Genome Editing and Differentiation of Embryonic Stem Cells

By
Brian R. Shy
B.A., Michigan State University, 2006

THESIS

Submitted as partial fulfillment of the requirements
for the degree of Doctor of Philosophy in Biochemistry and Molecular Genetics
in the Graduate College of the
University of Illinois at Chicago, 2017

Defense Committee:

Bradley Merrill, Chair and Advisor
Elizaveta Benevolenskaya
Yee-Kin Ho
Nissim Hay
Pradip Raychaudhuri
Angela Tyner
Zain Paroo, Biopharmaceutical Science

This thesis is dedicated to my wife, Kimberly, who makes everything possible.

Contribution of Authors

Chapter 1 is an introductory review written by myself that places the addressed research questions in proper context.

Chapter 2 was written by myself and contains a series of my unpublished experiments designed to address the contributions of Tcf/Lef transcription factors to embryonic stem cell differentiation. The final figure (2.9) was generated in collaboration with Matthew MacDougall.

Chapter 3 represents a published manuscript (*Shy, B.R., Wu, C., Khramtsova, G.F., Zhang, J.Y., Olopade, O.I., Goss, K.H., and Merrill, B.J. "Regulation of Tcf7l1 DNA binding and protein stability as principle mechanisms of Wnt/β-catenin signaling". Cell Reports, 2013*.**)** for which myself and Chun-I Wu shared primary authorship. We both contributed equally to included experiments. My research mentor, Brad Merrill, contributed to writing of the manuscript.

Chapter 4 represents a published manuscript (*Shy, B.R., MacDougall M.S., Clarke R., and Merrill, B.J. "Co-incident insertion enables high efficiency genome engineering in mouse embryonic stem cells". Nucleic Acids Res., September 2016.***)** for which I was the primary author and completed the bulk of experiments with some assistance from Matthew MacDougall and Ryan Clarke. My research mentor, Brad Merrill, contributed to writing of the manuscript.

Chapter 5 consists of concluding remarks written by myself that synthesizes the included research and addresses future directions.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF ABBREVIATIONS

CH	CHIR99021 GSK3 Inhibitor
ChIP	Chromatin Immuno-Precipitation
CHIR	CHIR99021 GSK3 Inhibitor
COIN	Co-Insertion
CRISPR	Clustered Regularly Interspaced Short Palindromic Repeats
DHS	DNAse Hypersensitivity Site
DSB	Double-Stranded Break
dsDNA	Double-Stranded DNA
ESC	Embryonic Stem Cells
EpiLC	Epiblast-Like Stem Cells
EpiSC	Epiblast Stem Cells
FBS	Fetal Bovine Serum
HDR	Homology Directed Repair
hESC	Human Embryonic Stem Cells
iPSC	Induced Pluripotent Stem Cells
LIF	Leukemia Inhibitory Factor
mESC	Mouse Embryonic Stem Cells
MMTV	Mouse Mammary Tumor Virus
NHEJ	Non-Homologous End Joining
PD	PD0325901 Mek Inhibitor
PTM	Post-Translational Modification
qPCR	Quantitative Real-Time Polymerase Chain Reaction
sgRNA	Single Guide RNA
ssDNA	Single-Stranded DNA
ssODN	Single-Stranded Oligonucleotide Donor
TALEN	TAL Effector Nuclease
ZFN	Zinc Finger Nuclease

SUMMARY

This work is divided into two parts. The first examines the effects and mechanisms of Wnt signaling in mouse embryonic stem cells (ESC). Wnt signaling has been demonstrated to prevent the differentiation of ESC grown in culture. I focus on the most downstream effectors of this pathway, the Tcf/Lef transcription factors. I find that Tcf/Lefs are not required for the maintenance of the ESC state but that Tcf3, in particular, is required for differentiation. Wnt stimulation prevents differentiation through the inactivation of Tcf3. This process involves removal of Tcf3 from chromatin, export from the nucleus, and subsequent degradation. In addition to being functionally important for the differentiation of ESC, this mechanism appears to be specific to Tcf3 and represents a divergence from the classic picture of Tcf/Lef activation. Inactivation of Tcf3 appears in other phases of development including eyelid closure, the developing forebrain, as well as in aggressive forms of breast cancer, indicating that this mechanism is not unique to ESC.

The second part of this work examines methods for genome editing in ESC. I make use of CRISPR/Cas9-based tools for the generation of a double stranded breaks in genomic DNA and subsequent stimulation of homology directed repair (HDR). Multiple properties of HDR donor DNAs were evaluated including length of homology arms, size of insert, and concentration of DNA, in order to identify optimal parameters for genome editing experiments. Furthermore, I identified a method of co-insertion (COIN) that increases HDR rates by up to 30-fold in both mouse and human ESC. Finally, I have generated a series of optimized reagents that take advantage of the COIN phenomenon to universally improve HDR targeting efficiencies in mouse and human cells, as well as streamline the process of removing selection cassettes from the genome.

Chapter 1: Introduction

Part I: Differentiation of Mouse Embryonic Stem Cells

Naïve and Primed Pluripotent Stem Cells

All mammals begin as a single large cell, the fertilized egg. This cell floats slowly down the fallopian tube, dividing seven times to generate around 128 cells in both mice[1] and humans[2,3]. The process takes 4-5 days at the end of which the embryo will implant in the wall of the uterus. During divisions 5-7, a hollow cavity called the blastocoele forms in the center and the embryo is thereafter referred to as a blastocyst. In the mouse blastocyst, a homogenous group of 10-20 cells (30-50 cells in humans[2]) called the inner cell mass clusters on one side of the blastocoele[4]. This clump of cells represents the entirety of the future adult animal. These stem cells are labeled "pluripotent" because they have the potential to become any of the cell types within an adult. Cells from the inner cell mass can be removed from the blastocyst and either maintained in the pluripotent state or differentiated into a variety of lineages *in vitro*. Alternatively, they can be re-injected into a new blastocyst where they have the capacity to form all the different lineages, including germline, in the resulting chimaeric mice[5]. This property confirms their pluripotency and has made them extremely useful for studying mammalian biology. Their ability to be genetically modified *in vitro* and then reintroduced into embryos is the foundation of modern gene targeting and mammalian genetics[6].

After implantation, cells from the inner cell mass will undergo a number of transcriptional and epigenetic changes that prepare them to respond appropriately to differentiation signals. This process has been called "priming" and thus cells at the end of this stage are referred to as being "primed" for differentiation, while those in earlier stages are called "naïve"[7]. Morphological changes also occur in the embryo at this time. The inner cell mass flattens out into a sheet of cells and wraps around a central cavity to form a cuplike structure called the epiblast.

Pluripotent cells can be isolated and maintained in culture from either the naïve blastocyst or the primed epiblast. Although both sets of cells are taken from the embryo, pluripotent cells were originally isolated from blastocyst stage embryos and so these are historically referred to as embryonic stem cells in mice (mESC). Cells taken from the primed epiblast are instead referred to as epiblast stem cells (EpiSC).

After priming, a group of signaling molecules including WNT3[8], BMP4[9], and Nodal[10] are released from the posterior of the embryo and drive mesendoderm formation from a region called the primitive streak[11]. This begins the specification of the primary germ layers and initial patterning of the embryo in a process called gastrulation. One of the interesting differences between naïve and primed stem cells is how they respond to these differentiation cues. WNT signaling, for example, promotes the self-renewal of naïve cells and inhibits their differentiation[12-16]. Conversely, WNTs stimulate the differentiation of primed cells as seen in the primitive streak during gastrulation[17-20].

LIF, FGF, and WNT signaling pathways

Naïve mESC have the ability to self-renew, a property defined as the capacity of a multipotent cell line to proliferate without sacrificing its differentiation potential. The self-renewal of mESC can be maintained in culture, indefinitely, by the manipulation of a few different signaling pathways. This includes the LIF/STAT3[21-27], FGF/ERK[28], and canonical WNT pathway[28,29].

The secreted peptide Leukemia Inhibitory Factor (LIF) was the first defined factor recognized to support the self-renewal of mESC[25-27]. LIF stimulates the downstream transcription factor STAT3[22-24] which in turn promotes the expression of pluripotency regulator TCFCP2L1[21]. Although effective, LIF alone cannot maintain mESC and supplementation with high levels of fetal bovine serum (FBS) is required for long term culture. FBS is what remains

when cells and clotting factors are removed from blood. It contains a complex mixture of proteins, carbohydrates, lipids and other bioactive molecules that varies tremendously. Different batches of FBS are better or worse at supporting the self-renewal of mESC and this variability stimulated the search for defined factors which could generate the same benefits[5,30]. Through this work it was discovered that FBS could be replaced by either inhibition of the FGF signaling pathway or stimulation of the WNT signaling pathway[14].

FGF4 and the FGFR2 receptor are both expressed at high levels in mESC and the inner cell mass of the blastocyst[31]. Based on findings that FGF4 was required for early embryogenesis, it was proposed to support self-renewal as an autocrine signal[32,33]. Instead, FGF4 -/- mESC had no defects in self-renewal, but exhibited a marked decrease in their ability to differentiate[31]. Later it was shown that FGF4 stimulates the downstream MAP kinase (MAPK) ERK2 to promote the differentiation of mESC and *Erk2* -/- mESC could be maintained with LIF alone[28]. Use of selective chemical inhibitors soon followed. These targeted either the FGF receptor directly (SU5402) or the MAP kinase kinases (MAPKK) MEK1/2 (PD184352 or PD0325901) which lie upstream of ERK2. In combination with LIF, these chemical inhibitors supported the long-term self-renewal of mESC without the addition of FBS, providing the first defined culture conditions for maintenance of pluripotency[14].

Like FGF4, a number of WNT peptides, including WNT3a, are expressed in mESC[29] and the inner cell mass of blastocyst stage embryos[34]. In contrast to FGF signaling, WNT signaling does directly stimulate the self-renewal of mESC and inhibits their differentiation. In the presence of LIF, WNT3a conditioned media[16] or recombinant WNT3a peptides[15] both support the long term maintenance of naïve mESC. Likewise, inhibition of Glycogen Synthase Kinase 3 (GSK3), a negative regulator of the WNT pathway, with the chemical CHIR99021 or genetic ablation of GSK3 also promotes mESC self-renewal[13,14,16,35]. By some accounts this effect is even greater than that of FGF pathway inhibition[29].

The Core Pluripotency Network

Observations that manipulation of the LIF, FGF, and WNT signaling pathways could prevent mESC differentiation led to a defined chemical cocktail that could maintain mESC indefinitely[14]. However, none of these manipulations provide a benefit in isolation. At least two pathways must be targeted together and any combination is sufficient[14]. This indicates an incomplete redundancy between the pathways and suggests an overlapping set of downstream effects.

These three signaling pathways are proposed to converge on a core circuit of transcription factors including OCT4, SOX2, NANOG, and later extended to include others like ESRRB, TCFCP2L1, KLF2, and KLF4[21,36-42]. The transcription factors co-regulate each other's expression and bind to a highly overlapping set of sites within the genome[43-46]. This is often at large multi-enhancer complexes referred to as "super" or "stretch" enhancers, which appear to be key regulatory points for a particular differentiation state[47-49].

Nearly universally, these core transcription factors form a feed forward circuit that stimulates each other's expression to maintain the naïve state. The exception to this is the transcription factor TCF3 (gene name *Tcf7l1*), a component of the WNT signaling pathway. TCF3 binds to almost identical genomic locations as OCT4, SOX2, and NANOG[43,44,46], but appears to be the only member that provides negative feedback, repressing expression of NANOG, ESRRB, and TCFCP2L1 directly and promoting the differentiation of mESC[5,21,37].

Thus, the WNT signaling pathway is entwined with differentiation at multiple stages of early embryogenesis. Wnt stimulation promotes the self-renewal of mESC and is also involved in shutting down the naïve circuit through TCF3. Upon transition to the primed state, WNT stimulation promotes further differentiation to the mesendoderm fate[17-20]. This central role and

apparent complexity make the WNT pathway particularly interesting for studying the mechanisms of differentiation.

The WNT Signaling Pathway and TCF/LEF Transcription Factors

The WNT pathway is highly conserved and frequently responsible for cell fate decisions. In keeping with this role, the WNT pathway exists in all examined multi-cellular organisms but is absent from single celled organisms[50,51]. The mouse WNT1 gene and its *Drosophila* homolog were discovered in parallel[52,53]. WNT1 was shown to be an oncogene in mice, generating mammary tumors when driven by mouse mammary tumor virus (MMTV) integration[53]. In *Drosophila*, it was instead described as a developmental morphogen essential for wing development[52]. This dual role in cancer and development was an accurate prediction for the importance of WNT signaling. From early mesoderm formation[17-20] to the generation and maintenance of the hair follicle[54-56], the WNT pathway is required at every stage of embryonic development and plays a substantial role in the continued renewal of adult tissues[57]. Not surprisingly, mutations in this pathway cause extremely diverse human diseases[50] and are frequent drivers of cancer in the colon, blood, liver, brain, and breast[58].

Although there are variations, the canonical WNT pathway centers on the intracellular protein β-Catenin. β-Catenin is an essential component of adherens junctions, where it anchors the transmembrane protein E-Cadherin to the cytoskeleton through interaction with α-Catenin [59,60]. Free β-Catenin is a potent intracellular signaling molecule, but is usually maintained at low levels by a destruction complex including the scaffold protein AXIN, the CK1 and GSK3 kinases, the tumor suppressor APC, and the E3 ubiquitin ligase β-TRCP. Free β-Catenin is recruited into the complex and phosphorylated first by CK1 and then by GSK3. This phosphorylation stimulates ubiquitination by β-TRCP and β-Catenin is subsequently degraded in the

proteasome. In the presence of extracellular WNT ligands this destruction is inhibited and β-Catenin accumulates in the cell[50].

In mice and humans there are 19 different WNT genes which produce extracellular peptides typically involved in autocrine or paracrine signaling[50]. Their short range of action is secondary to a covalent lipid attachment which prevents their free diffusion and ties them to cell membranes or carrier proteins[61-63]. They bind to a receptor complex on the cell surface which includes any of ten different Frizzled proteins and the LRP5/6 co-receptors[50,64-66]. WNT ligands initiate binding to Frizzled via their lipid moiety, leading to conformational changes that result in phosphorylation of the intracellular tail of LRP[67]. Phosphorylated LRP can bind AXIN and recruit the destruction complex to the cell membrane[57,68]. This leads to exclusion of β-TRCP as well as direct inhibition of GSK3, causing saturation of the complex with phosphorylated β-Catenin[68,69]. Newly made β-Catenin can no longer be degraded and accumulates in the cell. β-Catenin is then translocated to the nucleus by a still unclear mechanism which may involve binding to the FOXM1 transcription factor[70].

β-Catenin lacks a DNA binding domain and must bind to transcription factors, primarily from the TCF/LEF family[71,72], in order to stimulate WNT target genes. There are four TCF/LEF transcription factors in vertebrates[73]. Lymphoid enhancer factor 1 (LEF1, gene name LEF1) and T cell factor 1 (TCF1, gene name *Tcf7*) derive their names from being first identified in T and B cell lineages[74-76]. TCF3 (gene name *Tcf7l1*) and TCF4 (Gene name *Tcf7l2*) followed soon after[77]. β-Catenin binds to the TCF/LEFs at their amino terminus, and they bind a consensus sequence of (A/T)(A/T)CAAAG through an HMG DNA binding domain near their carboxy terminus[78-84]. Both of these regions are highly conserved among the TCF/LEFs[72]. WNT target gene expression is driven by transactivation domains on β-Catenin, so the combination of TCF/LEF and β-Catenin forms a bi-partite activator which binds to DNA through the TCF/LEF and activates transcription via β-Catenin[78-84].

TCF/LEFs are expressed in different combinations in each tissue and developmental stage[85]. Although their abilities to bind β-Catenin and DNA are highly conserved, other regions are variable. These variable regions are responsible for recruiting different co-factors such as the Groucho family of co-repressors in the central region or C-terminal binding protein (CtBP) in the carboxy terminus[51,86,87], and also contain regulatory regions responsible for altering the stability of protein-protein interactions or stability of the TCF/LEFs themselves in response to post-translational modifications (PTM)[73,88]. Thus the TCF/LEFs have both similarities and differences. Single knockout studies produce unique phenotypes and confirm an independent role for each factor. Among the TCF/LEFs, TCF3 ablation alone is embryonic lethal, exhibiting early gastrulation defects that lead to abnormal anterior/posterior axis specification[89]. TCF1 knockout mice are normal apart from defects in thymocyte differentiation[90], TCF4 knockout mice die shortly after birth due to defects in the differentiation of intestinal epithelium[77], and LEF1 knockouts lack teeth, mammary glands, and hair follicles[91]. Double knockout combinations produce more severe phenotypes, which suggests some level of redundancy. TCF1 and LEF1, for example, exhibit a partially overlapping expression pattern which includes the primitive streak[92]. In contrast to the individual knockouts, the TCF1/LEF1 double knockout is embryonic lethal and exhibits a loss of paraxial mesoderm[85].

Analysis of WNT signaling in mESC is complicated by the combined expression of all four TCF/LEFs[93]. As discussed above, WNT signaling promotes the self-renewal of mESC and numerous studies have implicated the TCF/LEFs as central players in this response[5,72]. TCF3, in particular, has been identified in each of three genome-wide loss of function screens for factors which effect mESC differentiation[94-96]. Intriguingly, TCF3 opposes WNT pathway stimulation [12,13,37] while other TCF/LEFs, notably TCF1, stimulate mESC self-renewal and appear to act in a more classical fashion[12]. TCF/LEF independent effects have also been proposed to occur through either GSK3[5] or β-Catenin. β-Catenin can interact with numerous

transcription factors including the SOX family, FOXO proteins, Hypoxia Inducible Factor 1a (HIF1a), and many others[72]. OCT4, in particular, has been proposed to directly interact with β-Catenin to promote self-renewal[97-99], and a β-Catenin/KLF4 complex has been shown to stimulate Telomerase expression for long term maintenance of the mESC genome[100].

Teasing out the roles of individual TCF/LEFs, their combined response, and TCF/LEF independent effects, has been confounded by this complexity. Conclusions are always tempered with caution that some unanticipated combination of factors may be at play. While difficult, understanding the mechanism by which the WNT pathway regulates self-renewal and differentiation may lead to fundamental insights on how cells regulate cell fate decisions. A complete understanding requires a methodical analysis of each combination of factors in isolation. Recent breakthroughs in genome editing, namely CRISPR/Cas9 based tools, make these complex genetic analyses possible.

Part II: Genome Editing

DNA Damage and Mechanisms of Repair

Transmission of accurate genetic information to the next generation is the most important job of a living organism. Without faithful replication of the genome, it will cease to exist. The task of copying the genome is complex in part because of its enormous size. In human cells, a nearly exact copy of three billion base-pairs needs to be made every time the cell divides. A single base change may eliminate any chance of progeny survival, or more rarely confer a selective advantage. Replication of the genome is made even more difficult by the constant bombardment

of DNA with mutagenic stimuli. These can come from internal sources, such as reactive oxygen species (ROS) produced during normal cellular metabolism, or from external sources, such as ultraviolet radiation from the sun. These mutagenic factors change the chemical structure of individual bases and interfere with proper base-pairing. Each lesion must therefore be removed before accurate replication can continue. The chemical alterations produced in each case vary widely and cells have evolved distinct repair pathways to recognize and repair each of these lesions as efficiently as possible. For example, ROS most frequently produce highly reactive base radicals which are corrected by the Base Excision Repair (BER) pathway[101] while ultraviolet radiation typically produces pyrimidine dimers which undergo more elaborate dissection through the nucleotide excision repair (NER) pathway[102].

One of the most severe forms of damage is a double-stranded break (DSB). This can be caused by ionizing radiation or ROS, but in dividing cells most frequently occurs during the replication of unrepaired single-stranded breaks or lesions[103-105]. Whereas other lesions may lead to single base-pair mutations, DSBs are particularly dangerous because they can induce large deletions and chromosome rearrangements. Larger changes are more likely to cause disease, and rearrangements are frequent drivers of oncogenesis through altered regulation of oncogenes and tumor suppressors[106]. Cells have evolved multiple repair pathways to deal with DSBs quickly and efficiently.

Mechanisms of DSB repair can be broken down into two major types depending on whether or not they rely on a template DNA to direct the repair. The template DNA, typically provided by the sister chromatid[107,108], is recognized by having regions of DNA that are homologous to those surrounding the DSB site. Pathways that use this strategy are therefore referred to as homology directed repair (HDR) or homologous recombination (HR). Conversely, repair pathways which do not require any homologous DNA are referred to as Non-Homologous-End Joining (NHEJ) because in addition to not requiring a template, the free ends

are joined together using a DNA Ligase[109]. NHEJ appears to be the most frequently used form of DSB repair in mammalian cells[110-112].

The choice of repair pathway centers on whether the free DNA ends undergo 5'->3' resection to produce long single-stranded DNA (ssDNA) overhangs. The presence of these overhangs stimulates HDR and inhibits NHEJ[113,114]. NHEJ is initiated, instead, by heterodimers of KU70 and KU80 which encircle the free ends and protect them from extensive resection[115-117]. The KU proteins additionally act as a scaffold to recruit other components of the repair complex[117]. One of the initial factors recruited is DNA-Dependent Protein Kinase (DNA-PK)[118]. The exact role of DNA-PK is not clear but its kinase activity is required for NHEJ and is stimulated by combined interaction with the KU proteins and DNA[118,119]. Pairs of DNA-PK are likely involved in stabilizing the gap between free ends and holding them in opposition[120].

Prior to ligation, the free ends are frequently modified by a diverse set of enzymes. Single or multiple bases may be removed by nucleases such as ARTEMIS or APRAXIN, and small extensions may be generated by polymerases of the POL X family including Terminal deoxy Transferase (TdT)[121-125]. These overhangs can produce small regions of homology that stabilize the break prior to closure by DNA Ligase IV (LIG4)[114,126-128]. Due to the end processing that occurs, NHEJ frequently produces small insertions or deletions (indels) around the break site and is thus considered an error-prone repair mechanism[110-112,129].

In contrast to NHEJ, HDR pathways are stimulated by 5'->3' resection[130] of free DNA ends to produce long stretches of ssDNA that are stabilized by strands of Replication Protein A (RPA)[131]. This resection can be mediated by multiple nucleases including Exonuclease I (EXOI), MRE11, and another unidentified nuclease[132-134]. For homologous recombination to proceed, RPA must be replaced by strands of RAD51. RAD51 then mediates strand invasion into double stranded DNA (dsDNA) of the sister chromatid, and scans for homologous DNA sequences[135]. In yeast, approximately 100bp stretches of RAD51 coated ssDNA are required for strand

invasion[136]. Once a homologous region is located, RAD51 dissociates and the 3' end of the ssDNA acts as a primer for DNA synthesis[135]. This recovers any lost sequence and produces regions of overlapping ssDNA. Resolution depends on whether the other end of the break is also captured and can create crossover products where the complementary strands of sister chromatids are exchanged, or non-crossover products where the original pairs are maintained. Single stranded nicks present after resolution are sealed by DNA Ligase I (LIG1)[137].

Due to the lack of mutagenic indel formation, HDR is a higher fidelity repair mechanism and would presumably be the preferred pathway. However, HDR typically requires the presence of a sister chromatid to act as the repair template. Cells have thus adapted mechanisms to promote 5'->3' end resection during phases of the cell cycle when the sister chromatid is present, namely S and G2. This occurs through the protein CtIP which directly stimulates end resection at DSBs[138]. During S/G2, CDK2 interaction leads to phosphorylation of CtIP. This stimulates interaction with BRCA1[139] which then recruits CtIP to DSBs. CtIP can in turn stimulate end resection, leading to increased HDR frequencies during S and G2 phases of the cell cycle[139-143].

Early Gene Targeting

In 1980, Mario Capecchi demonstrated that foreign DNA, injected into the nucleus by micropipette, could be stably integrated into the mammalian genome. The addition of short sequences from the human SV40 virus increased the insertion frequency by more than 100-fold. It was suggested that the presence of degenerate SV40 sequences in the human genome could mediate this effect and therefore that the integration was mediated by HDR[6,144]. Five years later came the first demonstration of legitimate gene targeting, where foreign DNA was inserted into a particular locus rather than randomly in the genome. The β-globin gene was targeted at a low

frequency by including large arms of homology, about 4.6 kb total, flanking the inserted sequence[145]. Two years later, a similar strategy of targeted integration had been performed in mESC[146]. These experiments provided confirmation that homologous recombination machinery could be harnessed for making targeted changes to the genome and provided a blueprint for how to target other regions. However, the widespread use of these technologies were prevented by two factors. First, the low frequency of on-target insertion and second, the relatively high frequency of off-target insertions. The rate of homologous recombination in mammalian cells, even using large arms of homology, is in the range of 1 in 10^5-10^7 while the rate of random integration is around 1 in 10^2-10^4. This produces a 1,000-100,000-fold higher frequency of random integration[6,147-149], requiring that huge numbers of clones be screened to get the desired product.

