
129. Porteus, M.H. & Carroll, D. Gene targeting using zinc finger nucleases. *Nat Biotechnol* **23**, 967-73 (2005).
130. Bogdanove, A.J., Schornack, S. & Lahaye, T. TAL effectors: finding plant genes for disease and defense. *Curr Opin Plant Biol* **13**, 394-401 (2010).
131. Moscou, M.J. & Bogdanove, A.J. A simple cipher governs DNA recognition by TAL effectors. *Science* **326**, 1501 (2009).
132. Boch, J. & Bonas, U. Xanthomonas AvrBs3 family-type III effectors: discovery and function. *Annu Rev Phytopathol* **48**, 419-36 (2010).
133. Christian, M. *et al.* Targeting DNA double-strand breaks with TAL effector nucleases. *Genetics* **186**, 757-61 (2010).
134. Miller, J.C. *et al.* A TALE nuclease architecture for efficient genome editing. *Nat Biotechnol* **29**, 143-8 (2011).
135. Li, T. *et al.* TAL nucleases (TALNs): hybrid proteins composed of TAL effectors and FokI DNA-cleavage domain. *Nucleic Acids Res* **39**, 359-72 (2011).

136. Mahfouz, M.M. *et al.* De novo-engineered transcription activator-like effector (TALE) hybrid nuclease with novel DNA binding specificity creates double-strand breaks. *Proc Natl Acad Sci U S A* **108**, 2623-8 (2011).
137. Boch, J. *et al.* Breaking the code of DNA binding specificity of TAL-type III effectors. *Science* **326**, 1509-12 (2009).
138. Engler, C., Gruetzner, R., Kandzia, R. & Marillonnet, S. Golden gate shuffling: a one-pot DNA shuffling method based on type IIs restriction enzymes. *PLoS One* **4**, e5553 (2009).
139. Engler, C., Kandzia, R. & Marillonnet, S. A one pot, one step, precision cloning method with high throughput capability. *PLoS One* **3**, e3647 (2008).
140. Mussolino, C. *et al.* A novel TALE nuclease scaffold enables high genome editing activity in combination with low toxicity. *Nucleic Acids Res* **39**, 9283-93 (2011).
141. Wiedenheft, B., Sternberg, S.H. & Doudna, J.A. RNA-guided genetic silencing systems in bacteria and archaea. *Nature* **482**, 331-8 (2012).
142. Cong, L. *et al.* Multiplex genome engineering using CRISPR/Cas systems. *Science* **339**, 819-23 (2013).
143. Mali, P. *et al.* RNA-guided human genome engineering via Cas9. *Science* **339**, 823-6 (2013).
144. Hwang, W.Y. *et al.* Efficient genome editing in zebrafish using a CRISPR-Cas system. *Nat Biotechnol* **31**, 227-9 (2013).
145. Fu, Y. *et al.* High-frequency off-target mutagenesis induced by CRISPR-Cas nucleases in human cells. *Nat Biotechnol* (2013).
146. Yamamoto, Y., Miura, K. & Komiyama, M. Site-selective hydrolysis of huge DNA by artificial restriction DNA cutter. *Nucleic Acids Symp Ser (Oxf)*, 265-6 (2006).
147. Chen, W., Kitamura, Y., Zhou, J.M., Sumaoka, J. & Komiyama, M. Site-selective DNA hydrolysis by combining Ce(IV)/EDTA with monophosphate-bearing oligonucleotides and enzymatic ligation of the scission fragments. *J Am Chem Soc* **126**, 10285-91 (2004).
148. Liu, P., Jenkins, N.A. & Copeland, N.G. A highly efficient recombineering-based method for generating conditional knockout mutations. *Genome Res* **13**, 476-84 (2003).
149. Doyle, E.L. *et al.* TAL Effector-Nucleotide Targeter (TALE-NT) 2.0: tools for TAL effector design and target prediction. *Nucleic Acids Res* **40**, W117-22 (2012).
150. Cermak, T. *et al.* Efficient design and assembly of custom TALEN and other TAL effector-based constructs for DNA targeting. *Nucleic Acids Res* **39**, e82 (2011).
151. Sanjana, N.E. *et al.* A transcription activator-like effector toolbox for genome engineering. *Nat Protoc* **7**, 171-92 (2012).
152. Karasawa, S., Araki, T., Yamamoto-Hino, M. & Miyawaki, A. A green-emitting fluorescent protein from Galaxeidae coral and its monomeric version for use in fluorescent labeling. *J Biol Chem* **278**, 34167-71 (2003).

153. Meyer, L.R. *et al.* The UCSC Genome Browser database: extensions and updates 2013. *Nucleic Acids Res* **41**, D64-9 (2013).
154. Kim, H., Kim, M.S., Wee, G., Lee, C.I. & Kim, J.S. Magnetic separation and antibiotics selection enable enrichment of cells with ZFN/TALEN-induced mutations. *PLoS One* **8**, e56476 (2013).