The ability to target different regions of the genome efficiently was made possible by the introduction of selectable markers. These provided resistance to chemicals such as Neomycin or Puromycin, and were the gold standard for the next 25 years[145-147]. A selection cassette would be placed in the middle of long homology arms and integrants could be selected based on the presence of resistance. This enriches for both on- and off-target insertions, so a number of methods were generated to select against off-target insertions. One such method relies on using an endogenous promoter to drive expression of the resistance gene rather than including one in the cassette. This improves the on:off ratio by as much as 100-fold but requires targeting a region that has an amenable promoter[150]. Negative selection cassettes were also incorporated outside of the homology arms to select against off-target insertions, which are assumed to use non-homologous repair mechanisms such as NHEJ. This negative selection is used to kill cells which incorporate any of the DNA outside of the homology arms. Positive and negative selection (PNS) in combination enriched for positive clones by as much as 2000-fold[147] and paved the way for modern mammalian genetics. Most mammalian biology labs now utilize

knock-out or knock-in mice targeting their gene of interest, and large scale projects have been organized to knockout every known mouse gene[151].

While a great success, gene targeting in mice was expensive and time consuming. The complexity associated with generating large donor constructs, screening and housing large numbers of mice led to costs of $10,000-20,000 and delays in the range of 1 year to generate a single new mutant[152]. In addition, while PNS gave 2,000-fold increases in mESC, its effects were much less substantial in other cell types, giving as little as 2-fold enrichment[153]. mESC further exhibit a relatively high rate of HDR compared to other cell types, potentially because they spend the majority of their time in the S phase of the cell cycle[154]. Thus, the application of gene targeting was pigeon-holed to the generation of mice, with extremely limited use in the many well established experimental cell lines[149].

Stimulation of HDR with targeted DSBs and CRISPR/Cas9 Based Tools

The next breakthroughs in genome editing came from the ability to target DSBs to a specific location in the genome. One of the major limitations of gene targeting was the low frequency of on-target insertions. This was true even with the use of PNS which enrich for rare HDR events, rather than increasing their frequency. As discussed, damage to DNA is quickly recognized in cells and stimulates repair. Therefore, intentionally damaging DNA by generating a DSB could potentially increase the frequency of repair at that site. Proof of principle was provided using the HO and Sce-I homing endonucleases from S. Cerevisiae, which recognize large 18 and 24 bp sequences, respectively. Generation of a DSB with these endonucleases stimulated both NHEJ and HDR in yeast and mammalian cells, increasing gene targeting by more than two orders of magnitude[155-158]. In mouse 3T3 cells, this was accomplished by first integrating the Sce-I recognition site into the genome, then transfecting cells with an expression plasmid for the Sce-I

nuclease[157]. HDR rates of up to 10% were achieved in these cells following Sce-I expression, however, the requirement for first introducing the recognition site by low efficiency HDR defeats the purpose for most applications.

A method of generating a DSB in the unmodified genome was needed, and came nearly ten years later with the introduction of zinc finger nucleases (ZFN)[159]. These engineered proteins are composed of modular zinc finger domains adapted from endogenous transcription factors. Endogenous zinc fingers recognize a variety of 3 bp sequences and can be combined to target larger regions. Fusion to a Fok-I endonuclease provided the first programmable endonucleases that could be used to generate a DSB in a particular location. ZFN and the similar TAL effector nucleases (TALENs) were able to increase targeting efficiencies by several thousand fold in mammalian cells[160-162] and led to the first demonstrated correction of a genetic mutation in human derived induced pluripotent stem cells (iPSC)[163]. Despite the power of these tools, they have not been adopted for widespread use because of the cost and complexity associated with engineering a new protein and validating its activity for each desired DNA target site[164].

A simpler and less expensive alternative came with the introduction of CRISPR/Cas9 based tools for genome editing in 2012. CRISPR (Clustered Regularly Interspaced Short Palindomic Repeats) was first recognized as an unusual genomic sequence at the 3' end of a gene in E. Coli. It was noted that there were five 29bp repeats interspersed with variable 32bp spacer regions[165]. This unusual configuration was unlikely to occur randomly and suggested some function in the cell. The role of CRISPR was not described until twenty years later, when it was shown to be a bacterial adaptive immune system used to digest the genetic material of viral invaders[166]. If a bacterial cell survived a viral invasion, a small piece of the viral genome would be stored for later reference. These fragments represented the 32bp spacers between the

originally described repeats. Upon future infection with the same virus, the fragments could be used as a reference to target and digest the viral genome.

The type II CRISPR system was first adapted as a programmable tool for the generation of DSB by Martin Jinek and colleagues in the lab of Jennifer Doudna[167]. In the endogenous CRISPR system, digestion is performed by a single endonuclease called Cas9. Cas9 is directed to a particular nucleotide sequence by a pair of RNAs that incorporate into the protein. The first RNA is transcribed from the spacers in the CRISPR locus and confers the site specificity. Because a single stranded RNA is generated that is complementary to the viral genome, it is able to direct Cas9 to that location through Watson-Crick base pairing. The second RNA, the trans-activating crRNA (tracrRNA), is a structural component in the final ribonucleoprotein. To simplify the system, a fusion of the two RNAs was engineered to obtain a single guide RNA (sgRNA). Thus, a two component tool was generated: the Cas9 endonuclease and a sgRNA that directs Cas9 to a particular sequence where it then generates a DSB[164,167]. To program Cas9 to target a new site, only the first 20 bp of the sgRNA need be changed. Single plasmids were later generated that contained both Cas9 and an easily modified sgRNA, making the cloning process simple and inexpensive. In addition, no cipher was needed to convert between nucleotide and amino acid sequences as with ZFN or TALENs. Three papers were published a year later showing that this system could be used to generate site-specific DSB in mammalian cells and stimulate repair through both NHEJ and HDR by several orders of magnitude[110,111,168].

The power, ease-of-use, and inexpensive nature of CRISPR/Cas9 based tools has led to their rapid incorporation by the scientific community in only three years. CRISPR/Cas9 apparently works in every organism, from bacteria, to plants, to humans. The list of confirmed applications is exhaustive and, so far, there have been no publications describing a species in which CRISPR is not effective. Nonetheless, there are some concerns with current methods of genome editing. The efficiency of HDR even after the generation of a DSB is still quite low in

many cell types, perhaps most notably in human iPSC[111] and human embryonic stem cells (hESC)[169]. This requires the generation of donor constructs with large homology arms (2-14 kb is typical) and the screening of many clones to obtain the desired outcome. In combination, this increases the time and expense tremendously.

In addition, there are a number of concerns over safety, particularly if genome editing is to be used for cell therapy in humans. While the Cas9/sgRNA ribonucleoprotein is relatively specific for a particular site, it also exhibits off-target digestion at degenerate sequences throughout the genome. A number of methods have been developed that significantly reduce the frequency of off-target breaks[112,168,170], however, the rules governing specificity are still not clear and the position of off-target breaks can't be accurately predicted[171]. The high frequency of off-target integrations with HDR donors has also not been adequately addressed. As many as 1% of cells obtain an off-target integration under typical HDR conditions[6,146-148] and while negative selection can reduce this frequency as much as 10-fold[172], this still leaves many cells with an off-target integration. With the latest generation of advancements, genome editing has broken through to widespread use in the scientific community. Although promising, the use of current techniques for human medicine will require further improvements in safety and increases in targeting efficiency.

Chapter 2: TCF/LEF expression and requirement during mouse embryonic stem cell differentiation

Abstract:

The WNT signaling pathway stimulates the self-renewal of mouse embryonic stem cells (ESC) and prevents their differentiation to an epiblast like cell (EpiLC) *in vitro*. Following this transition, WNTs promote further differentiation to mesendoderm lineages both *in vitro* and *in vivo*. There are four TCF/LEF transcription factors in vertebrates (LEF1, TCF1, TCF3, and TCF4) that may mediate the effects of WNT in addition to potential TCF/LEF independent mechanisms. We hypothesize that one or more TCF/LEF transcription factor is required for the WNT response at each stage of differentiation.

To address the complexity of this system and definitively address the role of TCF/LEFs, we have taken a systematic approach to evaluate and disrupt TCF/LEF expression. We have first examined TCF/LEF expression patterns during EpiLC differentiation and in response to manipulation of the core signaling pathways that maintain ESC self-renewal. Following this we have generated a series of TCF/LEF knockout cell lines using CRISPR/Cas9 based tools to examine requirement. Results indicate that none of the TCF/LEF transcription factors are required for maintenance of embryonic stem cell self-renewal and TCF3 alone was required for differentiation to EpiLC.

Introduction:

The exit of embryonic stem cells from pluripotency begins after the blastocyst stage embryo attaches to the uterine wall to become the epiblast. During this transition epiblast cells undergo a series of transcriptional, epigenetic, and morphological changes that prepare them to respond to stimuli during the process of gastrulation. Gastrulation is initiated by a group of signaling molecules including WNT3[8], BMP4[9], and Nodal[10] that are released from the posterior of the embryo and drive mesendoderm formation from a region called the primitive streak[11]. Cells isolated after implantation and prior to gastrulation are called epiblast stem cells (EpiSC) in reference to their origin. These are distinct from the naïve pluripotent stem cells isolated from the earlier blastocyst stage embryo, which are historically referred to as embryonic stem cells (ESC) in mice.

ESC can be artificially maintained in culture through chemical inhibition or allowed to spontaneously differentiate to an epiblast like state. These cells are referred to as epiblast like cells (EpiLC) if the differentiation is performed *in vitro*, to distinguish them from EpiSC which come directly from the epiblast. Differentiation of ESC can be blocked by inhibition of the FGF signaling pathway with PD0325901, a MEK1/2 inhibitor, stimulation of the WNT pathway with CHIR99021, an inhibitor of GSK3, or stimulation of the LIF/STAT3 pathway through exogenous LIF peptides[14]. The combination of these two inhibitors plus LIF (2iL) is commonly used to maintain ESC *in vitro*.

The WNT pathway is interesting, in particular, because in addition to blocking ESC differentiation to EpiLC, it promotes mesendoderm differentiation once cells reach the EpiLC state[17-20]. This mimics the role of WNT stimulation *in vivo*. WNT3 or β-Catenin knockout embryos fail to generate the primitive streak and completely lack subsequent mesoderm formation[8,173]. The requirement for WNT stimulation at the blastocyst stage is more controversial. WNT peptides are released in the blastocyst, and expression of WNT target

genes such as AXIN2 as well as stimulation of WNT reporter constructs suggest WNT pathway activation[29]. However, β-Catenin knockout embryos appear to progress normally through the blastocyst stage and don't exhibit defects until the start of gastrulation[173]. Similar arguments can be made for the LIF and FGF pathways. LIF knockout mice progress normally to adulthood[174] and FGF knockout mice develop normally until after implantation[33].

A possible explanation for these discrepancies involves the process of diapause, a reproductive strategy in many species, including mice, that allows the pausing of embryogenesis prior to implantation. This is thought to allow pregnancy to better align with abundant food supplies[175]. During diapause, pluripotent stem cells in the blastocyst are prevented from further differentiation and undergo self-renewal. Self-renewal describes the capacity of these cells to proliferate without losing their pluripotency. It has been proposed that *in vitro* cultures of ESC maintained in 2iL represent capture of this diapause state, and that these signaling pathways help maintain diapause[176]. In agreement with this, although LIF is not required for normal embryogenesis, it is required to maintain pluripotent cells during diapause[177].

Whether or not WNT stimulation is required in the blastocyst, its ability to block the EpiLC transition suggests components of this pathway are required for differentiation. The canonical WNT pathway centers on the intracellular protein β-Catenin. In the absence of WNT stimulation, free β-Catenin is recruited into a destruction complex where it is sequentially phosphorylated by the kinases CK1 and GSK3, ubiquitinated by the E3 ligase β-TRCP, and degraded by the proteasome. In the presence of WNT peptides, the destruction of β-Catenin is inhibited. β-Catenin accumulates in the cell, undergoes nuclear translocation, and interacts with transcription factors to stimulate target gene expression[50]. The classic description is that β-Catenin converts the TCF/LEF transcription factors from repressors to activators by virtue of a trans-activation domain on β-Catenin[78-84].

There are four TCF/LEF transcription factors in vertebrates: Lymphoid enhancer factor 1 (LEF1, gene name *Lef1*), T cell factor 1 (TCF1, gene name *Tcf7*), T cell factor 3 (TCF3, gene name *Tcf7l1*) and T cell factor 4 (TCF4, gene name *Tcf7l2*)[73]. All bind to β-Catenin through a highly conserved domain at their amino terminus and bind a consensus sequence of (A/T)(A/T)CAAAG through an HMG DNA binding domain near their carboxy terminus[78-84]. While expression patterns in the defined ESC and EpiLC states have not been described, all four of the TCF/LEFs are expressed in mixed cultures of differentiated and undifferentiated ESC grown in fetal bovine serum (FBS) and LIF[93].

Two of the TCF/LEFs, TCF1 and TCF3, have been directly implicated in regulation of EpiLC differentiation and ESC self-renewal. The evidence for TCF3 is strongest. TCF3 is required *in vivo* and embryos exhibit early gastrulation defects that lead to abnormal anterior/posterior axis specification[89]. It has been identified in a number of genome-wide loss of function screens for factors which effect ESC differentiation[94-96], and shown to oppose WNT pathway stimulation of ESC self-renewal *in vitro*[12,13,37]. TCF1 knockout embryos are grossly normal[90] but evidence suggests that WNT pathway stimulation of ESC self-renewal, *in vitro*, is at least partly mediated by TCF1 based activation[12]. Roles for LEF1 and TCF4 are unknown. There have been further reports suggesting TCF/LEF independent effects of WNT pathway stimulation. OCT4 in particular has been proposed to interact with β-Catenin directly to promote ESC self-renewal[97-99], and a β-Catenin/KLF4 complex has been shown to stimulate Telomerase expression for long term maintenance of the mESC genome[100].

We hypothesize that one or more TCF/LEF transcription factor are required for the WNT promotion of mESC self-renewal, differentiation to EpiLC, and differentiation to mesendoderm. To address the complexity of this system and definitively address the role of TCF/LEFs, we have taken a systematic approach to evaluate and disrupt TCF/LEF expression. We have first examined TCF/LEF expression patterns during EpiLC differentiation and in response to

manipulation of the core signaling pathways that maintain ESC self-renewal. Following this we have generated a series of TCF/LEF knockout cell lines using CRISPR/Cas9 based tools to establish their requirement alone or in combination. Here we find that, surprisingly, none of the TCF/LEF transcription factors are required for maintenance of embryonic stem cell self-renewal. TCF3 alone was required for differentiation to EpiLC and cells lacking TCF3 maintain the mESC in the absence of Wnt signaling.

Results:

***In vitro* differentiation of ESC**

To maintain cells in the naïve ESC state, cells were cultured for greater than five passages in 2iL. This combination has been shown to maintain highly homogenous colonies of pluripotent stem cells resembling the inner cell mass of blastocyst stage embryos[14]. Within 48-72 hours of removing 2iL and adding bFGF to the media, cells differentiate to epiblast-like stem cells (EpiLC)[178,179]. The morphology shifts from round, three-dimensial colonies of closely packed cells in the ESC state, to flat sheets of individual cells in the EpiLC state (Fig 2.1A). RNA was collected from each of these states and analyzed by microarray to identify differentially expressed genes. More than 2,000 genes were identified with greater than 2-fold expression differences (Fig2.1B).

To clarify the changes relevant for differentiation, we focused on a subset of genes that have been previously identified as differentially expressed in these two states and exhibit either some level of functional significance for differentiation, or are commonly used markers for the ESC or EpiLC state (Figure 2.1C)[178,180]. We see elevated expression of ESC genes in 2iL (Fig 2.1C, compare blue bars to genes with blue text) and elevated expression of EpiLC genes after differentiation (Fig 2.1C, compare red bars to genes with red text) suggesting efficient capture of the ESC and EpiLC states.

We next performed a functional assessment of differentiation. After differentiation, the majority of EpiLC are unable to support growth in 2iL, potentially due to a dependence on FGF signaling for survival[5]. Upon replating, the majority of well differentiated EpiLC will not form colonies, in contrast to naïve ESC (Figure2.1D)[181]. Interestingly, although exogenous bFGF is frequently added to stimulate EpiLC differentiation[178,179], it does not appear to be required. Cells differentiated in the absence of bFGF underwent identical morphological changes (not shown)

Figure 2.1 In vitro differentiation of mESC.

(A) Representative colony morphology before and after 72 hours of differentiation.

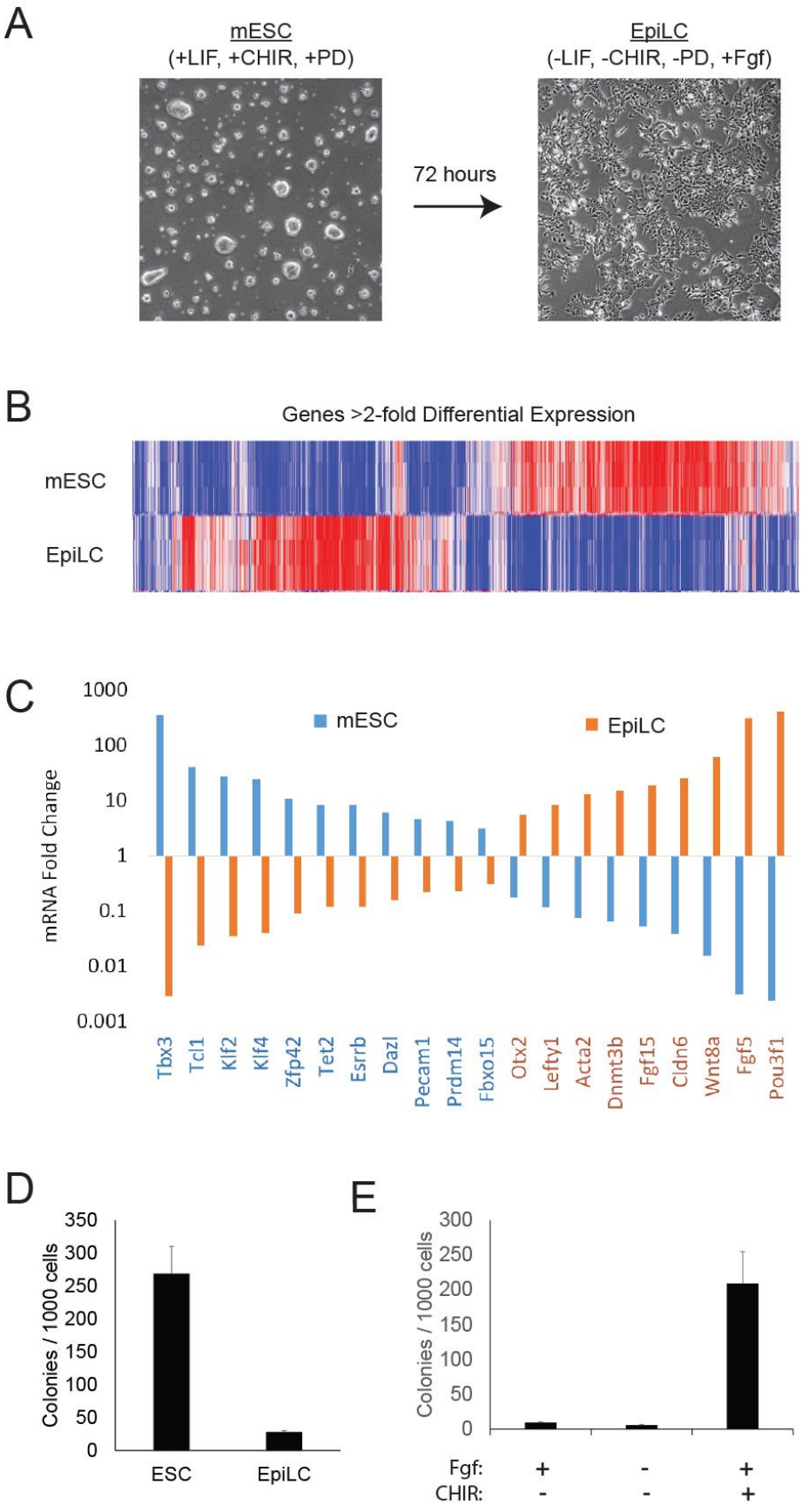
(B) Microarray analysis of WT cells examining genes with > 2-fold increased expression in either the ESC or EpiLC state (red = upregulated, blue = downregulated).

(C) mRNA fold change in log scale for a subset of genes previously associated with the mESC state (blue text) or the EpiLC state (red text). Blue bars represent ESC/EpiLC. Red bars represent EpiLC/ESC.

(D) Replating assay showing the number of colonies that form when cells are replated in naïve culture conditions (2iL). ESC were maintained in 2iL throughout experiment. EpiLC were differentiated for 4 days prior to replating through the removal of 2iL and addition of bFGF.

(E) Replating assay showing the number of colonies that form when cells are replated in naïve culture conditions (2iL). Cells were differentiated for 4 days prior to replating through the removal of 2iL. During this time cells were maintained +/- exogenous bFGF and +/- CHIR99021 as indicated.

*Data for Figure 2.1D was generated by Matthew MacDougall


A
mESC
(+LIF, +CHIR, +PD)
72 hours
EpiLC
(-LIF, -CHIR, -PD, +Fgf)
B
Genes >2-fold Differential Expression
mESC
EpiLC
C
mRNA Fold Change
1000
100
10
1
0.1
0.01
0.001
mESC
EpiLC
Tbx3
Tcl1
Klf2
Klf4
Zfp42
Tet2
Esrrb
Dazl
Pecam1
Prdm14
Fbxo15
Otx2
Lefty1
Acta2
Dnmt3b
Fgf15
Cldn6
Wnt8a
Fgf5
Pou3f1
D
Colonies / 1000 cells
350
300
250
200
150
100
50
0
ESC
EpiLC
E
Colonies / 1000 cells
300
250
200
150
100
50
0
Fgf: + - +
CHIR: - - +

and were unable to generate colonies in 2iL (Figure2.1E). This indifference to exogenous FGF is probably due to the release of endogenous FGF4 and FGF5[5]. In contrast, inhibition of GSK3 with 3 uM CHIR99021 (CHIR) prevents EpiLC morphological changes (Fig 2.2A, bottom row) and maintains a large population of cells capable of forming naïve colonies in 2iL (Fig 2.1E, 2.2B).

Inhibition of the FGF/ERK pathway with PD0325901 (PD) also prevents efficient differentiation. Populations of 2iL competent cells are maintained after differentiation with either 1 or 6 uM PD at approximately the same frequency as CHIR treated cells (Figure 2.2 B). The combination of both PD and CHIR appears more potent but the differences are not statistically significant. The morphological changes associated with EpiLC differentiation are not inhibited by PD at 1 uM, and only slightly at 6 uM. This suggests a separation of the morphological and functional changes associated with differentiation (Fig 2.2A).

Effects of differentiation, WNT pathway stimulation, and FGF pathway inhibition on TCF/LEF expression.

The role of the WNT pathway on ESC self-renewal, differentiation to EpiLC, and further differentiation to mesendoderm lineages prompted us to examine the most downstream effectors in this pathway, the TCF/LEF transcription factors. We initially examined their expression at each stage. ESC cells were maintained for three passages in CHIR plus LIF, and then allowed to differentiate to EpiLC for 72 hours by removing these factors. Changes in TCF/LEF protein expression were examined by western blot in ESC versus EpiLC (Fig 2.3A column 1 vs. 4, Fig 2.3B column 1 vs. 7). Strikingly, LEF1 is not expressed at all in ESC but exhibits substantial expression after EpiLC differentiation. TCF1 is expressed at both stages with slightly lower expression in EpiLC. TCF3 is also expressed at both stages with higher

expression in EpiLC. At the mRNA level, TCF4 has the lowest expression (not shown), however protein is detectable in both ESC and EpiLC.

The effects of the three primary signaling pathways (LIF, WNT, and FGF) on TCF/LEF expression were also examined. Inhibition of FGF/ERK signaling with PD exhibited opposing effects on TCF3 and LEF1 expression (Fig 2.3A, 2.3B). TCF3 levels were increased with increasing levels of PD in both ESC and EpiLC. It is counter-intuitive that TCF3 promotes differentiation[12,13] and is simultaneously upregulated by PD, which inhibits differentiation[14]. This suggests that either a compensatory mechanism exists to inactivate TCF3 or that protein levels of TCF3 do not correlate with its functional activity in this context.

LEF1 is expressed only in EpiLC and shows a striking dependence on FGF/ERK signaling. LEF1 expression is substantially diminished in the presence of PD and is abolished with complete inhibition (Fig 2.3A columns 4-6 and 7-9, Fig 2.3B compare p-Erk1/2 to LEF1 in columns 7-12). This strong correlation suggests that FGF signaling may directly regulate LEF1 expression in EpiLC.

To examine the effects of WNT pathway stimulation, EpiLC cells were treated without or with CHIR for 24 hours (Fig 2.3A, compare columns 4-6 vs. 7-9). In contrast to PD, CHIR led to increased LEF1 and decreased TCF3 protein. In addition, CHIR appears to increase the expression of TCF1. Since CHIR prevents differentiation into EpiLC[14], these expression changes correlate well with data suggesting that TCF1 promotes self-renewal while TCF3 promotes differentiation.

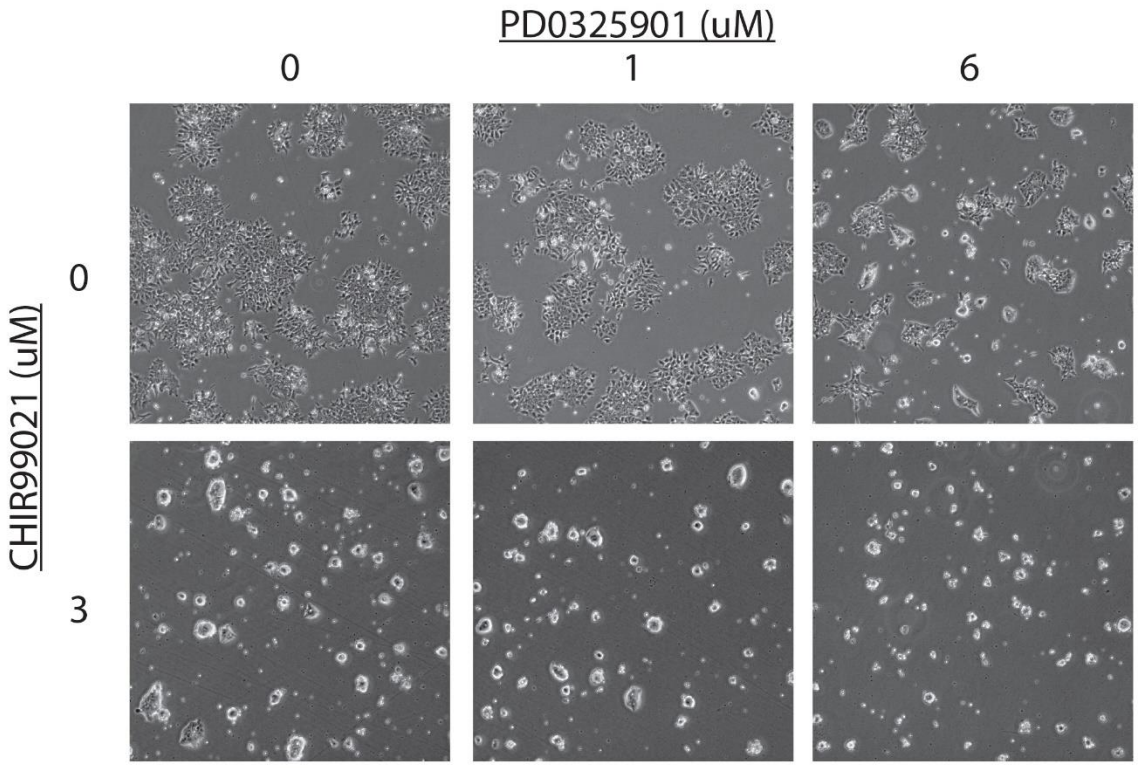
Cells grown in knockout DMEM supplemented with 15% FBS were also examined (Fig 2.3C). These culture conditions were the standard method for maintaining mouse ESC prior to the introduction of 2iL. Although it supports robust growth and some maintenance of naïve ESC, the cultures are heterogenous and contain a variable mixture of undifferentiated and

Figure 2.2 Differentiation to EpiLC is inhibited by PD0325901 (Mek inhibitor) or CHIR99021 (Gsk3 inhibitor).

(A) Representative colony morphology after 72 hours of differentiation in the presence of 0-6 uM PD0325901 (top) and 0-3 uM CHIR99021 (bottom).

(B) Replating assay showing the number of colonies that form in naïve culture conditions after differentiation in the presence of 0-6 uM PD0325901 and 0-3 uM CHIR99021.

A


PD0325901 (uM)
0
1
6
CHIR99021 (uM)
0
3

B

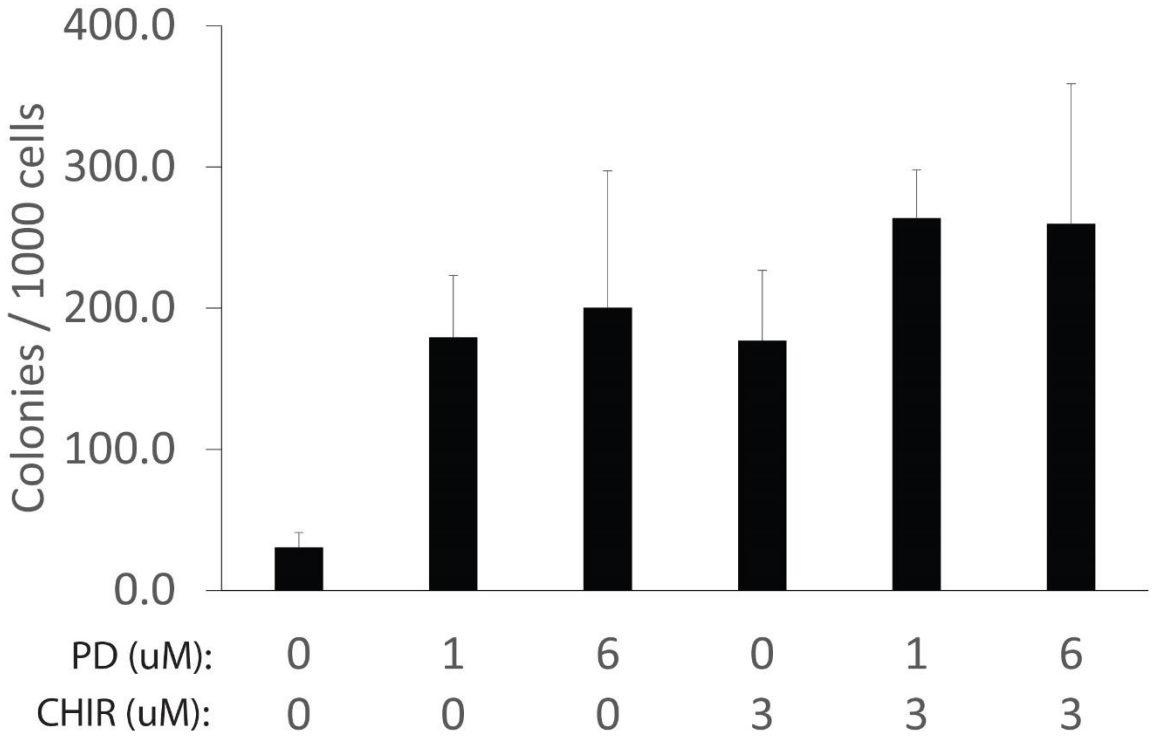
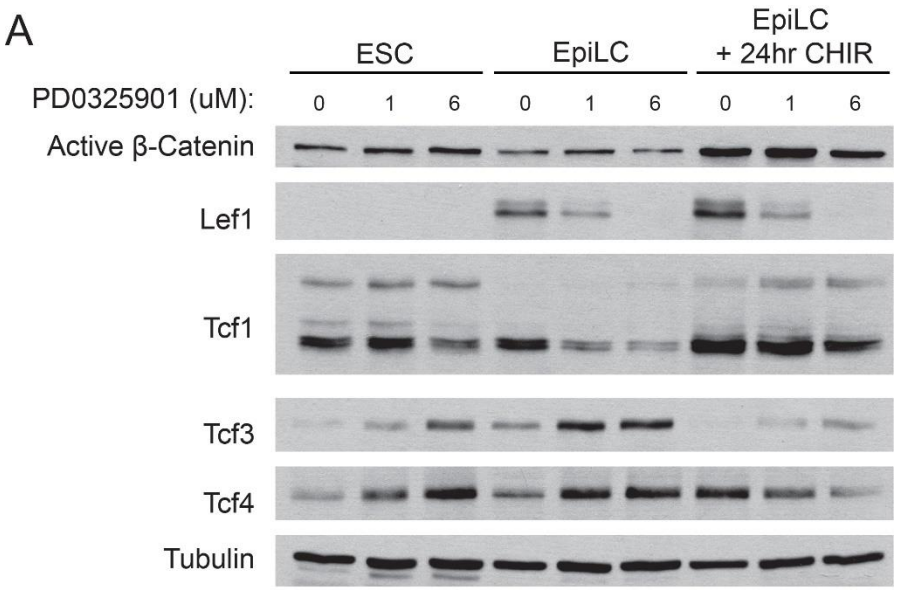

400.0
300.0
200.0
100.0
0.0
Colonies / 1000 cells
PD (uM): 0 1 6 0 1 6
CHIR (uM): 0 0 0 3 3 3

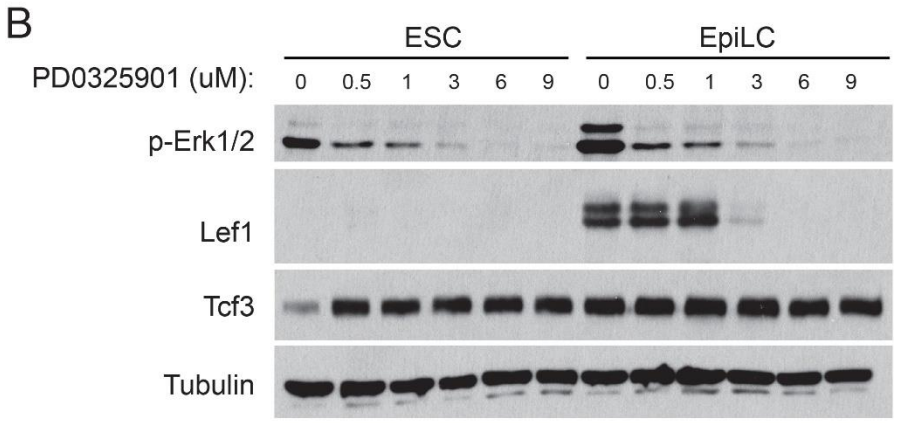
Figure 2.3 Effects of differentiation, FGF signaling, and Wnt signaling on TCF/LEF expression.

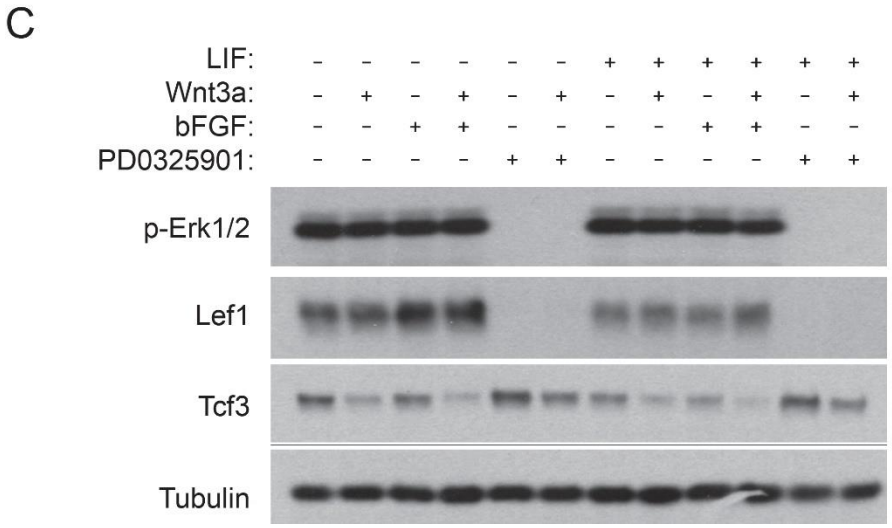
(A) Western blot showing protein levels of active (de-phosphorylated) β-catenin and TCF/LEF transcription factors. Cells were maintained in naive culture conditions (LEFt 3 columns) or differentiated for 72 hours (right 6 columns). Differentiated cells were treated without (middle 3 columns) or with 3 uM CHIR99021 (Gsk3 inhibitor, right 3 columns) for the last 24 hours of differentiation. For each condition cells were exposed tp 0,1, or 6 uM PD0325901 (Mek inhibitor) for the last 24 hours.

(B) Western blot showing protein levels of phosphorylated Erk1 (top band of p-Erk1/2), phosphorylated Erk2 (bottom band of p-Erk1/2), LEF1, and TCF3. Cells were maintained in naive culture conditions (LEFt 6 columns) or differentiated for 72 hours (right 6 columns). Cells were treated with 0-9 uM PD0325901 for the last 24 hours of differentiation.

(C) Western blot showing protein levels of phosphorylated Erk1 (top band of p-Erk1/2), phosphorylated Erk2 (bottom band of p-Erk1/2), LEF1, and TCF3. Cells were maintained in Knockout DMEM + 15% FBS in the presence (right 6 columns) or absence of LIF (LEFt 6 columns) for 72 hours. Cells were treated with Wnt3a, bFGF, or PD0325901 for 24hours prior to collection.


A
ESC
EpiLC
EpiLC + 24hr CHIR
PD0325901 (uM):
0 1 6 0 1 6 0 1 6
Active β-Catenin
Lef1
Tcf1
Tcf3
Tcf4
Tubulin


B
ESC
EpiLC
PD0325901 (uM):
0 0.5 1 3 6 9 0 0.5 1 3 6 9
p-Erk1/2
Lef1
Tcf3
Tubulin


C
LIF:
Wnt3a:
bFGF:
PD0325901:
p-Erk1/2
Lef1
Tcf3
Tubulin

differentiated cells. Cells were additionally grown in presence or absence of LIF for 72 hours (Fig 2.3C, columns 1-6 vs. 7-12). LIF had little impact on the expression of TCF3 and LEF1 expression appears lower in the presence of LIF. The most likely explanation for this is that LIF increases the percentage of cells in the naïve state in these mixed cultures[21-27], and LEF1 is not expressed in naïve cells (Fig 2.3A,B). Treatment of cells with WNT3a for 24 hours gave results comparable to CHIR, substantially decreasing the protein levels of TCF3. Inhibition of FGF/ERK signaling with PD again completely eliminates expression of LEF1 and increases expression of TCF3. Conversely, addition of exogenous bFGF had relatively little effect. This may be because FGF signaling is maintained at a high baseline in these cells due to autocrine and paracrine release of FGF4 and FGF5[5,28].

Generation of TCF/LEF knockout ESC

To systematically evaluate the contributions of each TCF/LEF protein, isogenic knockout cell lines were generated using CRISPR-Cas9 based tools. Cas9 was targeted to each of the TCF/LEFs using a high quality gRNA with more than three mismatches in all potential off-target sites, identified using the CRISPR design tool at crispr.mit.edu[182]. Cas9 generates a double-stranded break (DSB) in the targeted genomic region, which is repaired most frequently by non-homologous end joining (NHEJ). NHEJ is an error-prone repair mechanism, frequently producing small insertions or deletions (indels) which may lead to frameshift mutations in the targeted exon[164]. For each of the TCF/LEFs, DSBs were targeted to an exon in the HMG DNA binding domain, a strategy that has been previously successful for producing each of the single TCF/LEF knockout mice[77,90,91].

Cells were initially grown in knockout DMEM supplemented with 15% FBS and LIF. These conditions provide robust growth, expression of all TCF/LEFs, and can be easily

converted to 2iL. Cells were transfected with plasmids encoding Cas9 and the relevant gRNA. Clonal populations were isolated from single cells plated at low density and grown into colonies. Individual colonies were then picked and expanded for subsequent analysis. Colonies were initially screened by western blot for loss of protein expression. Bi-allelic frameshift mutations in an upstream exon leads to loss of protein primarily through nonsense-mediated decay of mRNA[183]. Analysis by western blot was successful for LEF1, TCF1, and TCF3. TCF4 protein is difficult to detect because of its low expression and TCF4 western blots exhibit a high level of background which confounds analysis. TCF4 was thus analyzed first by Sanger sequencing. Clones with bi-allelic frameshifts were identified at the DNA level and then loss of protein was confirmed by western blot. For each TCF/LEF, bi-allelic frameshifts and loss of protein were obtained in slightly less than 40% of clones (data not shown).

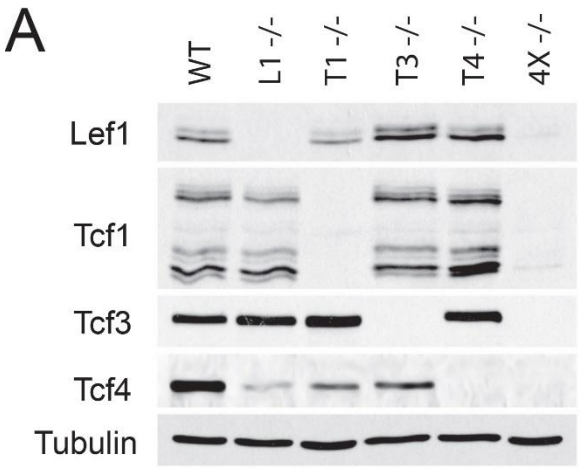
Using this method, isogenic mouse ESC cell lines were successfully generated with all sixteen possible combinations of TCF/LEFs. Western blots are shown for each of the single knockouts (Fig 2.4A), double knockouts (Fig 2.4B), triple knockouts (Fig 2.4C), and the quadruple knockout (4X, Fig 2.4A).

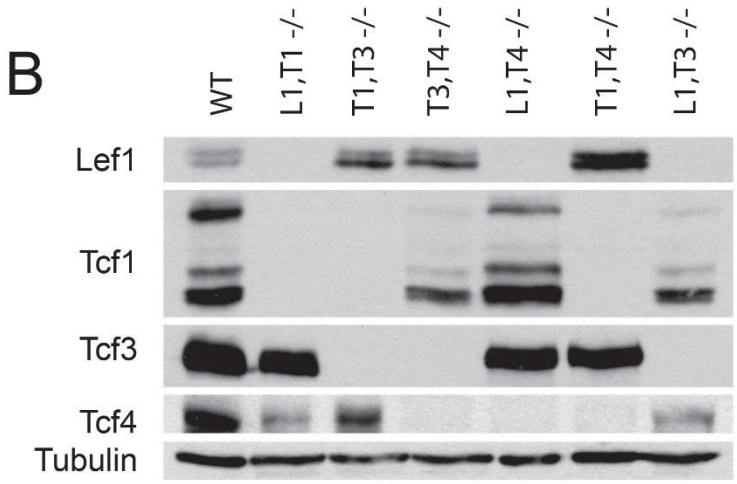
Characterization of WNT pathway stimulation in 4X -/- mESC

In 15% FBS plus LIF, 4X cell lines formed numerous spherical colonies with ESC morphology (Fig 2.5C), but were more prone to differentiation as compared to WT (Fig 2.6E). Regulation of β-Catenin was unaffected by the absence of TCF/LEF proteins. WT and 4X cell lines exhibited identical levels of total β-Catenin (Fig 2.5A), and both responded to GSK3 inhibition by increasing the level of un-phosphorylated (active) β-Catenin (Fig 2.5B). Upon stimulation with CHIR, β-Catenin was appropriately trafficked to the nucleus in both WT and 4X cell lines, indicating that TCF/LEFs are not required for nuclear import (Fig 2.5C).

Figure 2.4 Generation of all TCF/LEF knockout combinations in mESC.

Western blots showing protein levels of TCF/LEFs in (A) single knockout cell lines and the 4X knockout (far right), (B) double knockout cell lines, and (C) triple knockout cell lines. All cell lines were generated and isolated from knockout DMEM media supplemented with 15% FBS and LIF.


A
WT
L1 -/-
T1 -/-
T3 -/-
T4 -/-
4X -/-
Lef1
Tcf1
Tcf3
Tcf4
Tubulin


B
WT
L1,T1 -/-
T1,T3 -/-
T3,T4 -/-
L1,T4 -/-
T1,T4 -/-
L1,T3 -/-
Lef1
Tcf1
Tcf3
Tcf4
Tubulin

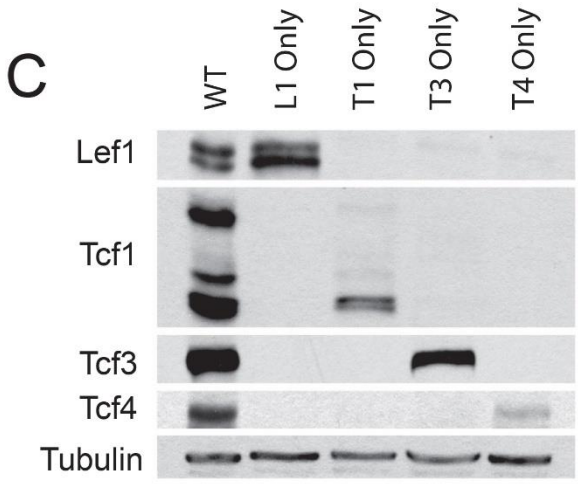

C
WT
L1 Only
T1 Only
T3 Only
T4 Only
Lef1
Tcf1
Tcf3
Tcf4
Tubulin

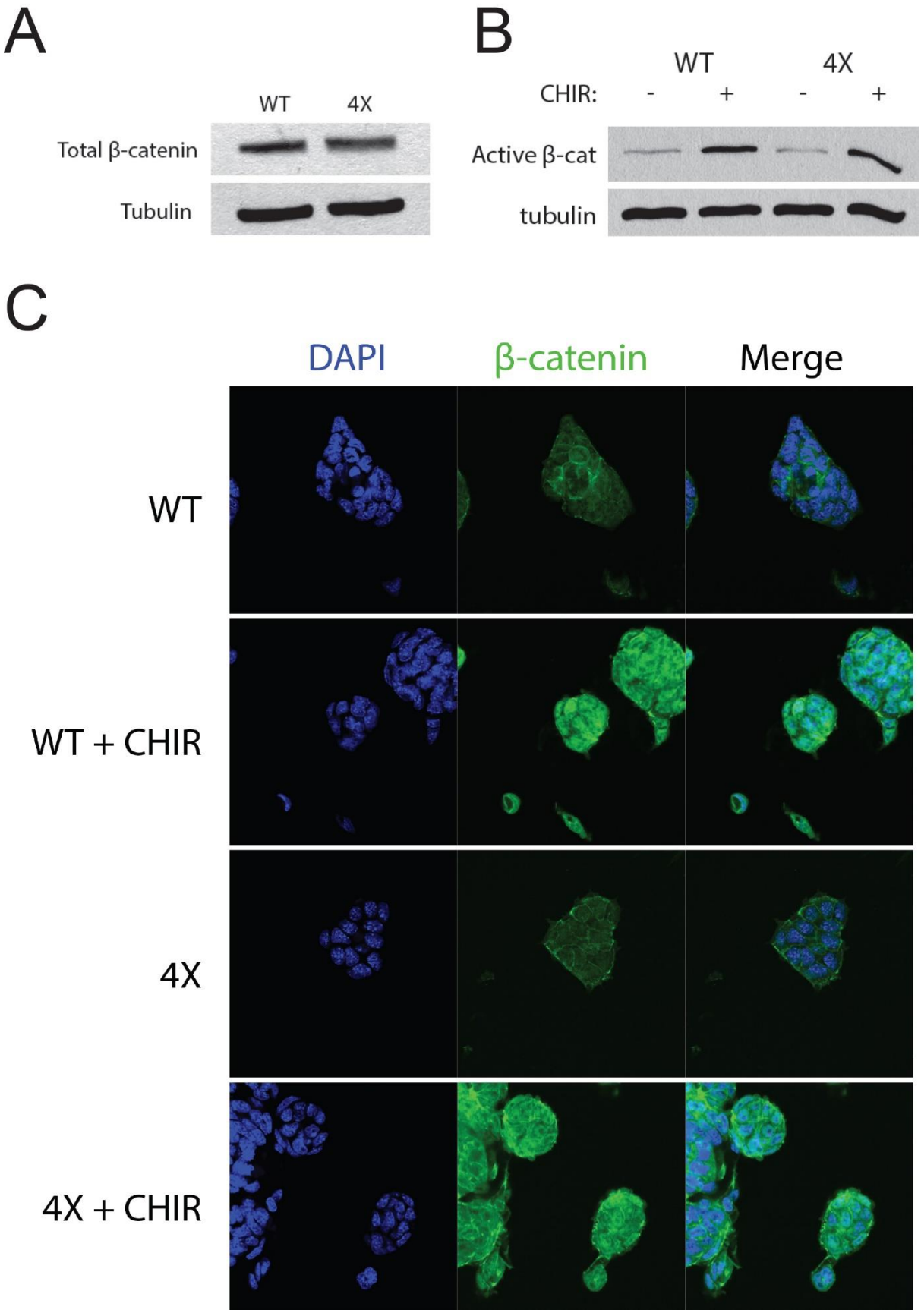
Figure 2.5 Normal regulation of β-catenin in WT and 4X TCF/LEF knockout cell lines.

Comparison of WT and 4X -/- cell lines grown in knockout DMEM supplemented with 15% FBS and LIF.

(A) Western blots showing protein levels of total β-catenin.

(B) Western blots showing protein levels of active β-catenin (un-phosphorylated).

(C) Immunofluorescence showing nuclear (DAPI, blue) and β-catenin (green) localization in colonies of mESC. Both WT and 4X cell lines exhibit increased nuclear β-catenin following treatment with 6 uM CHIR99021 for 24 hours.


A
WT
4X
Total β-catenin
Tubulin
B
WT
4X
CHIR:
- + - +
Active β-cat
tubulin
C
DAPI
β-catenin
Merge
WT
WT + CHIR
4X
4X + CHIR

Within the nucleus, the response to CHIR was markedly different. Whereas nuclear β-Catenin is recruited to enhancers of WNT target genes *Cdx1* and *Axin2* following CHIR treatment in WT cells, this effect is abolished in the 4X cell line (Fig 2.6A). This suggests that TCF/LEFs mediate this recruitment and subsequent activation of target genes. In agreement with this, increased *Cdx1* or *Axin2* mRNA is not stimulated by CHIR in the 4X cell lines (Fig 2.6C). Similarly, 4X cell lines are unable to stimulate a WNT reporter construct, TOP-Flash, which contains multiple TCF/LEF binding sites driving expression of luciferase. In contrast, WT cells exhibit substantial luciferase expression in response to CHIR and knockout of TCF3 alone increases this further (Fig 2.6B).

Intriguingly, some genes do maintain WNT responsiveness. A notable example is *Esrrb*, which has been identified as one of the core transcription factors regulating ESC pluripotency[5,37]. It is expressed exclusively in naïve ESC and was shown to be directly repressed by TCF3 during EpiLC differentiation. In defined culture conditions such as 2iL, overexpression of ESRRB can replace the requirement for WNT pathway stimulation in the maintenance of ESC self-renewal[37]. Baseline ESRRB expression is lower in 4X cell lines as compared to WT, which may be secondary to the higher frequency of differentiation in 4X cultures (Fig 2.6C). Nonetheless, treatment with CHIR causes a significant increase in ESRRB expression in the 4X cell lines. CHIR also stimulates an increase in β-Catenin binding at a central ESRRB enhancer indicating TCF/LEF independent recruitment (Fig 2.6D).

To provide a functional analysis of differentiation we performed an alkaline phosphatase assay. As described, colonies grown in 15% FBS plus LIF exhibit a range of differentiation. This can be quantified by a combination of morphology and alkaline phosphatase staining, a marker present at high levels in naïve ESC. Colonies are counted as undifferentiated if they exhibit a spherical morphology and stain darkly for alkaline phosphatase (Fig 2.6E). Using this criteria the percentage of undifferentiated versus differentiated colonies can be quantified (Fig 2.6F). The

percentage of undifferentiated colonies is substantially higher in WT cell lines treated with CHIR. In agreement with ESRRB expression changes, this pattern is also true in 4X cell lines. Although they have a lower baseline of undifferentiated colonies, there is still a substantial increase in response to CHIR. This indicates a TCF/LEF independent inhibition of differentiation in these conditions.

TCF/LEF requirement in defined culture conditions

To examine TCF/LEF activity in homogenous populations of naïve ESC, knockout cell lines were converted to 2iL culture conditions and maintained for greater than 5 passages. RNA was then collected from WT, T3 -/-, 4X -/-, and T3 Only cell lines for microarray analysis. These were compared to *in vitro* differentiated WT EpiLC as described in figure 2.1B. A large-scale analysis of differentially expressed genes indicates relatively little difference between any of the knockout cell lines maintained in 2iL (Fig 2.7A ESC) as compared to WT EpiLC. There are some notable exceptions for T3-/- and T3 only cell lines (see far left and far right in Fig 2.7A). However, looking at the subset of differentiation associated genes described in figure 2.1C, there are no significant differences between any of the knockout cell lines grown in 2iL (Fig 2.7B). In addition, the knockout cell lines appear morphologically identical (Fig 2.7C). Altogether, these results indicate that TCF/LEF expression is not required for maintenance of the naïve ESC state.

In contrast, there is a substantial shift during EpiLC differentiation. Gene expression changes are similar between either of the cell lines with TCF3 (WT and T3 Only) both globally (Fig 2.8A) and in the subset of differentiation genes (Fig 2.8B). However, gene expression changes are significantly altered in cell lines lacking TCF3 (T3 -/- and 4X -/-). Cell lines lacking TCF3 are defective in down-regulation of all ESC specific genes as well as up-regulation of many EpiLC specific genes (Fig 2.8C). Morphological changes also show a requirement for

Figure 2.6 Downstream Wnt response in 4X knockout mESC.

Comparison of WT and 4X -/- cell lines grown in knockout DMEM supplemented with 15% FBS and LIF.

(A) Chromatin immuno-precipitation for β-catenin at negative control site and enhancers for Cdx1 and Axin2 genes. WT and 4X cell lines were treated with 3uM CHIR99021 for 24 hours.

(B) TOPFlash reporter assay in WT, T3 -/-, and 4X -/- cell lines treated with 3uM CHIR99021 for 24 hours.

(C) mRNA expression measured by qRT-PCR for Cdx1 (LEFt), Axin2 (middle), and Esrrb (right). All were normalized to GAPDH expression. WT and 4X cell lines were treated -/+ 3 uM CHIR for 24 hours.

(D) Chromatin immuno-precipitation for β-catenin at Esrrb enhancer. WT and 4X cell lines were treated with 3uM CHIR99021 for 24 hours.

(E) Colony morphology and alkaline phosphatase (AP) staining of representative undifferentiated (spherical, dark AP staining) and differentiated ESC colonies (flat morphology, light AP staining).

(F) AP assay for WT, T3 -/-, and 4X -/- cell lines treated -/+ 3uM for 5 days.

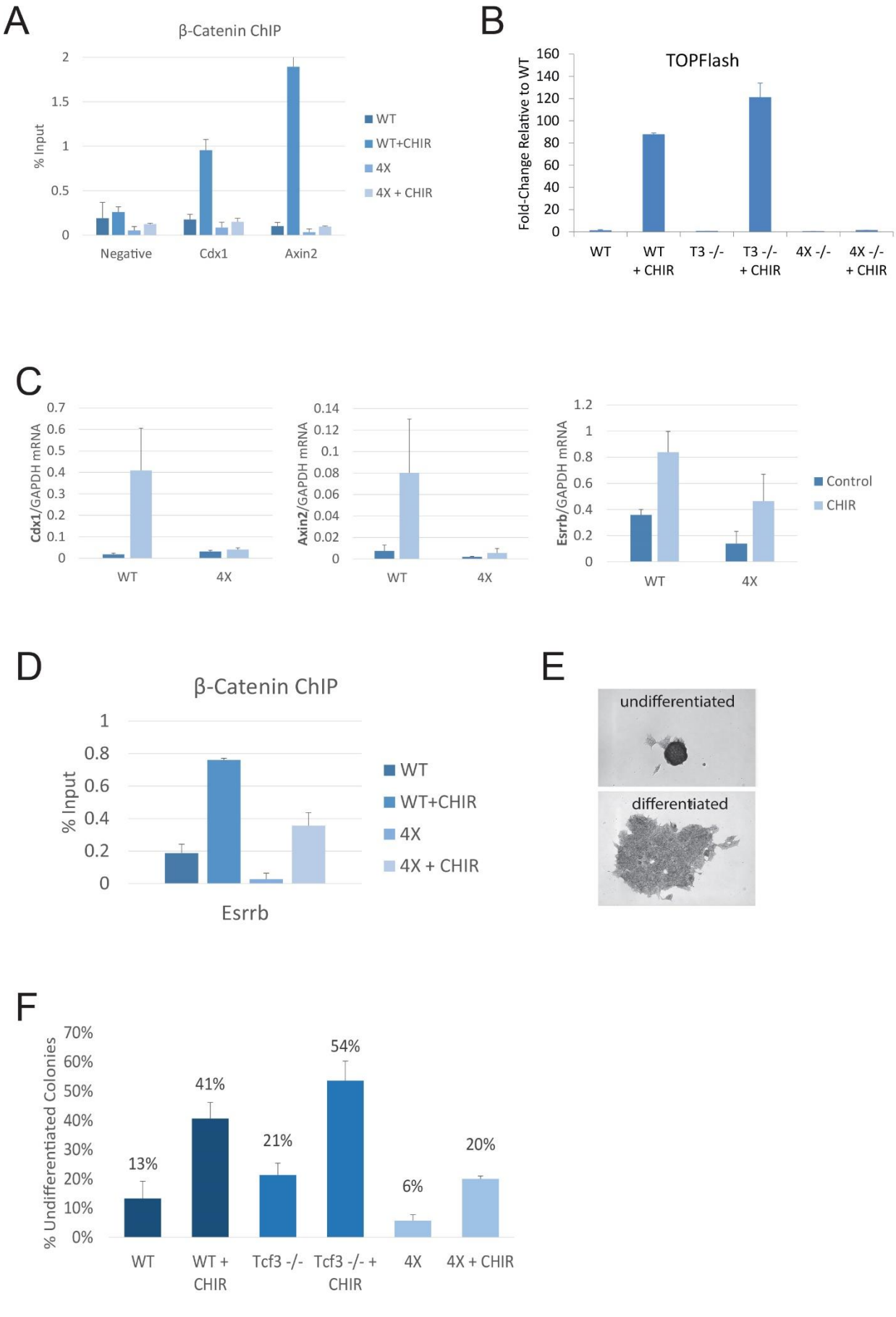

A
β-Catenin ChIP
% Input
2
1.5
1
0.5
0
Negative
Cdx1
Axin2
WT
WT+CHIR
4X
4X + CHIR
B
TOPFlash
Fold-Change Relative to WT
160
140
120
100
80
60
40
20
0
WT
WT + CHIR
T3 -/-
T3 -/- + CHIR
4X -/-
4X -/- + CHIR
C
Cdx1/GAPDH mRNA
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
WT
4X
Axin2/GAPDH mRNA
0.14
0.12
0.1
0.08
0.06
0.04
0.02
0
WT
4X
Esrrb/GAPDH mRNA
1.2
1
0.8
0.6
0.4
0.2
0
WT
4X
Control
CHIR
D
β-Catenin ChIP
% Input
1
0.8
0.6
0.4
0.2
0
Esrrb
WT
WT+CHIR
4X
4X + CHIR
E
undifferentiated
differentiated
F
% Undifferentiated Colonies
70%
60%
50%
40%
30%
20%
10%
0%
13%
41%
21%
54%
6%
20%
WT
WT + CHIR
Tcf3 -/-
Tcf3 -/- + CHIR
4X
4X + CHIR

Figure 2.7 TCF/LEFs are not required for maintenance of naïve state

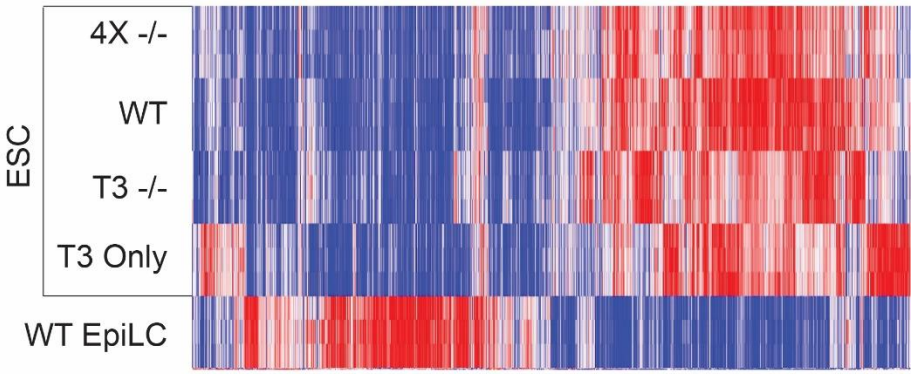
Comparison of WT and TCF/LEF knockout cell lines grown in naïve culture conditions (DMEM supplemented with CHIR99021 + PD0325901 + LIF).

(A) Microarray analysis of WT and TCF/LEF knockout cell lines examining genes with > 2-fold increased expression in either the ESC or EpiLC state (red = upregulated, blue = downregulated). Expression levels are shown in naïve culture conditions for all four cell lines (ESC, top 4 rows) or after 72 hours of EpiLC culture conditions for WT cells (bottom row).

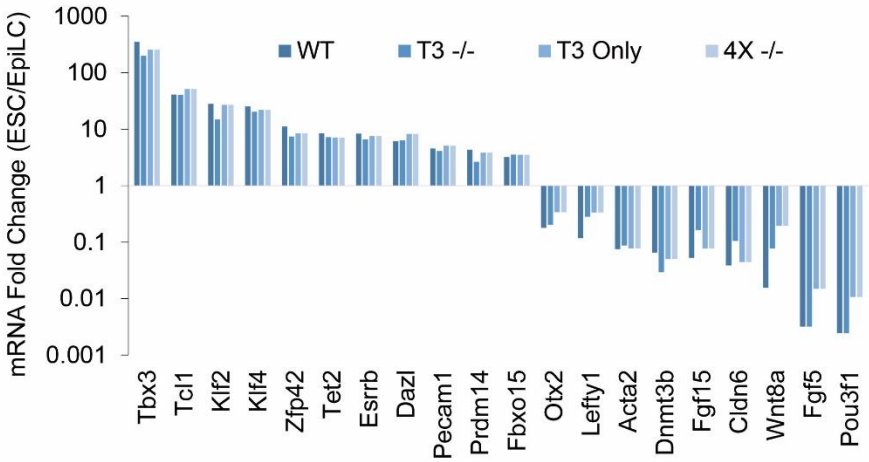
(B) mRNA fold change in log scale for a subset of genes previously associated with the mESC state (LEFt side: Tbx3-Fbxo15) or the EpiLC state (right side: Otx2-Pou3f1). Bars = ESC expression for each cell line / WT EpiLC expression.

(C) Representative colony morphology for each cell line in naïve culture conditions.

A


4X -/-
WT
T3 -/-
T3 Only
ESC
WT EpiLC

B


mRNA Fold Change (ESC/EpiLC)
1000
100
10
1
0.1
0.01
0.001
WT
T3 -/-
T3 Only
4X -/-
Tbx3
Tcl1
Klf2
Klf4
Zfp42
Tet2
Esrrb
Dazl
Pecam1
Prdm14
Fbxo15
Otx2
Lefty1
Acta2
Dnmt3b
Fgf15
Cldn6
Wnt8a
Fgf5
Pou3f1

C

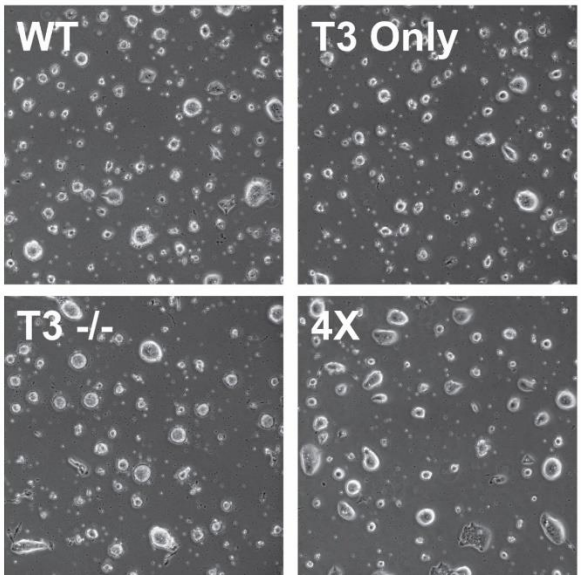

WT
T3 Only
T3 -/-
4X

Figure 2.8 TCF3 is required for EpiLC differentiation

Comparison of WT and TCF/LEF knockout cell lines after 72 hours of EpiLC differentiation (DMEM supplemented with bFGF and without CHIR99021, PD0325901, or LIF).

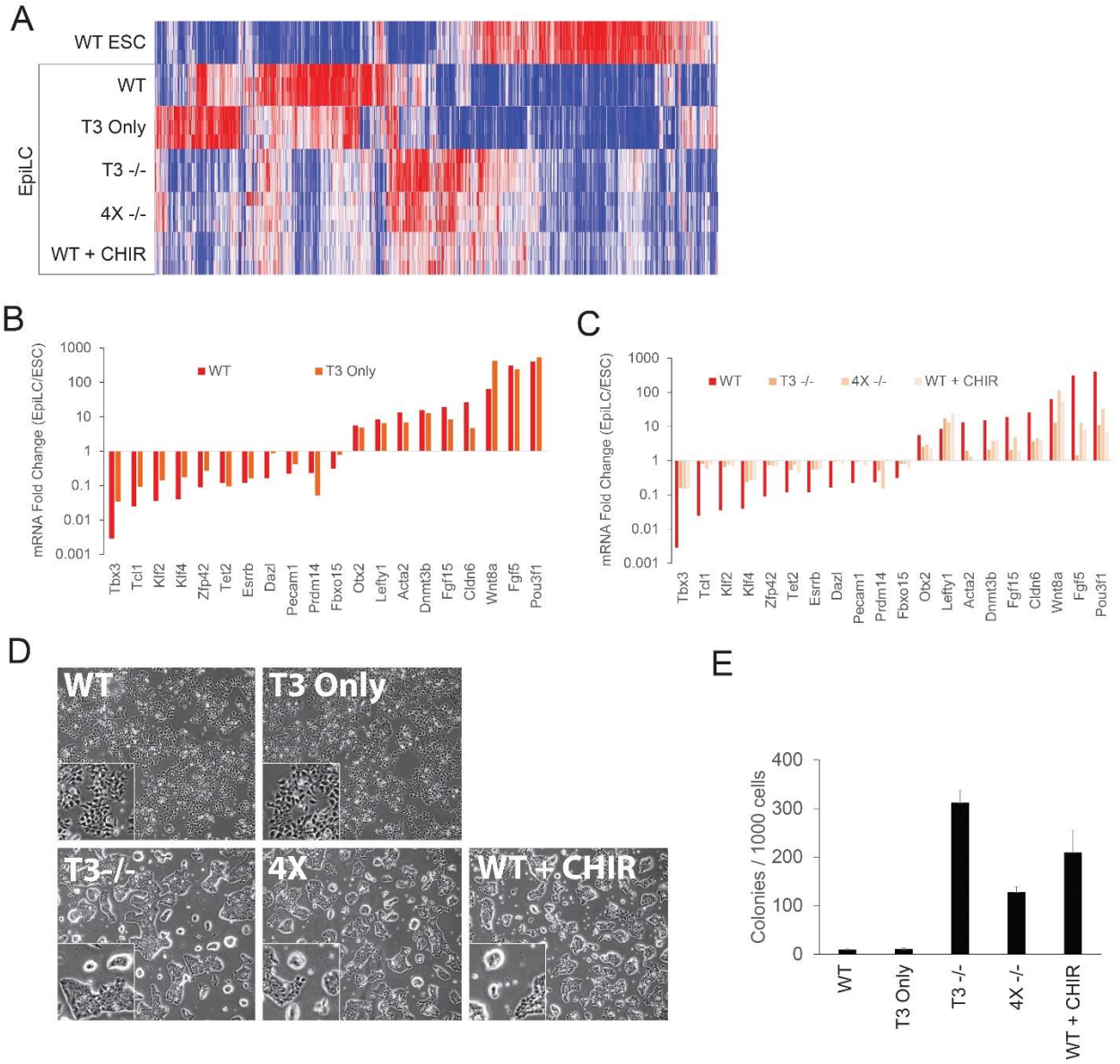
(A) Microarray analysis of WT and TCF/LEF knockout cell lines examining genes with > 2-fold increased expression in either the ESC or EpiLC state (red = upregulated, blue = downregulated). Expression levels are shown in for WT cells naïve culture conditions (WT ESC) or for all four cell lines after 72 hours of EpiLC differentiation. WT cells maintained in 3 uM CHIR99021 during EpiLC differentiation are also shown (bottom row).

(B) mRNA fold change in log scale for a subset of genes previously associated with the mESC state (LEFt side: Tbx3-Fbxo15) or the EpiLC state (right side: Otx2-Pou3f1). Bars = EpiLC expression for each cell line / WT ESC expression.

(C) mRNA fold change in log scale for a subset of genes previously associated with the mESC state or the EpiLC state. Bars = EpiLC expression for each cell line / WT ESC expression.

(D) Representative colony morphology for each cell line after EpiLC differentiation.

(E) Replating assay showing the number of colonies that form in naïve culture conditions after EpiLC differentiation.


A
WT ESC
EpiLC
WT
T3 Only
T3 -/-
4X -/-
WT + CHIR
B
mRNA Fold Change (EpiLC/ESC)
1000
100
10
1
0.1
0.01
0.001
WT
T3 Only
Tbx3
Tcl1
Klf2
Klf4
Zfp42
Tet2
Esrrb
Dazl
Pecam1
Prdm14
Fbxo15
Otx2
Lefty1
Acta2
Dnmt3b
Fgf15
Cldn6
Wnt8a
Fgf5
Pou3f1
C
mRNA Fold Change (EpiLC/ESC)
WT
T3 -/-
4X -/-
WT + CHIR
D
WT
T3 Only
T3-/-
4X
WT + CHIR
E
Colonies / 1000 cells
400
300
200
100
0
WT
T3 Only
T3 -/-
4X -/-
WT + CHIR

TCF3. WT and T3 Only cell lines differentiate properly to form flat sheets of individual cells (Fig 2.8D top row) while T3 -/- and 4X -/- retain a three-dimensional colony morphology without clear separation between cells (Fig 2.8D bottom row). Replating cells in 2iL also demonstrates defective differentiation in cell lines lacking TCF3 (Fig 2.8E). For WT and T3 only cell lines, less than ten cells out of a thousand retain the capacity to form colonies in 2iL. Conversely, more than three hundred colonies form per thousand T3 -/- cells. This percentage is comparable to undifferentiated cells maintained in 2iL (Fig 2.1D) and indicates nearly all of these cells fail to differentiate. 4X -/- cell lines also retain the capacity to form naïve colonies but at a slightly lower level than T3 -/-. This suggests that other TCF/LEFs may act opposite TCF3 and inhibit differentiation. This is consistent with previous work suggesting TCF1 and TCF3 exhibit opposing effects on ESC self-renewal[12].

Maintaining WT cells in 3 uM CHIR during EpiLC differentiation also prevents differentiation (Fig 2.8E). Morphological changes with CHIR treatment are identical to those seen in T3 -/- cell lines (Fig 2.8E). Gene expression changes are also similar between CHIR treated WT cells and T3 -/- cells at both the global level (Fig 2.8A) and for differentiation associated genes (Fig 2.8C). Altogether, this suggest an inverse relationship between WNT pathway stimulation and TCF3 activity. CHIR inhibits differentiation to EpiLC whereas TCF3 is required for this transition. The changes seen by stimulating cells with CHIR or ablating TCF3 are nearly identical.

Unsupervised clustering of TCF/LEF knockout cell lines

Much of this data can be summarized and a few conclusions can be drawn from the unsupervised clustering of gene expression patterns as measured by microarray. A dendrogram

clustering the knockout cell lines by similarity identifies three major groups (Fig 2.9A). These correspond to undifferentiated ESC (all cell lines in the ESC culture conditions, far left), well differentiated EpiLC (WT Epi and T3 Only Epi, far right), and the poorly differentiated EpiLC (T3 -/- Epi, 4X -/- Epi, and WT Epi maintained in CHIR, middle). This indicates that global gene expression differences correlate well with the differentiation status of the cells. In the naïve state, all cell lines cluster together, but there is separation that occurs during EpiLC differentiation. Here the pools correlate with the TCF3 status of cells. Cells without TCF3 group together, and cells with TCF3 group together.

This unsupervised clustering can be extended from a single dimension as in the dendrogram (Fig 2.9A), to multiple dimensions with a principal component analysis (PCA) (Fig 2.9B). A PCA identifies orthogonal sets of changes that best identify the overall variation in the data. In other words, it asks what group of gene expression changes best separate the individual samples. The answer to this defines principal component 1 (PC1). Of the remaining gene expression changes once the components of PC1 are removed, one can again ask what changes best separate the data, to define principal component 2 (PC2). This pattern can continue to define additional principal components as long as each is not correlated with any other principal component. Here is shown a principal component analysis with only the top 2 principal components, PC1 and PC2 (fig 2.9B). This allows an unsupervised separation of the data along two axes.

Intriguingly, the data naturally separates on the PC1 and PC2 with the major experimental variables: differentiation status and TCF/LEF expression (Fig 2.9B). PC1 is very similar to the single dimension dendrogram shown in figure 2.9A. Well differentiated EpiLC are shifted to the far right (WT Epi and T3 Only Epi), undifferentiated are on the far left (all cell lines in ESC conditions), and poorly differentiated EpiLC are near the middle (T3 -/- Epi, 4X -/- Epi, and WT + CH Epi). Within this initial grouping, shifts along the vertical PC2 axis are correlated

Figure 2.9 Unsupervised clustering of TCF/LEF knockout cell lines

Cells were clustered in R based on gene expression differences as measured by microarray analysis.

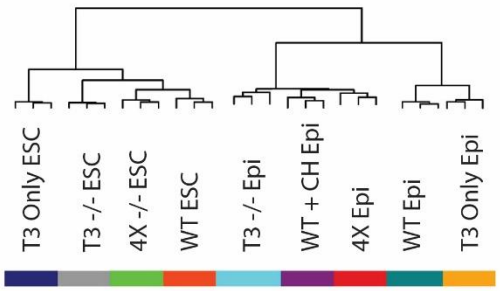
(A) Dendrogram shows clustering of WT and TCF/LEF knockout lines in naïve culture conditions (ESC) and after 72 hours of EpiLC differentiation (Epi). WT + CHIR Epi sample represents WT cells maintained in 3 uM CHIR99021 during 72 hours of EpiLC differentiation.

(B) Principal component analysis (PCA) shows clustering in two dimensions (Principal Component 1 and Principal Component 2) of WT and TCF/LEF knockout lines. Arrows show correlation of position with presence of TCF/LEFs and differentiation status.

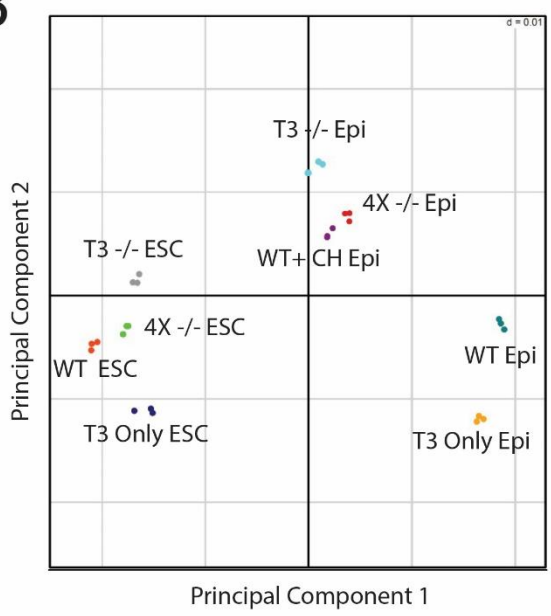
(C) Principal component analysis (PCA) shows correlation of all differentially expressed genes along Principal Component 1 and Principal Component 2 axes as described in (B). Arrows show correlation of position with presence of TCF/LEFs and differentiation status.

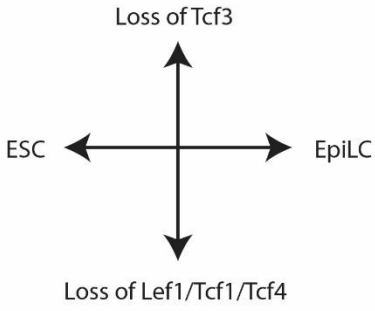
*All data, analyses, and figures were generated in collaboration with Matthew MacDougall

A

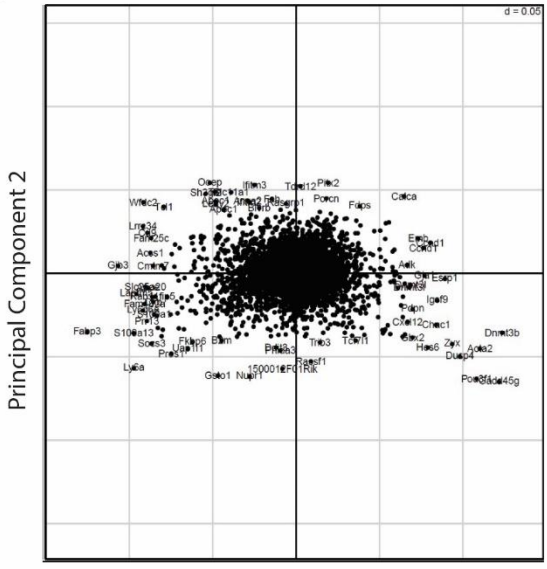

T3 Only ESC
T3 -/- ESC
4X -/- ESC
WT ESC
T3 -/- Epi
WT + CH Epi
4X Epi
WT Epi
T3 Only Epi

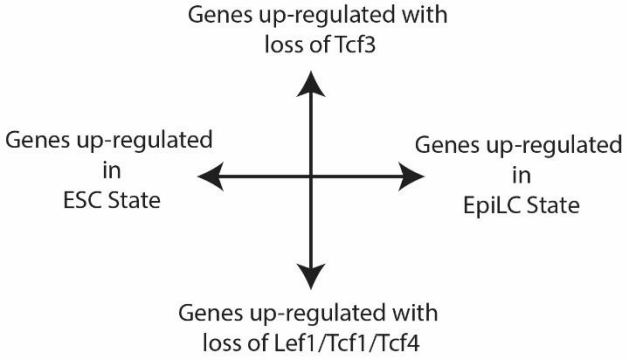
B


d = 0.01
T3 +/- Epi
4X -/- Epi
T3 -/- ESC
WT+ CH Epi
4X -/- ESC
WT ESC
WT Epi
T3 Only ESC
T3 Only Epi
Principal Component 2
Principal Component 1


Loss of Tcf3
ESC
EpiLC
Loss of Lef1/Tcf1/Tcf4

C


d = 0.05
Principal Component 2
Principal Component 1


Genes up-regulated with loss of Tcf3
Genes up-regulated in ESC State
Genes up-regulated in EpiLC State
Genes up-regulated with loss of Lef1/Tcf1/Tcf4

with TCF/LEF expression. Loss of TCF3 is associated with an upward shift on PC2, while loss of the other TCF/LEFs are associated with a downward shift. For example, T3 -/- ESC are higher as compared to WT ESC, and T3 only ESC are lower than WT ESC. Likewise T3 -/- are higher and T3 only are lower for EpiLC conditions. This analysis confirms the correlation between TCF3 status and EpiLC differentiation, but also suggests an opposing relationship between TCF3 and the other TCF/LEFs.

Using the same principal components we can also plot each gene represented on the microarray and how well its expression correlates with each axis. Expression of those in the middle, the majority, are not correlated with either principal component. However, those at the extremes are well correlated with that position for each principal component. This means that genes on the far right are associated with EpiLC differentiation while those on the far left are associated with the ESC state. Similarly, genes near the top are correlated with a loss of TCF3 and genes near the bottom with a loss of the other TCF/LEFs. This may be useful, for example, in identifying groups of genes which are upregulated during differentiation in a TCF/LEF independent (middle-right) or TCF dependent manner (top-right or bottom-right).

Discussion:

We have analyzed the expression and requirement for TCF/LEF transcription factors during the *in vitro* differentiation of ESC to EpiLC and in response to signaling pathways which regulate that differentiation. We see first that TCF/LEFs are not required for maintenance of the naïve ESC state in the presence of 2iL. This does not mean that WNT pathway stimulation is not required. TCF3 expression is reduced in response to stimulation with WNT3a or chemical inhibition of GSK3. Thus the WNT pathway may be required to reduce TCF3 in normal cells, but not in cells where it is genetically removed. In addition, TCF/LEF independent effects of WNT stimulation may stimulate ESC self-renewal. Indeed, a TCF/LEF independent effect of CHIR was seen in 4X cell lines grown in 15% FBS plus LIF.

In 4X cell lines, CHIR increases the percentage of undifferentiated colonies and increases expression of the pluripotency factor ESRRB. It would be informative to verify these changes in defined culture conditions. Examining whether CHIR increases the replating capacity of 4X cells in 2iL after EpiLC differentiation could provide a functional analysis of differentiation. Gene expression changes of 4X cell lines in response to CHIR could additionally be evaluated in naïve ESC conditions as well as EpiLC. Although we typically maintain ESC in PD, CHIR, and LIF, any combination of two is sufficient for maintaining the naïve state[14]. Thus, CHIR can be removed and then added back later to evaluate expression changes. Knowing the set of gene expression changes may provide clues for what regulates them. If changes are seen, the requirement for β-Catenin can be assessed by knocking it out in 4X cell lines. β-Catenin -/- cell lines have been previously generated and maintained in both serum plus LIF and 2iL culture conditions, although they have some defects in cell adhesion, are more prone to differentiation, and exhibit less robust growth in 2iL[5,13,184].

Although TCF/LEF independent effects may exist, the primary factor for differentiation is TCF3. TCF3 is required for differentiation to EpiLC, and among the TCF/LEFs is sufficient alone to mediate this transition. The mechanism of this is interesting to consider. On one hand, it may act simply as a transcriptional repressor of core pluripotency transcription factors[5,21,37]. It is intriguing, however, that even in those genes that TCF3 does not repress directly, it appears to bind to enhancers which disappear during differentiation. Looking at the subset of ESC associated genes that are downregulated during EpiLC differentiation, as described in figure 2.1B, TCF3 binds to a DNAse hypersensitivity site (DHS) within or a little upstream of each gene (Fig 2.10B). DHS represent regions of open chromatin that are more accessible to DNA binding proteins. Because they are more accessible to transcription factors, they are frequently important regulatory regions for gene expression or chromatin structure. Examining the DHS pattern in ESC as compared to the more differentiated mesoderm cells, many of the peaks are lost. These genes are not expressed in mesoderm and no longer get activated in response to various stimuli. Conversely, and not shown here, many new regulatory regions are opened up during differentiation. This process appears to be partially mediated by OCT4 and OTX2 during the EpiLC transition[178,185].

Intriguingly, those peaks associated with TCF3 are preferentially lost during differentiation. The mechanism of enhancer loss or “decommissioning” is not understood. However, the presence of TCF3 at decommissioned sites and the requirement for TCF3 for efficient differentiation suggests it could play an active role in this process. TCF3 might recruit a co-factor to drive epigenetic modifications or possess some unique biochemical activity itself. Alernatively, TCF3 may drive differentiation indirectly through transcriptional represson of NANOG, ESRRB, and TCFCP2L1[5,21,37].

During development, decommissioning of naïve regulatory regions may be important to alter the response to external stimuli. An example of this is the WNT pathway itself. In the ESC

state, WNT stimulation maintains self-renewal and inhibits differentiation[13,14,16,3529,34]. However, after differentiation to EpiLC, WNT stimulation promotes differentiation to mesendoderm lineage[17-20]. At the EpiLC state, naïve genes important for maintaining the ESC state, such as *Esrrb*, are no longer upregulated in response to WNT stimulation[181]. Instead, genes important for differentiation, such as the pan-mesodermal marker *Brachyury*, become newly responsive[18,19]. The pattern of gene expression changes can be visualized *in vitro* when comparing the response to CHIR in the naïve ESC state versus the EpiLC state (Fig 2.10A). One question is whether the genes that lose their responsiveness have lost WNT responsive enhancers, and if so whether this decommissioning is driven by TCF3.

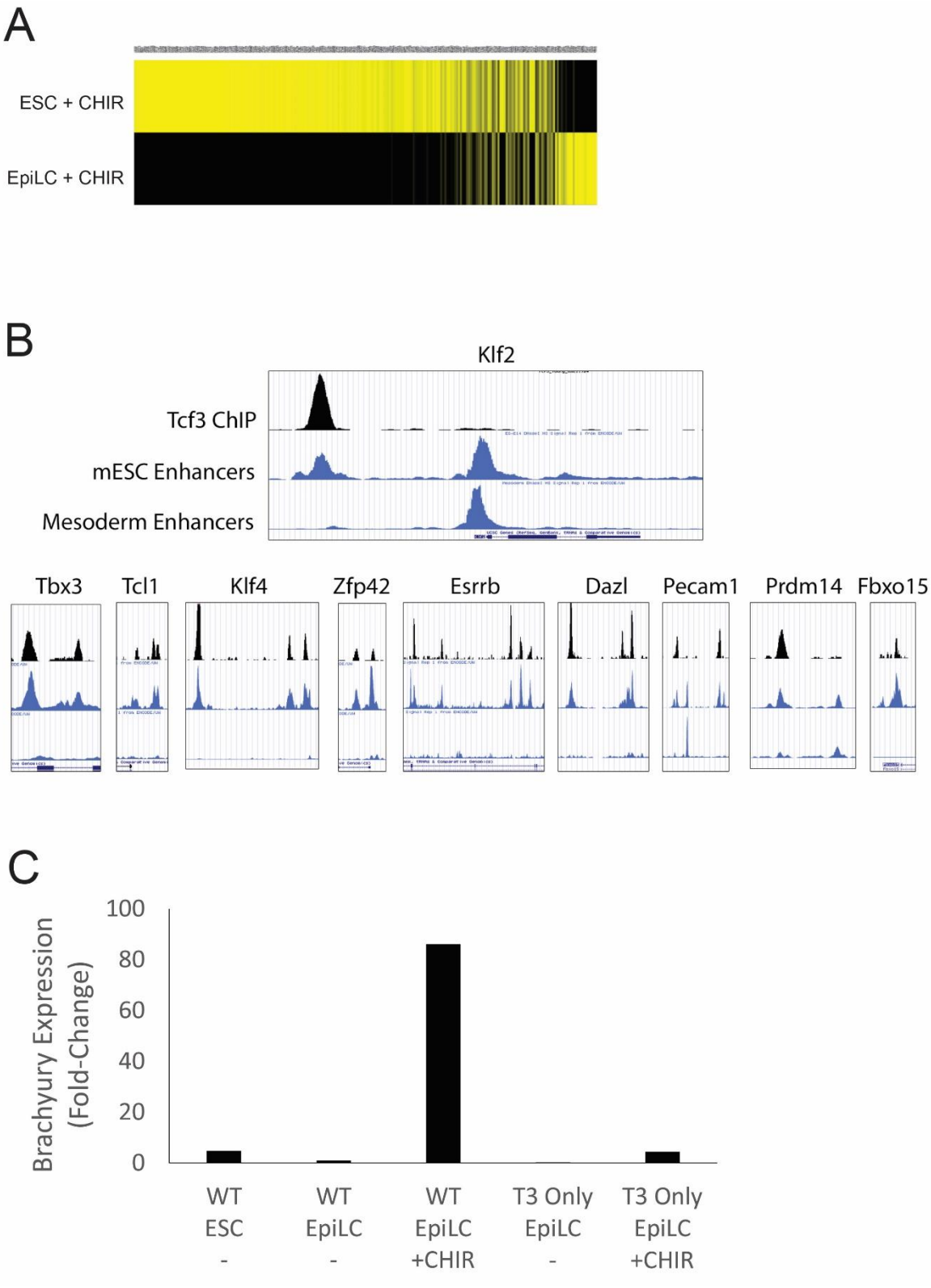
In the opposite direction, what is driving the newly WNT responsive genes such as *Brachyury*? Do these contain new WNT responsive regulatory regions of open chromatin and, if so, are these generated by OCT4 and OTX2 or a distinct mechanism? Alternatively, they may simply require a WNT responsive transcription factor which is not present in ESC. LEF1, for example, is newly expressed in EpiLC. In potential agreement with this, although T3 Only cell lines can differentiate to the EpiLC state, they are deficient in stimulating *Brachyury* expression in response to CHIR as compared to WT (Fig 2.10C). Thus LEF1 could be required to drive mesoderm differentiation. *In vivo* data, however, does not support this conclusion. LEF1 and TCF1 are expressed specifically in the primitive streak during gastrulation, and LEF1/TCF1 double knockout mice are deficient in paraxial mesoderm production. However axial, intermediate, and lateral mesoderm fates develop normally, suggesting the defect occurs after the initial mesoderm specification[85]. This phenotype is identical to that of WNT3a -/- mice[186]. In contrast, ablation of WNT3, a different gene with an unfortunately similar name, generates none of the mesoderm fates and completely fails to form a primitive streak, suggesting WNT stimulated factors other than LEF1 and TCF1 are involved at earlier stages of mesoderm induction[8].

Figure 2.10 Changes in Wnt pathway stimulation and DNAse hypersensity associated with EpiLC differentiation.

(A) mRNA expression changes exhibiting greater than 2-fold increase in response to CHIR stimulation in naïve culture conditions (ESC + CHIR) or after 72 hours of EpiLC differentiation (EpiLC + CHIR). EpiLC + CHIR represents WT cells that were allowed to differentiate in the absence of CHIR for 48 hours and then CHIR was reintroduced for the last 24 hours.

(B) Comparison of TCF3 chromatin binding as measured by ChIP-seq and DNAse hypersensitivity sites (labeled Enhancers) in either ESC or Mesoderm. Binding patterns were examined for genes that lose expression during EpiLC differentiation as described in Figure 2.1C.

(C) Induction of Brachyury expression as measured by microarray in WT and TCF3 Only cell lines after EpiLC differentiation and subsequent stimulation with 3 uM CHIR for 24 hours. Fold-change is calculated in comparison to untreated WT EpiLC.


A
ESC + CHIR
EpiLC + CHIR
B
Klf2
Tcf3 ChIP
mESC Enhancers
Mesoderm Enhancers
Tbx3
Tcl1
Klf4
Zfp42
Esrrb
Dazl
Pecam1
Prdm14
Fbxo15
C
Brachyury Expression (Fold-Change)
100
80
60
40
20
0
WT ESC -
WT EpiLC -
WT EpiLC +CHIR
T3 Only EpiLC -
T3 Only EpiLC +CHIR

Experimental Procedures:

ESC culture and differentiation

Naive-state wild type and Tcf/Lef-mutants underwent at least 5 passages in N2B27 (Gibco) media supplemented with 1μM PD0325901 (Meki, Stemolecule), 3μM CHIR99021 (Gsk3i, Sigma), and 1000 units/mL LIF (EMD Millipore) on tissue culture (TC) plates pre-coated with 7.5μg/ml poly-ornithine (Advanced Biomatrix) and 5μg/ml ultrapure laminin (Corning; Novus). 400,000 to 750,000 cells were seeded onto 10 cm^2 TC plates and grown for 3-5 days, changing media every 2 days, prior to sample collection. For differentiation to EpiLC, 200,000-500,000 ESC were seeded on to 10cm^2 TC plates pre-coated with 15 μg/mL fibronectin (Millipore) in N2B27 media supplemented with PD, LIF, and 1% knockout serum replacement ("KOSR", Gibco). After 24 hours, cells were treated with N2B27 media supplemented with 12 ng/mL Fgf2 (R&D) and 1% KOSR for 2 days, with media changed daily.

Generation of Tcf/Lef mutant ESC lines via CRISPR/Cas9 mutagenesis

For ESC lines which required Tcf7l1-ablation, a previously described Tcf7l1-/- 129/Sv ESC line was used to mutate other, additional Tcf/Lef genes[93]. To facilitate knock-out, sgRNA targeting the HMG domain of each of the Tcf/Lef gene were designed using a published sgRNA design algorithm (crispr.mit.edu)[182]. Bsb1 restriction enzyme overhangs were appended to the 5' ends (CACC to sense, AAAC to anti-sense) of the guide sequences for easy integration downstream of the U6 promoter in the px330 plasmid[168]. ESC were transfected with equimolar amounts of the sgRNA expressing px330 plasmid and a plasmid expressing puromycin

resistance using Opti-MEM (Gibco) and Lipofectamine 2000 (Invitrogen). In ESC lines, in which ablation of multiple Tcf/Lefs was desired, transfection with all appropriate guides occurred simultaneously. After puromycin selection, ESC were seeded at low density for clonal isolation (~2000 cells per well of 6-well TC dish). Individual clones were screened by western blot to identify those positive for Tcf/Lef-ablation. All in all, 14 new ESC lines were generated, making all possible combinations of Tcf/Lef knockouts.

Western blot for Tcf/Lef factors

Protein lysates were collected using 2x Laemmli buffer (0.4 g/mL SDS, 0.2 mg/mL bromophenol blue, 0.16 v/v glycerol, 20 mM Tris 0.5M pH6.8, 200 uM DTT). The following antibodies were used to detect Tcf/Lefs: anti-Tcf7 (C63D9), anti-Lef1 (C12A5), anti-Tcf7l2 (C9B9), and a homemade polyclonal rabbit anti-Tcf7l1.

TOPFlash Luciferase reporter assay

Twenty four hours prior to transfection, 100,000 ESC maintained in serum+LIF media were seeded per well of a 24-well plate in duplicates. Cells were transfected with SuperTOPFlash Firefly and Renilla luciferase reporter plasmids using Opti-MEM (Gibco) and Lipofectamine 2000 (Invitrogen). One day after transfection, cells were lysed with 1X passive lysis buffer (Promega). A Clarity luminometer (Bio-Tek) was used to measure luciferase activities using a dual luciferase reporter assay kit (Promega). Relative activity was calculated as the ratio of the reporter plasmid Firefly luciferase activity to Renilla luciferase activity (pRL-CMV).

Chromatin Immunoprecipitation (ChIP) with quantitative PCR (qPCR)

ESC were grown to confluency in a 10cm^2 dish in serum+LIF conditions (roughly ~15 million cells to be divided into 3 replicates). To facilitate cross-linking, 37% formaldehyde (Fisher) was added for a final concentration of 1%, followed by incubation under overnight at 4°C. 1/20 volume of glycine (Fisher) was added to quench the reaction. Cells were then washed with cold phosphate buffered saline (PBS) twice prior to harvesting by silicon cell scrapers and centrifugation (4°C, 4000 rpm), followed by flash freezing with liquid nitrogen and storage at -80°C. After thawing, all subsequent steps were performed at 4°C or on ice. Cells were resuspended in lysis buffer (LB) 1 (50 mM Hepes pH 7.7, 140 mM NaCl, 1 mM EDTA, 10% Glycerol, 0.5% NP-40, 0.25% Trtiton-X-100). After incubation for 10 minutes under constant rotation, cells were pelleted by by centrifugation (2500 rpm) and resuspended in LB 2 (200 mM NaCl, 1 mM EDTA, 0.5 mM EGTA, 10 mM Tris pH 7.5). After incubation for 10 minutes under constant rotation, cells were pelleted by by centrifugation (2500 rpm) and resuspended in LB 3 (1 mM EDTA, 0.5 mM EGTA, 10 mM Tris pH 7.5, 100 mM NaCl, 0.1% Na-deoxycholate). Protease inhibitors were added to all lysis buffers prior to resuspension. Sonication was performed using a Branson Digital Sonifier 450 at 60% amplitude on ice with 20 cycles of 30 seconds ON, 60 seconds OFF. After sonication, 1/10 volume of 10% Triton-X-100 was added before centrifugation to remove cellular debris. 50 uL of chromatin extract was uncrosslinked for subsequent input DNA isolation. Resulting chromatin extracts were incubated overnight at 4°C with 20 ul of Protein G Dynabeads (Invitrogen) pre-incubated with 5 µg of β-catenin (Invitrogen 71–2700). Beads were washed 4 times with RIPA buffer (50 mM Hepes, 1mM EDTA, 0.7% Na-deoxycholate, 1% NP-40, 0.5 M LiCl) and 1 time with TBS (50

mM Tris, 150 mM NaCl, pH 7.6). Bound complexes were eluted from the beads by heating at 65°C with occasional vortexing. Crosslinks were reversed by incubation at 65°C for 6 hours to overnight. DNA was isolated using phenolchloroform extraction and ethanol precipitation, followed by resuspension in Tris-EDTA.

Microarray sample collection and data analysis

Biological triplicates were prepared for RNA isolation with TRIZOL, followed by column purification with the Qiagen RNeasy kit per manufacturer protocol. Subsequently, RNA was hybridization to an Illumina Mouse Whole Genome 8-V2 Beadchip microarray. Resulting microarray data were analyzed using the Lumi R package.

Chapter 3: Regulation of Tcf7l1 DNA binding and protein stability as principle mechanisms of Wnt/β-catenin signaling.

*****With the exception of several formatting changes, the content of chapter has been published in the following manuscript:

Shy, B.R., Wu, C., Khramtsova, G.F., Zhang, J.Y., Olopade, O.I., Goss, K.H., and Merrill, B.J. "Regulation of Tcf7l1 DNA binding and protein stability as principle mechanisms of Wnt/β-catenin signaling". **Cell Reports, 2013.**

Abstract:

Wnt/β-catenin signal transduction requires direct binding of β-catenin to Tcf/Lef proteins, an event classically associated with stimulating transcription by recruiting co-activators. This molecular cascade plays critical roles throughout embryonic development and normal postnatal life by affecting stem cell characteristics and tumor formation. Here, we show this pathway utilizes a fundamentally different mechanism to regulate Tcf7l1 (formerly named Tcf3) activity. β-catenin inactivates Tcf7l1 without a switch to a co-activator complex by removing it from DNA, an effect leading to Tcf7l1 protein degradation. Mouse genetic experiments demonstrate that Tcf7l1 inactivation is the only required effect of the Tcf7l1-β-catenin interaction. Given the expression of Tcf7l1 in pluripotent embryonic and adult stem cells, and in poorly differentiated breast cancer, these finding provide new mechanistic insights into the regulation of pluripotency and the role of Wnt/β-catenin in breast cancer.

Introduction:

Canonical Wnt/β-catenin signaling impacts a wide range of biological activities, including stem cell self renewal, organ morphogenesis, and tumor formation [50,187]. Regulation of the pathway centers on the stability of β-catenin, which is targeted for proteasome-mediated degradation by a complex containing Adenomatous Polyposis Coli (APC), Axin structural proteins, and Glycogen Synthase Kinase 3 (GSK3) [188]. Phosporylation of β-catenin by GSK3 stimulates degradation dependent upon APC, Axin, and the β-TrCP E3 ligase [189-192]. Wnt signaling inhibits degradation of β-catenin by blocking its ubiquitination (Li et al., 2012). Pharmacological GSK3 inhibitors similarly inhibit β-catenin degradation by blocking β-catenin phosphorylation.

An important downstream mechanism of the Wnt/β-catenin pathway occurs as β-catenin binds to the amino terminal of Tcf/Lef proteins thereby displacing co-repressor proteins bound to the Tcf/Lef [86,87,193]. Tcf-β-catenin binding subsequently recruits transactivator proteins to the genomic sites previously occupied by co-repressors [194-197]. This accepted model of canonical Wnt/β-catenin signaling is consistent with observed effects of Tcf/Lef proteins in many contexts [72], however, it is not consistent with recent observations for mammalian Tcf7l1 (formerly Tcf3). In cells where Lef1 and Tcf7 (formerly Tcf1) act as β-catenin-dependent transactivators, only transcriptional repressor activity for Tcf7l1 was detected [198,199]. Here, we show that β-catenin-binding to Tcf7l1 does not form a transactivation complex, but instead initiates a fundamentally distinct mechanism. β-catenin binding inactivates Tcf7l1 by reducing its chromatin occupancy and secondarily stimulates its protein degradation. Mouse genetic experiments demonstrate that this inactivation is the only necessary function of the Tcf7l1-β-catenin interaction. These molecular and genetic findings provide new insights into the role of Wnt/β-catenin signaling in cells where Tcf7l1 expression is prominent, including embryonic stem cells (ESC) and poorly differentiated breast cancer.

Results:

β-catenin reduces Tcf7l1 protein levels by stimulating protein degradation

Molecular support for the conversion of Tcf/Lef proteins into transactivators by β-catenin includes the ability of a β-catenin-Tcf7 fusion protein to activate target genes without Wnt pathway stimulation (Staal et al., 1999). If Tcf7l1 was switched to a transactivator by β-catenin binding, one would expect a β-catenin-Tcf7l1 fusion protein to similarly activate target genes. In ESC, the β-catenin-Tcf7l1 fusion was unable to activate TOPFlash and LRH-1 reporter genes, and instead repressed Wnt3a-stimulation of reporter genes (Figure 3.1A). Rather than converting Tcf7l1 to a transactivator, Wnt/β-catenin stimulation notably decreased Tcf7l1 protein in ESC treated with recombinant Wnt3a or the GSK3 inhibitor, Chiron99021 (CH) (Figure 3.1B). These results indicate a significant difference in the downstream effects of Tcf7-β-catenin and Tcf7l1-β-catenin interaction.

To elucidate the transactivation independent effects of β-catenin on Tcf7l1, we investigated how Tcf7l1 protein levels were reduced. Wnt3a- and CH-treated ESC displayed increased Lef1 and Tcf7 mRNA levels that correlated with increased protein levels (Figure 3.1B), consistent with Lef1 and Tcf7 being Wnt/β-catenin target genes (Filali et al., 2002; Hovanes et al., 2000; Roose et al., 1999; Waterman, 2004). In contrast, decreased Tcf7l1 protein was not paralleled by a significant change in mRNA levels (Figure 3.1B), indicating that β-catenin regulation of Tcf7l1 does not occur transcriptionally. Because Dgcr8 is a required component of the microprocessor complex necessary for biogenesis of microRNAs (Wang et al., 2007), CH-stimulated reduction of Tcf7l1 in DGCR8-/- ESC showed that reduction of Tcf7l1 protein was also not microRNA-mediated (Figure 3.1C). Treatments with the proteasome inhibitors MG-132 or MG-115 effectively blocked the Wnt3a- and CH-stimulated reduction of Tcf7l1 protein (Figure 3.1D, 3.2A), demonstrating that reduction of Tcf7l1 required protein

Figure 3.1 Wnt/β-catenin stimulates Tcf7l1 protein degradation.

A) Transient transfection of Tcf7l1-/- ESC with β-catenin-Tcf fusion plasmids and SuperTOPFlash (top) or LRH1 promoter (bottom) luciferase reporter plasmids. Values represent mean +/- standard deviation for triplicate transfections.

B) Western blot (top) and Quantitative RT-PCR (bottom) analyses of ESC treated with 50ng/ml recombinant Wnt3a (left) or 3μM CH (right) for the indicated time.

C-E) Western blot analysis of ESC treated with 3μM CH for 24 hrs in DGCR8 mutant cells (C), for 6 hrs together with MG-132 (5μM) in Tcf7l1 +/+ and Tcf7l1Δn/ΔN mutant cells (D), and for 12 hrs together with leptomycin B (E)

F) Distribution of relative levels of nuclear Tcfl71 immunoreactivity in Tcf7l1+/+ (top) and Tcf7l1ΔN/ΔN (bottom) cells expressing either GFP (red bars) or ΔNβ-catenin (blue bars). A total of 200 nuclei were counted for each condition. Data are representative of three separate experiments (see Figure 3.2C for example of immunofluorescence and Figure 3.2D for treatment with ΔNΔCβ-catenin).

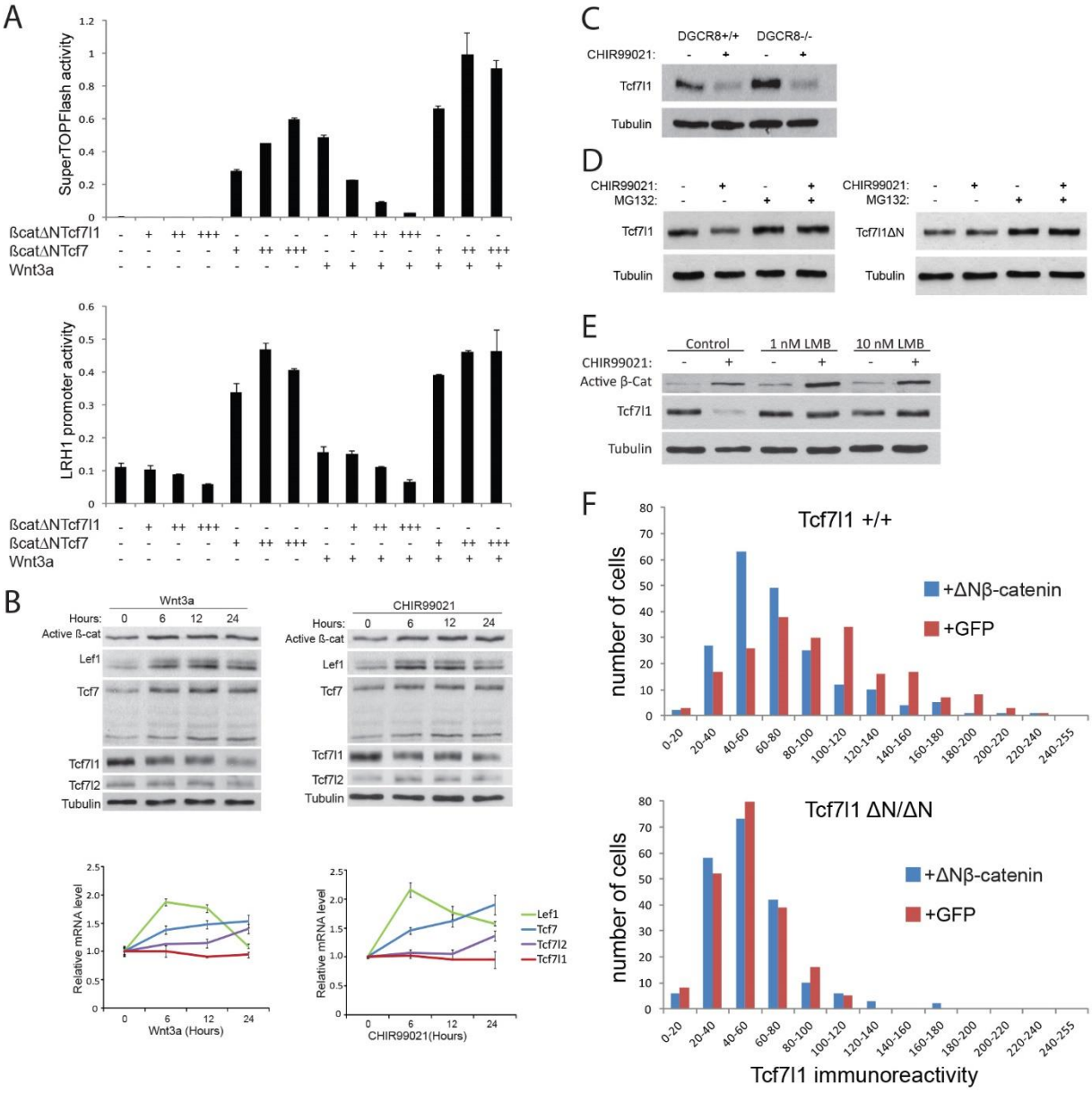

A
SuperTOPFlash activity
ßcatΔNTcf7l1
ßcatΔNTcf7
Wnt3a
LRH1 promoter activity
B
Wnt3a
CHIR99021
Hours:
Active ß-cat
Lef1
Tcf7
Tcf7l1
Tcf7l2
Tubulin
Relative mRNA level
Wnt3a (Hours)
CHIR99021(Hours)
C
DGCR8+/+
DGCR8-/-
CHIR99021:
D
MG132:
Tcf7l1ΔN
E
Control
1 nM LMB
10 nM LMB
F
Tcf7l1 +/+
Tcf7l1 ΔN/ΔN
number of cells
+ΔNβ-catenin
+GFP
Tcf7l1 immunoreactivity

Figure 3.2 β-catenin interaction required for Tcf7l1 protein degradation.

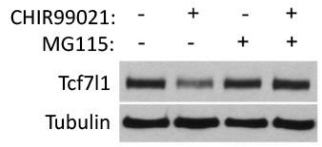
A) Western blot analysis of Tcf7l1 protein levels in ESC treated for 6 hours +/- 3μM CHIR and/or 5 μM MG115.

B) Western blot analysis and quantitation of Tcf7l1 protein levels in Tcf7l1ΔN/ΔN ESC treated for 12 hours +/- 50 ng/mL recombinant Wnt3a.

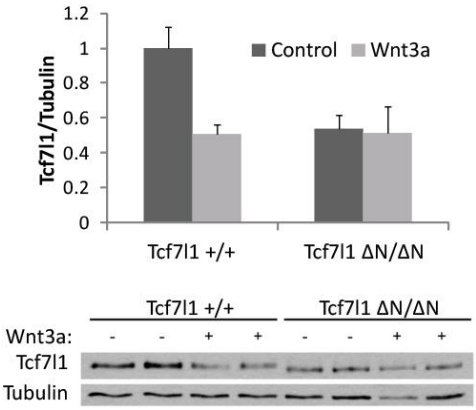
C) Transient transfection of GFP or ΔNβ-catenin expression plasmids was performed in Tcf7l1+/+ (top) and Tcf7L1ΔN/ΔN (bottom) ESC. Immunofluorescent detection of Tcf7l1 protein (cyan), GFP (green) or β-catenin (magenta), and DNA (blue) was measured in nuclei of transfected cells. Images were processed for quantitative immunofluorescence to generate data for graphs (Figure 3.1F, 3.2D).

D) Quantitative immunofluorescence for Tcf7l1 protein in ESC transiently transfected with ΔNβ-catenin (blue), ΔNΔCβ-catenin (green), or GFP only (red) expression plasmids. Bars indicate number of cells (of a total of 200 counted) display the indicated relative intensity of Tcf7l1 immunoreactivity.

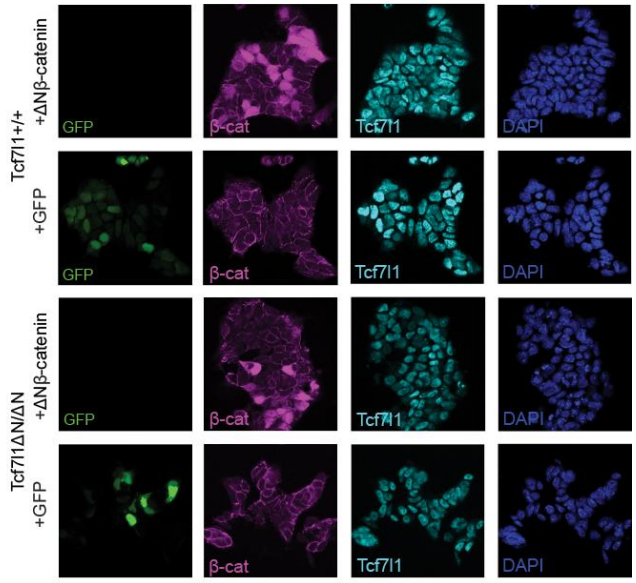
A


CHIR99021: - + - +
MG115: - - + +
Tcf7l1
Tubulin

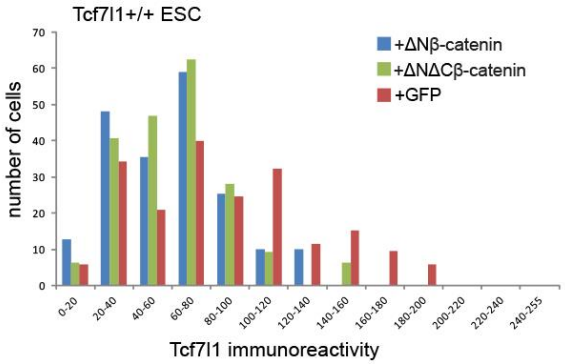
B


Tcf7l1/Tubulin
1.2
1
0.8
0.6
0.4
0.2
0
Control
Wnt3a
Tcf7l1 +/+
Tcf7l1 ΔN/ΔN
Tcf7l1 +/+
Tcf7l1 ΔN/ΔN
Wnt3a: - - + + - - + +
Tcf7l1
Tubulin

C


Tcf7l1+/+
+ΔNβ-catenin
+GFP
Tcf7l1ΔN/ΔN
+ΔNβ-catenin
+GFP
GFP
β-cat
Tcf7l1
DAPI

D


Tcf7l1+/+ ESC
+ΔNβ-catenin
+ΔNΔCβ-catenin
+GFP
number of cells
70
60
50
40
30
20
10
0
0-20
20-40
40-60
60-80
80-100
100-120
120-140
140-160
160-180
180-200
200-220
220-240
240-255
Tcf7l1 immunoreactivity

degradation. Finally, reduction of Tcf7l1 was blocked by leptomycin B, indicating that it required Exportin1-mediated nuclear transport (Figure 3.1E).

To determine the role of β-catenin binding to Tcf7l1, Tcf7l1ΔN/ΔN knockin ESC were used. In contrast to wild-type Tcf7l1, Tcf7l1ΔN was not degraded in response to CH or Wnt3a (Figure 3.1D, 3.2B), indicating the Tcf7l1-β-catenin interaction was necessary for degradation. To determine if the interaction was sufficient for degradation, ΔNβ-catenin was expressed in ESC and Tcf7l1 levels were measured by quantitative immunofluorescence. ΔNβ-catenin expression was sufficient to reduce nuclear Tcf7l1 levels in Tcf7l1+/+, but not in Tcf7l1ΔN/ΔN cells (Figure 3.1F, 3.2C). Interestingly, several recent studies showed that a mutant form of β-catenin (β-cateninΔC) supported self-renewal of mouse ESC, and complemented defects caused by ablation of β-catenin despite the lack of the C-terminal transactivation domain in the β-cateninΔC mutant [13,97,184]. Therefore, it is notable that expression of a ΔNβ-cateninΔC was also sufficient to reduce nuclear Tcf7l1 protein levels ESC (Figure 3.2D). Given the substantial effects of altering Tcf7l1 levels in ESC [12,13,93,200], the reduction of Tcf7l1 protein provides a mechanism for the poorly understood pro-self-renewal effects of β-cateninΔC in ESC.

Reduction of Tcf7l1 is sufficient to replace Tcf7l1-β-catenin interaction

If a principle mechanism of Wnt/β-catenin signaling functions through inactivation of Tcf7l1, and not conversion to a Tcf7l1-β-catenin transactivator complex, then reducing the level of Tcf7l1 should be sufficient to replace the Tcf7l1-β-catenin interaction. This hypothesis was first tested in ESC, where reducing the amount of Tcf7l1ΔN by siRNA stimulated reporter gene response to Wnt3a (Figure 3.4A,B). For examination of broader effects of reducing Tcf7l1 in mice, the level of Tcf7l1 was reduced by breeding for hemizygous mice (i.e. Tcf7l1+/- or Tcf7l1-/ΔN; Figure 3.4C). It is important to note that Tcf7l1-/- mice die shortly after gastrulation [89].Tcf7l1ΔN/ΔN

embryos progress normally through gastrulation, but later develop a constellation of morphogenetic defects causing death for all Tcf7l1ΔN/ΔN mice at or before birth [181,199]. Mating Tcf7l1+/- with Tcf7l1+/ΔN mice produced the Mendelian-expected ratio of Tcf7l1-/ΔN offspring, despite the genetic absence of a Tcf7l1 protein capable of interacting with β-catenin (Figure 3.3A, S2D). Moreover, Tcf7l1-/ΔN mice displayed none of the morphogenetic defects observed in Tcf7l1ΔN/ΔN mice, including poor vascular integrity, edema, oligodactyly, and opened eyelids (Figure 3.3C-D'', S2E,F). Indeed, Tcf7l1-/ΔN mice advanced to adulthood and appeared indistinguishable from Tcf7l1+/+ littermates throughout their ostensibly normal lifetimes (Figure 3.3B). Thus, removing one copy of Tcf7l1ΔN genetically rescued the defects caused by ablating the Tcf7l1-β-catenin interaction. These results demonstrate that inactivation of Tcf7l1 by β-catenin is the necessary effect downstream of Tcf7l1-β-catenin interaction for mouse embryogenesis and postnatal viability.

To determine the effects of reducing Tcf7l1 at the target gene level in mice, tissues previously shown to be affected in Tcf7l1ΔN/ΔN embryos were examined in Tcf7l1-/ΔN embryos harboring the BAT-Gal reporter. Compared to Tcf7l1+/+ e14.5 eyelids (Figure 3.3E-E'''), Tcf7l1ΔN/ΔN displayed a restricted domain of BAT-Gal activity and decreased expression of Lef1, a Wnt/β-catenin target, in the mucocutaneous junction of the eyelid (Figure 3.3F-F''') [199]. The domain of Lef1 expression and BAT-Gal activity was increased in Tcf7l1-/ΔN relative to Tcf7l1ΔN/ΔN, and BAT-Gal activity was detected only in cells expressing Lef1 (Figure 3.3 F-G''', Figure 3.4G,H). Given the inability of Tcf7l1ΔN to respond to β-catenin, the rescue of BAT-Gal activity in Tcf7l1-/ΔN embryos shows that activation is mediated by Lef1, and attenuation of this activation depends on the level of Tcf7l1-repressor.

Figure 3.3 Reducing Tcf7l1 levels replaces the requirement for β-catenin interaction in mice.

A-D'') Tcf7l1-/ΔN mice appear normal at birth (A) and through adult stages (B) (see also Figure 3.4D-F). Phenotypes observed in Tcf7l1ΔN/ΔN embryos, including oligodactyly (C-C'') and opened eyelids at birth (D-D''), were not observed in Tcf7l1-/ΔN embryos. (See also Figure 3.4E,F).

E-G''') Tcf/Lef-β-catenin activation of BAT-Gal reporter is restored in the Tcf7l1-/ΔN eyelid. Immunofluorescence for Tcf7l1 (green), Lef1 (red), and β-galactosidase (blue) displayed for e14.5 eyelids from BAT-Gal transgenic Tcf7l1+/+ (E-E'''), Tcf7l1ΔN/ΔN (F-F'''), and Tcf7l1-/ΔN (G-G''') embryos. Arrows point to Lef1 positive nuclei with low levels of Tcf3. Arrowheads point to Tcf7l1 and Lef1 double positive nuclei, which were observed in Tcf7l1+/+ (E-E''') and Tcf7l1-/ΔN (G-G''') but not Tcf7l1ΔN/ΔN (F-F'''). The dotted line denotes the BAT-Gal-positive region. Asterisks in (F) mark Tcf7l1 positive cells in the nearby cornea.

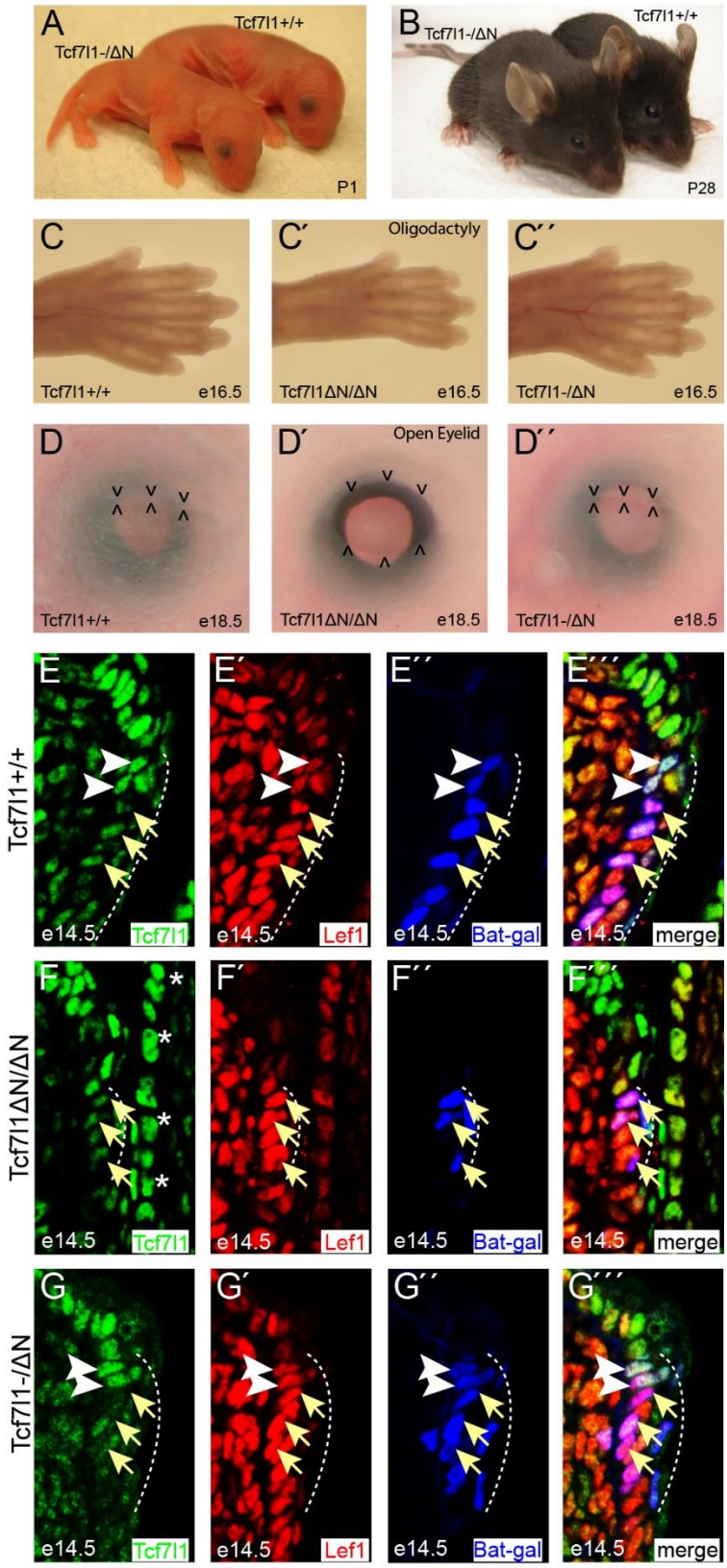

A
Tcf7l1+/+
Tcf7l1-/ΔN
P1
B
Tcf7l1-/ΔN
Tcf7l1+/+
P28
C
Tcf7l1+/+
e16.5
C´
Oligodactyly
Tcf7l1ΔN/ΔN
e16.5
C´´
Tcf7l1-/ΔN
e16.5
D
Tcf7l1+/+
e18.5
D´
Open Eyelid
Tcf7l1ΔN/ΔN
e18.5
D´´
Tcf7l1-/ΔN
e18.5
Tcf7l1+/+
E
E´
E´´
E´´´
e14.5
Tcf7l1
Lef1
Bat-gal
merge
Tcf7l1ΔN/ΔN
F
F´
F´´
F´´´
e14.5
Tcf7l1
Lef1
Bat-gal
merge
Tcf7l1-/ΔN
G
G´
G´´
G´´´
e14.5
Tcf7l1
Lef1
Bat-gal
merge

Figure 3.4 Reduction of Tcf7l1ΔN rescues Wnt activation.

A) Transient transfection of Tcf7l1+/+ (WT) or Tcf7l1ΔN/ΔN (KI) ESC with SuperTOPFlash Luciferase reporter and various concentrations (0.04nM to 5nM) of siRNA compounds. Control siRNAs (SCRM) and two independent Tcf7l1-specific (#1, #2) siRNAs were used. Reporter activity was stimulated by treating with Wnt3a-conditioned media or control conditioned media for 24hrs. Values represent means +/- standard deviations of biological triplicates.

B) (top) Western blot showing degree of Tcf7l1 protein knockdown caused by siRNA used for reporter assay (A) and (bottom) coomassie stained gel showing even loading of protein lysates samples.

C) Western blot analysis (top, middle) of whole e8.5 mouse embryos determined to have Tcf7l1+/+ and Tcf7l1+/- genotypes based on PCR genotyping reactions (bottom).

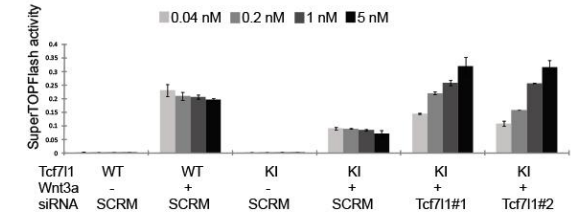
D) Recovery of offspring with the indicated genotypes from mating Tcf7l1+/- and Tcf7l1+/ΔN mice. Note the normal viability of Tcf7l1-/ΔN mice.

E) Images of whole e15.5 embryos of the indicated genotype. Note the lack of edema exhibited by Tcf7l1-/ΔN mice.

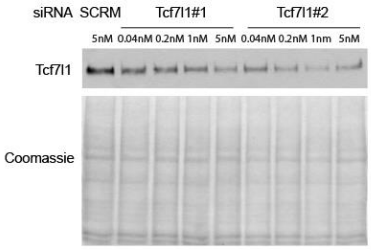
F) Images of whole e18.5 embryos of the indicated genotype. Note the lack of hemorrhage in the Tcf7l1-/ΔN mice.

G,H) BAT-Gal transgenic mice harboring the Tcf/Lef-β-catenin reporter transgene were X-gal stained to identify Tcf/Lef-β-catenin activity. (G) In the limb buds of e10.5 embryos, the intense domain of activity at the posterior region (arrows) is absent in Tcf7l1ΔN/ΔN embryos and returns in Tcf7l1-/ΔN embryos. (H) In 8um thin cryosections of the eyelids from e14.5 embryos, the domain of BAT-Gal activity at the mucocutaneous junction (dotted line) is reduced in Tcf7l1ΔN/ΔN and restored to normal in Tcf7l1-/ΔN embryos.

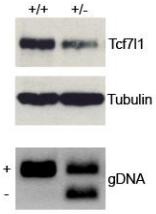
A



B



C

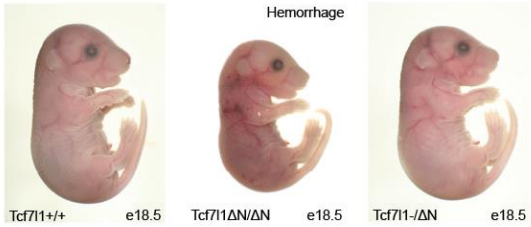


D

Stage	Total number of mice	Number of mice of each genotype			
		Tcf7l1+/+	Tcf7l1+/-	Tcf7l1+/ΔN	Tcf7l1-/ΔN
E10.5	70	16	17	16	21
E15.5	135	30	31	46	28
E18.5	14	2	5	3	4
P20	190	46	51	51	42

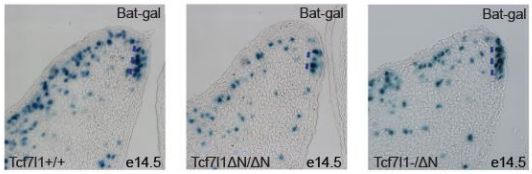
E

F



G

H



β-catenin stimulates Tcf7l1 inactivation in human breast cancer.

In addition to pluripotent cells in the early mammalian embryo, and ESC in vitro, Tcf7l1 mRNA expression has been noted in several types of adult stem cells and in poorly differentiated cancers [201-204]. We reasoned the novel aspects of the Tcf7l1-destabilization mechanism could provide new insights into the effects of Wnt/β-catenin in these important contexts. Breast cancer was chosen for further analysis, because: 1. Wnt/β-catenin has been known to effect mammary tumors since its discovery by Nusse and Varmus [205], 2) despite the long history, there remains a poor understanding of underlying mechanisms of Wnt/β-catenin effects in this disease [206], 3) poorly differentiated mammary tumors express high levels of Tcf7l1 mRNA (0.81 mean Tcf7l1 mRNA +/- 0.91 standard deviation for 270 Basal tumors, -0.32 +/- 0.67 for 941 non Basal; $p<0.0001$) (Figure 3.5A) [201], and 4) Altering the level of Tcf7l1 caused significant effects in xenograft tumor formation experiments [207].

Consistent with previous analyses, we noted a very high frequency of the basal molecular subtype among invasive mammary tumors expressing the highest levels of Tcf7l1 mRNA (83% of tumors with Tcf7l1 mRNA >1.1 were basal vs 22% of all tumors assessed were basal; Figure 3.5A)[201,207]. To examine patterns of Tcf7l1 protein expression, a Tcf7l1 specific antibody was used for immunohistochemistry (IHC) staining of an array of breast cancer tissue samples (Figure 3.6A)[93,208]. In contrast to Tcf7l1 mRNA, nuclear Tcf7l1 protein was not significantly higher in basal subtype tumors (55 +/- 60, n=23) relative to non-basal tumors (44 +/- 74, n=47) (Figure 3.5B, 3.6A), and the frequency of basal tumors displaying strong nuclear Tcf7l1 (i.e nuclear IHC score > 180) was lower than the overall frequency of basal tumors on the array (25% with high nuclear Tcf7l1 vs 32% of all tumors) (Figure 3.5C, 3.6B,C). Thus, although Tcf7l1 mRNA is highly elevated in basal subtype tumors, Tcf7l1 protein is not.

Previous analysis of β-catenin protein in breast cancer patient samples demonstrated that nuclear and cytoplasmic β-catenin was strongly associated with basal subtype tumors and

poor prognosis [208-210]. To determine if the disparity between Tcf7l1 mRNA and protein levels could be caused by elevated β-catenin, Tcf7l1 IHC results were compared to β-catenin IHC results among identical patient samples. Remarkably, tumors with strong nuclear Tcf7l1 had predominantly no nuclear or cytoplasmic β-catenin (β-catenin IHC score of 0) (Figure 3.5D), and tumors with cytoplasmic and/or nuclear β-catenin (β-catenin IHC score of 2 or 3) displayed predominantly diffuse or low levels of Tcf7l1 protein (Figure 3.5D). These data indicate that in human mammary tumors, stimulation of Wnt/β-catenin is strongly correlated with decreased nuclear Tcf7l1.

To test a causal relationship between elevated β-catenin and low Tcf7l1 protein in breast cancer, several breast cancer cell lines were used. Endogenous Tcf7l1 protein was reduced by CH in all breast cancer cells examined (MCF7, MDA-MB-231, MDA-MB-468, HS578T, HCC38) (Figure 3.5E). As in ESC, Tcf7l1 mRNA was not significantly diminished (Figure 3.6E), and the reduction of Tcf7l1 protein was blocked by the proteasome inhibitor MG-132 (Figure 3.6F). Treating cells with CH or CH +MG-132 increased cytoplasmic levels of Tcf7l1 detected by immunofluorescence (Figure 3.6G) As in ESC, Tcf7 is endogenously expressed and can stimulate TOPFlash activity (Figure 3.5F, 3.6H); overexpression of Tcf7l1 or β-catenin-Tcf7l1 fusion repressed TOPFlash activity (Figure 3.5F,G). Repression by Tcf7l1ΔN and Tcf7l1 HMG* mutants indicate that inhibition of TOPFlash was caused by a combination of β-catenin binding and DNA-binding dependent activities of Tcf7l1 (Figure 3.6I). These results suggest that the mechanism of β-catenin mediated inactivation of Tcf7l1, previously observed in ESC and mouse embryos, also occurs in human breast cancer.

Tcf7l1 inactivation occurs independently of phosphorylation by HIPK2 or NLK

Treating human breast cancer cells with CH caused a substantial shift in mobility of Tcf7l1 analyzed by SDS-PAGE (Figure 3.5E). Although more difficult to detect in mouse ESC, likely

Figure 3.5 Wnt/β-catenin inactivates Tcf7l1 protein in poorly differentiated breast cancer.

A) Heat map showing relative levels of Tcf7l1 mRNA and tumor subtype status for all individual tumors in a compendium of 1211 mammary tumors. Subtypes are displayed according to previously designation [201].

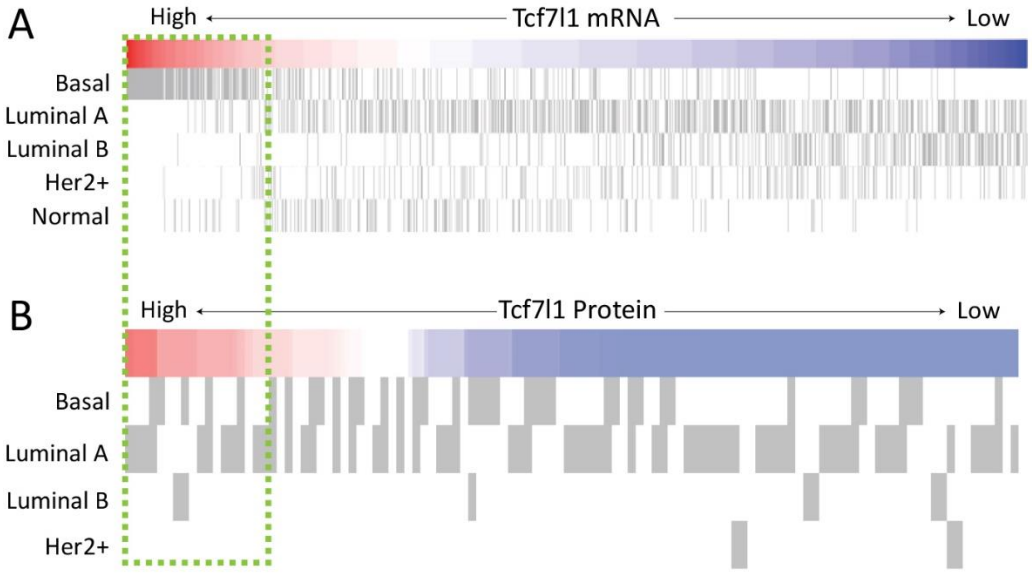
B) Heat map showing relative levels of Tcf7l1 protein nuclear immunoreactivity (intensity x area) for all individual invasive tumors in an array of samples from 71 individual patients [208]. Subtypes are displayed as they were determined previously [208].

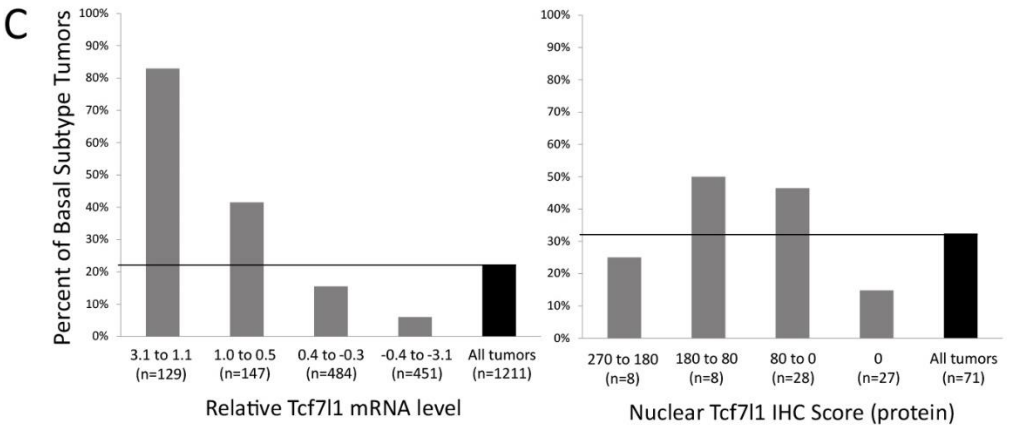
C) Graphs show distribution of basal subtype with respect to the level of Tcf7l1 mRNA (Left) or nuclear Tcf7l1 protein (Right).

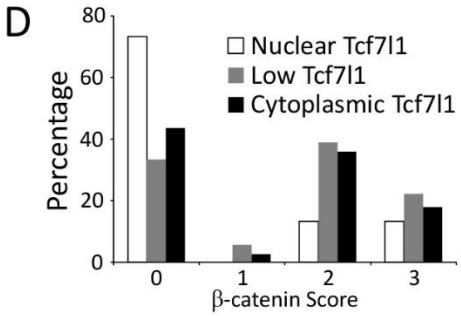
D) Distribution of tumors with nuclear Tcf7l1 (n=15; white), diffuse Tcf7l1 (n=38, black), or low Tcf7l1 (n=18, gray) classification relative to the immunohistochemistry (IHC) score for cytoplasmic β-catenin for each individual tumor. Values represent the percentage of tumors for each Tcf7l1 classification displaying the indicated cytoplasmic β-catenin IHC score.

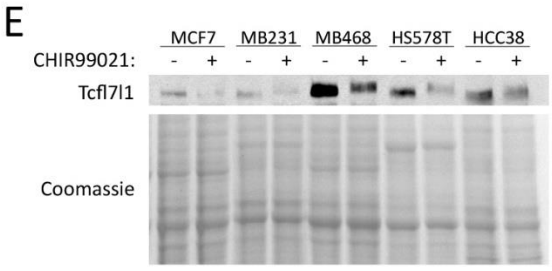
E) Western blot analysis Tcf7l1 protein levels relative to total protein (Coomassie) for the indicated breast cancer cell lines treated with vehicle (DMSO) or CH (3μM) for 12 hours.

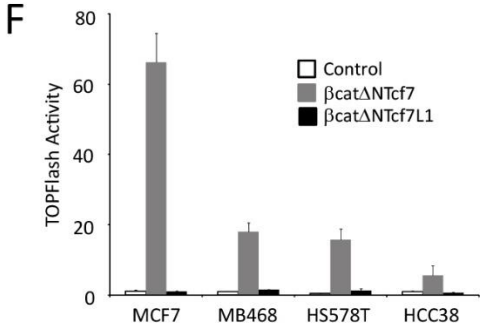
F,G) Relative SuperTOPFlash Luciferase reporter activity for breast cancer cell lines transiently transfected with the indicated β-catenin-ΔNTcf fusion plasmid, wild-type Tcf7l1 plasmid, or empty vector. Values represent mean +/- standard deviation of technical duplicates of biological duplicates.


A
High
Tcf7l1 mRNA
Low
Basal
Luminal A
Luminal B
Her2+
Normal
B
High
Tcf7l1 Protein
Low
Basal
Luminal A
Luminal B
Her2+


C
Percent of Basal Subtype Tumors
100%
90%
80%
70%
60%
50%
40%
30%
20%
10%
0%
3.1 to 1.1 (n=129)
1.0 to 0.5 (n=147)
0.4 to -0.3 (n=484)
-0.4 to -3.1 (n=451)
All tumors (n=1211)
Relative Tcf7l1 mRNA level
270 to 180 (n=8)
180 to 80 (n=8)
80 to 0 (n=28)
0 (n=27)
All tumors (n=71)
Nuclear Tcf7l1 IHC Score (protein)


D
Percentage
Nuclear Tcf7l1
Low Tcf7l1
Cytoplasmic Tcf7l1
β-catenin Score


E
MCF7
MB231
MB468
HS578T
HCC38
CHIR99021:
Tcfl7l1
Coomassie


F
TOPFlash Activity
Control
βcatΔNTcf7
βcatΔNTcf7L1
MCF7
MB468
HS578T
HCC38

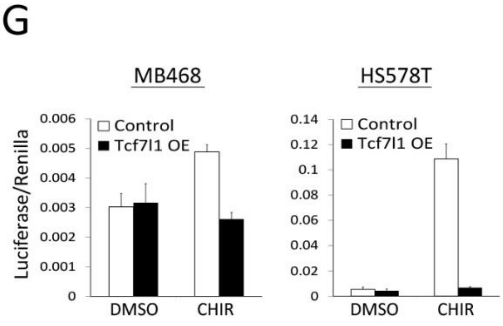

G
MB468
HS578T
Luciferase/Renilla
Control
Tcf7l1 OE
DMSO
CHIR

Figure 3.6 Tcf7l1 protein levels are reduced by Wnt/β-catenin in basal-type breast cancer.

A) Examples of Tcf7l1 immunohistochemical staining and scoring of patient samples on the breast cancer tissue microarray. Samples are arranged according to the intensity scores given for nuclear (left) and cytoplasmic (top) staining. Note: none of the patient samples had displayed a staining pattern of 0-cytoplasmic intensity score and 3-nuclear intensity score.

B) Graphs showing distribution of tumor subtypes with respect to the three classes of Tcf7l1 staining (nuclear, diffuse, low). Tumors with primarily nuclear Tcf7l1 were Luminal A. Tumors displaying diffuse Tcf7l1 were split between Luminal A and Basal subtypes. Tumors with low levels of Tcf7l1 protein were distributed among subtypes, with a preference for Luminal A subtype.

C) Each graph shows the distribution of scores for nuclear Tcf7l1 (0-3), cytoplasmic Tcf7l1 (0-3), or total Tcf7l1(0-6) for invasive tumors classified as Luminal A, Luminal B, Her2, and Basal subtypes based on marker staining [208].

D) Each graph shows the distribution of scores for nuclear Tcf7l1 (0-3), cytoplasmic Tcf7l1 (0-3), or total Tcf7l1(0-6) among the four classes of patient samples on the progressive tumor microarray corresponding to: non-tumor tissue (Normal), carcinoma in situ (DCIS), invasive tumor (Invasive), and lymph node metastases (LN Mets). Note high levels of nuclear Tcf7l1 (i.e. score of 2 or 3) predominate only in normal tissue.

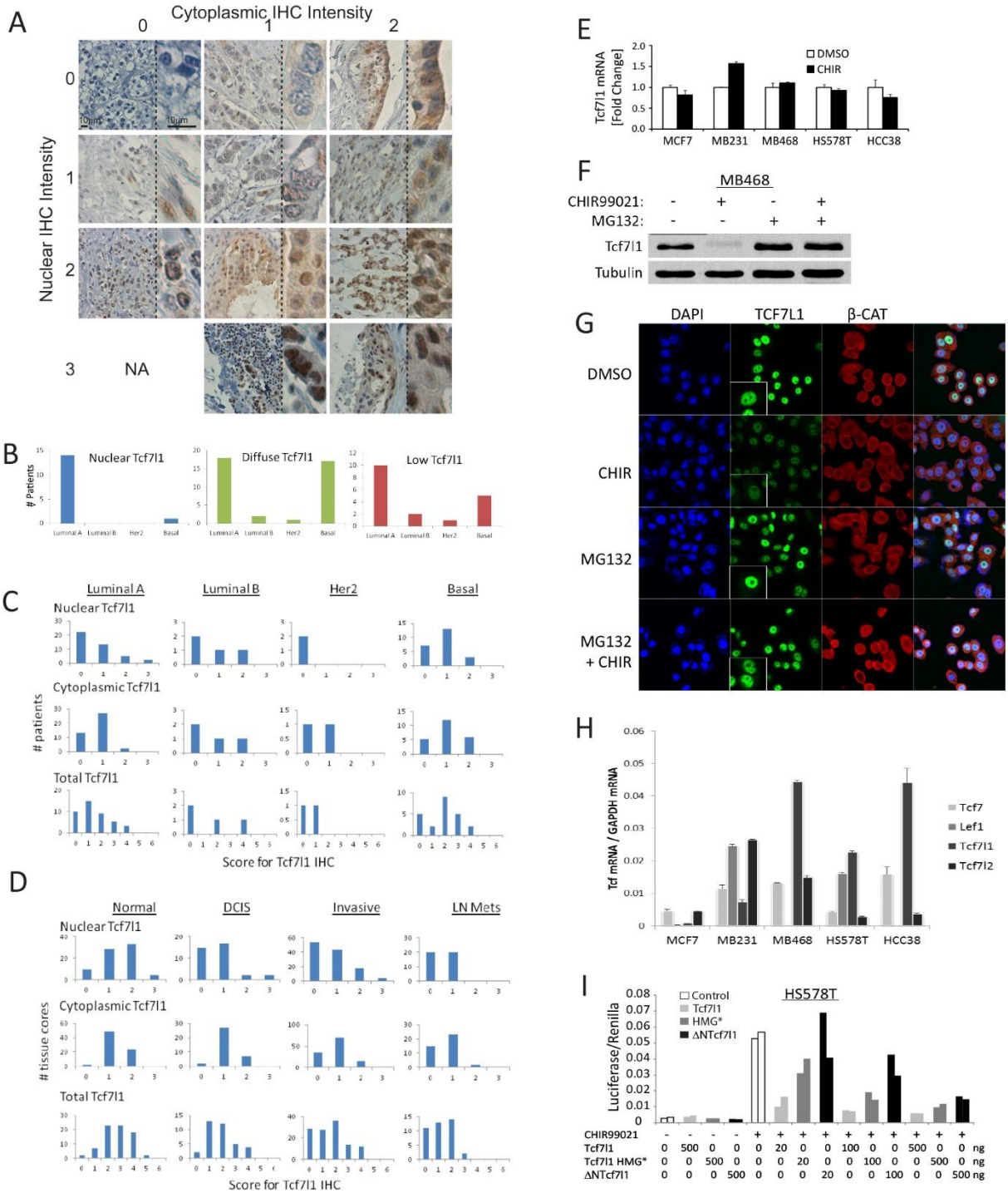
E) Quantitative RT-PCR analysis of Tcf7l1 mRNA levels after 12 hr. treatment with 3μM CH.

F) Western blot analysis of MB-468 breast cancer cells treated +/- 6 μM CH and +/- 5 μM MG132 for 12 hrs.

G) Immunofluorescence for Tcf7l1 (green), total β-catenin (red), and DAPI (blue) in MB-468 breast cancer cells treated +/- 6 μM CH and +/- 5 μM MG132 for 12 hrs. Tcf7l1 inset shows higher magnification.

H) qRT-PCR analysis of Tcf/Lef transcription factor mRNA levels in breast cancer cell lines. Tcf/Lef mRNA copy numbers were normalized to GAPDH mRNA copy numbers.

I) SuperTOPFlash activity was measured after HS-578T breast cancer cells were transiently transfected with 20-500ng of Tcf7l1, Tcf7l1 HMG*, or Tcf7l1ΔN expression plasmids. Biological duplicates are shown as two columns.


A
Cytoplasmic IHC Intensity
0
1
2
Nuclear IHC Intensity
0
1
2
3
NA
B
Nuclear Tcf7l1
Diffuse Tcf7l1
Low Tcf7l1
Patients
Luminal A
Luminal B
Her2
Basal
C
Luminal A
Luminal B
Her2
Basal
Nuclear Tcf7l1
Cytoplasmic Tcf7l1
Total Tcf7l1
patients
Score for Tcf7l1 IHC
D
Normal
DCIS
Invasive
LN Mets
Nuclear Tcf7l1
Cytoplasmic Tcf7l1
Total Tcf7l1
tissue cores
Score for Tcf7l1 IHC
E
Tcf7l1 mRNA [Fold Change]
DMSO
CHIR
MCF7
MB231
MB468
HS578T
HCC38
F
MB468
CHIR99021:
MG132:
Tcf7l1
Tubulin
G
DAPI
TCF7L1
β-CAT
DMSO
CHIR
MG132
MG132 + CHIR
H
Tcf mRNA / GAPDH mRNA
Tcf7
Lef1
Tcf7l1
Tcf7l2
MCF7
MB231
MB468
HS578T
HCC38
I
HS578T
Luciferase/Renilla
Control
Tcf7l1
HMG*
ΔNTcf7l1
CHIR99021
Tcf7l1
Tcf7l1 HMG*
ΔNTcf7l1
ng

because of multiple Tcf7l1 isoforms expressed in ESC [211], this shift also affected endogenous Tcf7l1 in mouse ESC treated with CH or Wnt3a (Figure 3.7A,B, S4). Interestingly, work in other systems demonstrated mobility shifts caused by β-catenin-dependent phosphorylation of Tcf7l1/Tcf3 proteins by homeodomain-interacting protein kinase 2 (HipK2)[88,212] and nemo-like kinase (NLK) [213-215]. In particular, HIPK2 has been proposed as a primary mediator of Tcf7l1 regulation by reducing chromatin binding after phosphorylation at conserved residues [88,212]. We previously showed that Tcf7l1 chromatin occupancy is reduced in ESC by Wnt3a, and the reduction required the Tcf7l1-β-catenin interaction [199]. Therefore, we tested whether phosphorylation of Tcf7l1 at conserved residues was needed for inactivation by using the Tcf7l1-P2/3/4 mutant, which harbors mutations at the residues phosphorylated by HipK2 and NLK [88]. Surprisingly, the Tcf7l1-P2/3/4 mutation did not affect CH-induced degradation of Tcf7l1 or Tcf7l1-repression of target gene expression in ESC (Figure 3.7C,D). Thus, it is unlikely that Wnt/β-catenin inactivation of Tcf7l1 in ESC requires phosphorylation by HipK2 or NLK. In addition, while Tcf7l1 was indeed phosphorylated, we detected no change in phosphorylation in the absence or presence of CH stimulation. The CH-induced mobility shift of Tcf7l1 was not phosphatase-sensitive, suggesting it was not mediated by increased phosphorylation (Figure 3.7B). The nature of this posttranslational modification is not known; however, it was blocked by MG-132 (Figure 3.7E), indicating that the modification required an active proteasome.

Reduction of chromatin occupancy provides the critical upstream point of Tcf7l1 regulation

To elucidate how the reduction of chromatin occupancy is causally linked to protein degradation, quantitative chromatin immunoprecipitation (ChIP) experiments measured Tcf7l1 chromatin occupancy following the combination of CH + MG-132 treatment. As expected from the changes to Tcf7l1 protein levels (Figure 3.1B), CH reduced Tcf7l1 occupancy on target genes (Axin2, Cdx1, Mycn), and MG-132 increased occupancy (Figure 3.7F). Importantly, CH treatment

reduced chromatin occupancy even when destabilization of Tcf7l1 was blocked by MG-132 (Figure 3.7F), indicating the reduction in DNA binding was upstream of degradation. Interestingly, since MG-132 prevented the mobility shift of Tcf7l1 (Figure 3.7E), this result also indicates that the reduction of chromatin binding does not require the posttranslational modification. Combined with the increased cytoplasmic Tcf7l1 staining after CH + MG-132 treatment (Figure 3.6G), these data indicate that Tcf7l1 is likely degraded after export from the nucleus.

To examine the role of chromatin occupancy in regulating Tcf7l1 stability, we used the Tcf7l1 HMG* mutation, which affects the DNA-binding HMG domain and disrupts DNA binding [198]. The HMG* mutation was sufficient to reduce Tcf7l1 protein stability in the absence of CH (Figure 3.7G). Moreover, stability of the mutant Tcf7l1 HMG* protein was not substantially decreased by CH (Figure 3.7H), indicating that destabilization of Tcf7l1 requires a change in chromatin occupancy. In support of the model focused on reduction of chromatin occupancy, co-immunoprecipitation experiments showed that Tcf7l1-β-catenin interaction was stimulated by CH + MG-132 (Figure 3.7I) and β-catenin chromatin occupancy increased (Figure 3.7J) while Tcf7l1 occupancy decreased (Figure 3.7F). Together these data are most consistent with the Tcf7l1-β-catenin interaction inhibiting chromatin occupancy as the primary effect. The secondary effect of Tcf7l1 degradation provides an additional mechanism by lowering Tcf7l1 levels, which further reduces the amount of Tcf7l1 available to bind to chromatin. Thus, the combination of reduced DNA binding and Tcf7l1 degradation combine for an additive reduction of Tcf7l1 repression in response to Wnt/β-catenin activity.

Figure 3.7 Inhibition of chromatin occupancy is upstream of Tcf7l1 protein degradation.

A) Western blot analysis of Tcf7l1 protein from wild-type ESC treated with 15μM CH for 0-2 hours.

B) Western blot analysis of Tcf7l1 protein. Cells were treated +/- 3μM CH for 12 hours. Lysates were treated +/- 15 U/μL lambda phosphatase for 30 minutes.

C) Western blot analysis from lysates of Tcf7l1-/- ESC stably expressing either Tcf7l1-WT or Tcf7l1-2/3/4. Cells treated +/- 6μM CH for 18 hours.

D) SuperTOPFlash Luciferase reporter activity of Tcf7l1-/- ESC transiently transfected with either Tcf7l1-WT or Tcf7l1-2/3/4. Approximately 200,000 cells were transfected with 0.5μg, 0.1μg, or 0.02μg vector. Cells were additionally treated +/- 3μM CH for 12 hours. Values represent mean +/- standard deviation of technical duplicates of biological duplicates.

E) Western blot analysis of Tcf7l1 protein from lysates of Tcf7l1-/- ESC stably expressing wild-type Tcf7l1. Cells were treated +/- 3μM CH and +/- 5μM MG-132 for 12 hours. Lysates treated +/- 15 U/μL lambda phosphatase for 30 minutes.

F) Quantitative chromatin immunoprecipitation using anti-Tcf7l1 antibody. Chromatin was isolated from ESC treated for 12 hours with 3μM CH and/or 5μM MG-132. qPCR measurement of Tcf7l1 bound DNA is shown for regions near AXIN2, CDX1, and MYCN genes. Values represent the mean plus standard deviation of percent of precipitated DNA relative to input for duplicate technical measurements of five biological replicates.**$p<0.01$; ***$p<0.001$; NS $p>0.05$.

G) (top) Western blot analysis of Tcf7l1 protein from lysates of Tcf7l1-/- ESC stably expressing either wild-type Tcf7l1 (WT) or the mutant Tcf7l1 HMG * (HMG*) protein. Cells were treated with 30μg/ml cyclohexamide to block new translation, and the stability of the two proteins was compared, relative to tubulin internal control, over a 12 hour periods. (bottom) Quantitation of Western blot and normalization of Tcf7l1 protein levels was calculated for WT (blue) and HMG* (red) proteins. Each data point represents the mean of biological triplicates.

H) Western blot analysis comparing the CH-mediated reduction WT and HMG* Tcf7l1 proteins using same ESC as (G).

I) Co-IP experiments using anti-Tcf7l1 (top) or anti-β-catenin (middle). Protein was immunoprecipitated from lysates of cells treated +/- 3 μM CHIR99021 and +/- 5 μM MG-132 (bottom).

J) Quantitative chromatin immunoprecipitation using anti-β-catenin antibody. Chromatin was isolated from ESC treated for 12 hours with 3μM CH and/or 5μM MG-132. qPCR measurement of β-catenin bound DNA is shown for regions near AXIN2 and CDX1 genes. Values represent the mean plus standard deviation of percent of precipitated DNA relative to input for duplicate technical measurements of 3 biological replicates.

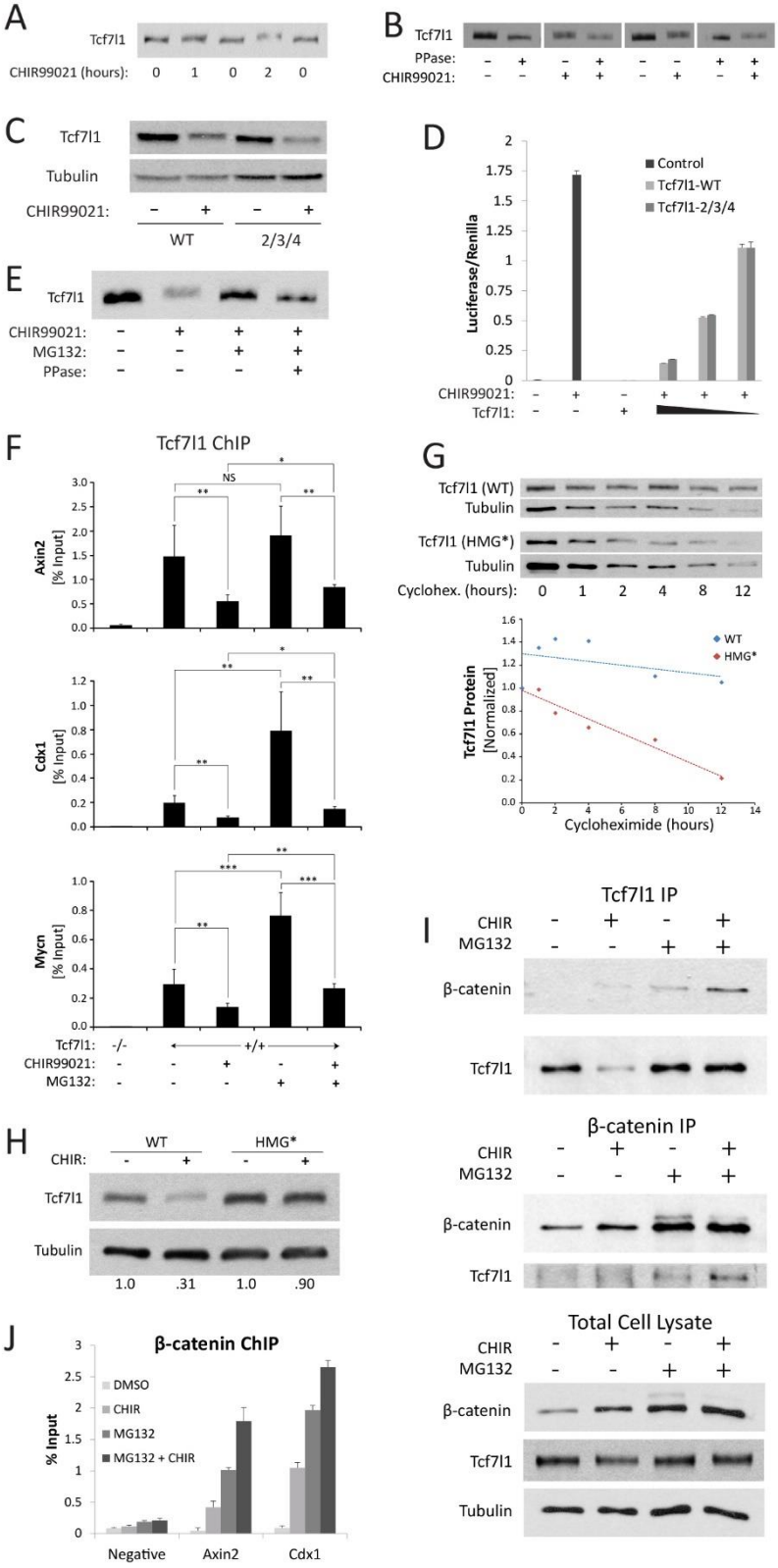

A
Tcf7l1
CHIR99021 (hours): 0 1 0 2 0
B
Tcf7l1
PPase:
CHIR99021:
C
Tcf7l1
Tubulin
CHIR99021:
WT
2/3/4
D
Control
Tcf7l1-WT
Tcf7l1-2/3/4
Luciferase/Renilla
CHIR99021:
Tcf7l1:
E
Tcf7l1
CHIR99021:
MG132:
PPase:
F
Tcf7l1 ChIP
Axin2 [% Input]
Cdx1 [% Input]
Mycn [% Input]
NS
Tcf7l1: -/- +/+
CHIR99021:
MG132:
G
Tcf7l1 (WT)
Tubulin
Tcf7l1 (HMG*)
Tubulin
Cyclohex. (hours): 0 1 2 4 8 12
WT
HMG*
Tcf7l1 Protein [Normalized]
Cycloheximide (hours)
H
WT
HMG*
CHIR:
Tcf7l1
Tubulin
1.0 .31 1.0 .90
I
Tcf7l1 IP
CHIR
MG132
β-catenin
Tcf7l1
β-catenin IP
CHIR
MG132
β-catenin
Tcf7l1
Total Cell Lysate
CHIR
MG132
β-catenin
Tcf7l1
Tubulin
J
β-catenin ChIP
DMSO
CHIR
MG132
MG132 + CHIR
% Input
Negative
Axin2
Cdx1

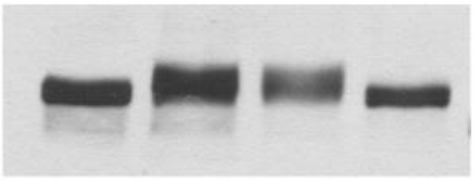
Figure 3.8 Post-translational modification of Tcf7l1 in response to Wnt/β-catenin and DNA binding.

A) Western blot analysis of endogenous Tcf7l1 from wild-type ESC treated +/- Wnt3a for 12 hours. Lysates were treated +/- 15 U/µL lambda phosphatase for 30 minutes. Endogenous Tcf7l1 is phosphorylated both in the presence and absence of Wnt3a and undergoes an upward mobility shift in response to Wnt3a.

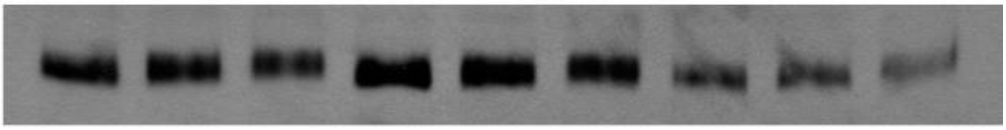
B) Western blot analysis of endogenous Tcf7l1 from wild-type ESC treated +/- Wnt3a or CH for 12 hours. Cells were grown +/- Serum and +/- LIF for 24 hours. Tcf7l1 mobility on SDS-PAGE is shifted by Wnt3a and CH independently of serum or LIF.

C) Western blot analysis of Tcf7l1 protein from lysates of Tcf7l1-/- ESC stably expressing either wild-type Tcf7l1 (WT) or the mutant Tcf7l1 HMG * (HMG*) protein. Cells were treated +/- CH for 12 hours. Lysates were treated +/- 15 U/µL lambda phosphatase for 30 minutes. Tcf7l1 phosphorylation and CH induced mobility shift both require DNA binding.

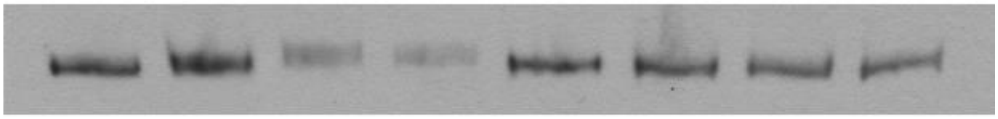
A


Tcf7l1
PPase: + – – +
Wnt3a: – – + +

B


Tcf7l1
LIF: – – – + + + + + +
Serum: + + + + + + – – –
Wnt3a: – + – – + – – + –
CHIR99021: – – + – – + – – +

C


Tcf7l1
Lane: 1 2 3 4 5 6 7 8
PPase: + – – + + – – +
CHIR99021: – – + + – – + +
WT
HMG*

Discussion:

The molecular effects of Wnt/β-catenin and Gsk3-inhibition on Tcf7l1 described here indicate that Tcf7l1 primarily functions outside of the classic model of canonical Wnt/β-catenin signaling. In ESC, inactivation of Tcf7l1 did not require phosphorylation of Tcf7l1 at conserved sites, and β-catenin was sufficient to reduce Tcf7l1 levels without exogenous pathway stimulation. These results were consistent with a novel mechanism of inactivation wherein β-catenin binding inhibits Tcf7l1-repression by reducing chromatin occupancy, consequently stimulating its degradation. This mechanism provides a simple explanation for the controversial pro-self renewal effects of the β-cateninΔC mutant in ESC [13,97,184]; the critical effect of inactivating Tcf7l1 is stimulated by the β-cateninΔC form, thus making β-catenin's C-terminal transactivation domain dispensable in ESC.

Experiments using human breast cancer tumors, breast cancer cell lines, and mouse genetics combine to indicate that inactivation of Tcf7l1 is the predominant mechanism whereby Wnt/β-catenin signaling interacts with this mammalian Tcf/Lef protein. The viability of Tcf7l1-/ΔN mice genetically demonstrates that inactivation of Tcf7l1 is the only effect of Tcf7l1-β-catenin binding required for normal mouse development and life. That said, additional activities downstream of Tcf7l1-β-catenin interaction likely exist. Indeed, reporter gene assays support rare Tcf7l1-β-catenin transactivator activity in some cell types, 293T, COS7 and human keratinocytes [198,199,207]; however, a biological significance for this effect remains to be determined.

Human breast cancer provides one important context where Tcf7l1-based activation has been suggested (Slyper et al., 2012). Basal subtype tumors are particularly relevant because they have been noted to share a gene expression signature with ESC [201] and fetal mammary stem cells [216]. These tumors, which express high levels of Tcf7l1 mRNA, have been suggested to arise following reprogramming to an earlier embryonic stage [217], making it important to

understand how Wnt/β-catenin and Tcf7l1 function. Previous direct experiments showed that ectopic Tcf7l1 expression and Wnt3a both stimulate xenograft tumor formation, mammosphere formation and colony formation in Matrigel from breast cancer cell lines [207]. Although it is formally possible that Tcf7l1-β-catenin complexes may act as transactivators for a set of target genes critical for breast cancer cells, data presented here do not support this possibility. Tcf7l1 displayed only repressor activity in reporter assay experiments, and Tcf7l1 was degraded following CH treatments. We propose two non-mutually exclusive possibilities of either, 1. Tcf7l1 and Wnt/β-catenin signaling mediate parallel effects, each stimulating tumor cells, or 2. Tcf7l1-β-catenin complexes have a novel biochemical activity distinct from the classic transactivator activity. The former is supported by recent demonstration of a Wnt/Gsk3/Slug/Snail signaling axis affecting triple negative breast cancers [218].

The β-catenin effects on Tcf7l1 are most parsimoniously explained by a mechanism of β-catenin directly inhibiting Tcf7l1 binding to chromatin. Experiments examining effects of β-catenin on Tcf/Lef interaction with naked DNA have shown there to be little or no effect on binding in vitro. By contrast, β-catenin interaction significantly affects binding to chromatin of the Lef1 protein [219]. Interestingly, mutational analysis of Lef1 indicated that a region in the Lef1 amino terminus provides an intramolecular inhibition of chromatin binding. β-catenin binding blocks the intramolecular inhibition, thus stimulating Lef1 binding to chromatin [219]. Although the effect of β-catenin on Lef1 is different than predicted for Tcf7l1, these previous findings demonstrate both positive and negative regulation of chromatin binding via the β-catenin interaction region of Tcf/Lef proteins. Further research is necessary to elucidate the biophysical and biochemical nature of β-catenin effects on the chromatin binding properties of Tcf7l1.

Experimental Procedures:

Immunohistochemistry Staining and Scoring of Mammary Tumor Microarray

Quantitative analysis (i.e. scoring) of IHC was performed without knowledge of specimen identification. Scoring was based on the combination of intensity of stained cells and percentage of tissue. Separate values for nuclear and cytoplasmic Tcf7l1 immunoreactivity were determined for each sample. Total Tcf7l1 IHC scores were calculated using a modified Reiner scoring system [220] by multiplying the intensity of staining (0-3 value) by the percentage of positive cells. Ranking and scores for β-catenin levels and localization were previously described in Khramstov et al [208].

Statistical Analyses of Tumor RNA and Protein Expression Data

The Kruskal-Wallis test was used to compare overall effect of stage on nuclear, cytoplasmic, and total Tcf7l1. The Wilcoxon Rank-Sum test with bonferroni correction was used to conduct pairwise comparisons among four stages. Two sample t-test was used to compare Tcf7l1 mRNA between basal and non-basal groups. The Wilcoxon Rank-Sum test was used to compare Tc7l1 nuclear, cytoplasmic, and total IHC-score between basal and non-basal groups and between luminal A and non-luminal A groups. Spearman correlations determined the relationships between nuclear Tcf7l1 and cytoplasmic β-catenin. P-values <0.05 were considered statistically significant.

Acknowledgements:

We thank Jackson Hoffman, Yuka Shimizu, and Alan Tseng for helpful discussions and work associated the manuscript, Weihua Gao at the UIC Center for Clinical and Translational Science (CCTS) for assistance with biostatistical services, Liza Benevolenskaya for assistance with tumor microarray studies, Sergei Sokol for kindly providing the Tcf3-P2/3/4 mutant DNA, and Rudolf Grosschedl for kindly provided β-cateninΔC plasmid DNA. The work was funded by grants from the National Institutes of Health (RO1-CA128571, B.J.M. and UL1TR000050, UIC CCTS) and the American Cancer Society Illinois Division (215889, K.H.G.).

Chapter 4: Co-incident insertion enables high efficiency genome engineering in mouse embryonic stem cells.

*****With the exception of several formatting changes, the content of chapter has been published in the following manuscript:

Shy, B.R., MacDougall M.S., Clarke R., and Merrill, B.J. "Co-incident insertion enables high efficiency genome engineering in mouse embryonic stem cells". Nucleic Acids Res., September 2016.

Abstract:

CRISPR/Cas9 nucleases have enabled powerful, new genome editing capabilities; however, the preponderance of non-homologous end joining (NHEJ) mediated repair events over homology directed repair (HDR) limits the ability to engineer precise changes in mammalian genomes. Here, we increase efficiency of generating precise HDR-mediated events in embryonic stem cells (ES) by more than 20-fold through the use of co-incidental insertion (COIN) of independent donor DNA sequences. COIN uses the insertion of a selectable marker at one control site to greatly enrich for cells with the desired precise HDR-mediated event at a second, independent site. In addition to providing a simple approach for efficiently engineering stem cell genomes, the COIN effect exemplifies a co-occurrence of independent events, which can be rationally applied to other aspects of genome engineering.

Introduction:

The recent adaptation of RNA-guided nucleases, such as *S. pyogenes* Cas9, for use in mammalian cells has yielded substantial benefits for stem cell and regenerative medicine research. The Cas9 protein uses structural RNAs from a clustered regularly interspaced short palindromic repeat (CRISPR) locus to target its nuclease activity to a DNA sequence [221]. The first 20bp of the sgRNA directs Cas9 nuclease activity to a DNA target by Watson-Crick base pairing, generating a double strand break (DSB) at that site [167,182]. This, in turn, stimulates endogenous DNA repair pathways, which can be harnessed for local genome editing. In mammalian cells, an engineered single guide RNA (sgRNA) is used in place of the processed CRISPR RNA [111,168]. This process has been used in stem cells to generate new cell models of genetic disease and to repair pathologic mutations in patients' stem cells for diverse human diseases including cancers, muscular dystrophy, HIV, and beta-thalassemia [222-227]. The profound therapeutic potential of this system depends on its ability to efficiently and safely engineer changes to genomes of patients' stem cells.

Precise genomic changes are made by using homology directed repair (HDR) to incorporate a donor DNA sequence at the DSB. In mammalian cells, this occurs at very low rates because Cas9-generated DSBs are far more frequently repaired by non-homologous end joining (NHEJ) than HDR [111,228]. To overcome the high frequency of NHEJ events when seeking HDR events, either a large number of clones need to be screened or the HDR-dependent events need to be enriched in a population of cells. A positive selection marker gene is commonly added to the donor DNA sequence to enrich for cells altered by HDR. Unfortunately, this leaves a heterologous DNA sequence inserted at or near the site of interest, which is incompatible with many applications of genome editing. In the absence of a positive selection marker gene, the process of screening, isolating, and expanding desired HDR events away from

the undesired NHEJ-mediated events constitutes a substantial barrier to successful genome engineering approaches.

Recent studies have identified some enhancements for stimulating HDR repair of Cas9-generated DSB in pluripotent stem cells. The greatest advances thus far involve using small single stranded oligos (ssODN) as donor DNAs. A screen for small molecules identified two compounds (Brefeldin A and L755507) that stimulated a 9-fold increase for insertion of a point mutation by ssODN, as well as a 2- and 3-fold increase in the frequency of insertion of GFP into the Nanog locus using a donor DNA with large arms of homology [229]. It remains unclear how this effect is caused by Brefeldin A, an inhibitor of ER to Golgi transport, or L755507, a β3-adrenergic receptor agonist. Delivering active Cas9 ribonucleoproteins (RNPs) directly into cells synchronized in the G2/M phase of the cell cycle also increased HDR events over 6-fold with a ssODN as the donor DNA. Use of a small linear dsDNA yielded a more modest 2-fold enrichment from timed RNP delivery [169]. These advances increase the utility of some HDR-mediated genome edits, but they are restricted by the limited capacity of ssODN to about 50bp insertions.

To develop technologies and methods for making HDR-mediated genome engineering more efficient in stem cells, we focused on optimizing the use of relatively easily generated dsDNA fragments as donor DNA. Initial experiments elucidating effects of homology arm length and donor concentration on the frequency of HDR-mediated insertion revealed an unexpectedly high frequency of cells with insertions at both alleles when a single gene was targeted. Importantly, similar to the abundant biallelic events at a single genomic site, a coincidental insertion (COIN) effect also occurred when unlinked genes were simultaneously targeted by distinct sgRNA/donor DNA combinations. Thus, use of positive selection for an HDR event at a safe harbor, such as Rosa26, provides substantial enrichment of HDR events at other genomic sites without the need for additional selectable markers. Using standard liposome-mediated

transfection of common plasmids and simple linear donor DNAs, we observed up to 20-fold increase in efficiency. In summary, the COIN effect makes it possible to expect 10-20% of transfected ES cells to contain the precisely altered locus.

Results:

To test the effect of donor DNA characteristics on insertion efficiency, we focused initial experiments on one genomic site in the Lef1 gene using a sgRNA sequence that we previously validated to function for NHEJ-based indel mutagenesis (Fig 4.1A). The frequency of HDR insertion of a PGK-Neo selection cassette at this Lef1 site was determined with a quantitative real time PCR (qPCR) assay, which accurately measures the number of Cas9-dependent insertion events at a frequency as low as one insertion per 1000 genomes (Fig 4.1B). The benefit of Cas9 and sgRNA was measured using the qPCR assay and exemplified by the 140-fold increased recovery of Lef1::PGK-Neo alleles (Fig 4.1C).

For initial experiments, the homology arm length was set near 200bp for two reasons. First, the qPCR assay is specific for on-target insertion of donor DNAs with a left arm of homology 168bp or less (Fig 4.1A). Second, a 200bp homology arm length has significant practical relevance, because it allows for attachment of homology arms by PCR amplification with primers possessing long 5' overhangs (Experimental Procedures). Therefore, we reasoned that methods enabling efficient insertion of donor DNAs with homology arms of about 200bp would provide significant benefits in their ease of use, simplicity of generation, and potential application to high-throughput experiments.

The effect of donor DNA concentration was assessed for on-target insertion at Lef1 and off-target insertions at unknown sites. Genomic DNA was isolated from cells transfected with the Cas9/sgRNA expression plasmid plus a Lef1-Neo donor DNA (Fig 4.1A,B). On-target insertions were directly measured by qPCR, and off-target insertions were indirectly determined by subtracting the percentage of on-target insertions from the total G418-resistant genome equivalents. Using 168/181bp arms of homology, the number of on-target insertions was proportional to the amount of donor DNA transfected into cells (Fig 4.1C). The highest level of insertion (2.5% of transfected cells at 1pmol donor DNA) did not appear to have reached

Figure 4.1 Quantitative PCR (qPCR) assay for optimization of HDR-mediated insertions

(A) Schematic of mouse *Lef1* gene showing sgRNA target sequence (green), PAM site (red), and beginning of arms of homology (**bold**). Locations of exons 7 and 8 are denoted with boxes. The size (bp) of homology arms and the Lef1-Neo DNA donor with a PGK-Neo selection cassette (blue) are indicated below each feature. Locations of primers used for quantitative PCR assay are noted.

(B) qPCR detection of varying ratios of genomic DNA from WT cells mixed with a heterozygous Lef1::PGK-Neo clone. Clones used for ladder generation are shown in Figure 4.2A.

(C) Percentage of on-target insertions following transfection of Cas9, donor DNA, and sgRNA from Figure 4.1A were determined using qPCR assay from Figure 4.1B after positive selection with G418 media. Bars represent mean +/- SD for biological and technical duplicates.

(D,E) qPCR analysis of increasing donor DNA concentration on on-target insertions among all transfected cells (D) or only those cells surviving G418 selection (E). Bars represent mean +/- SD for biological and technical duplicates.

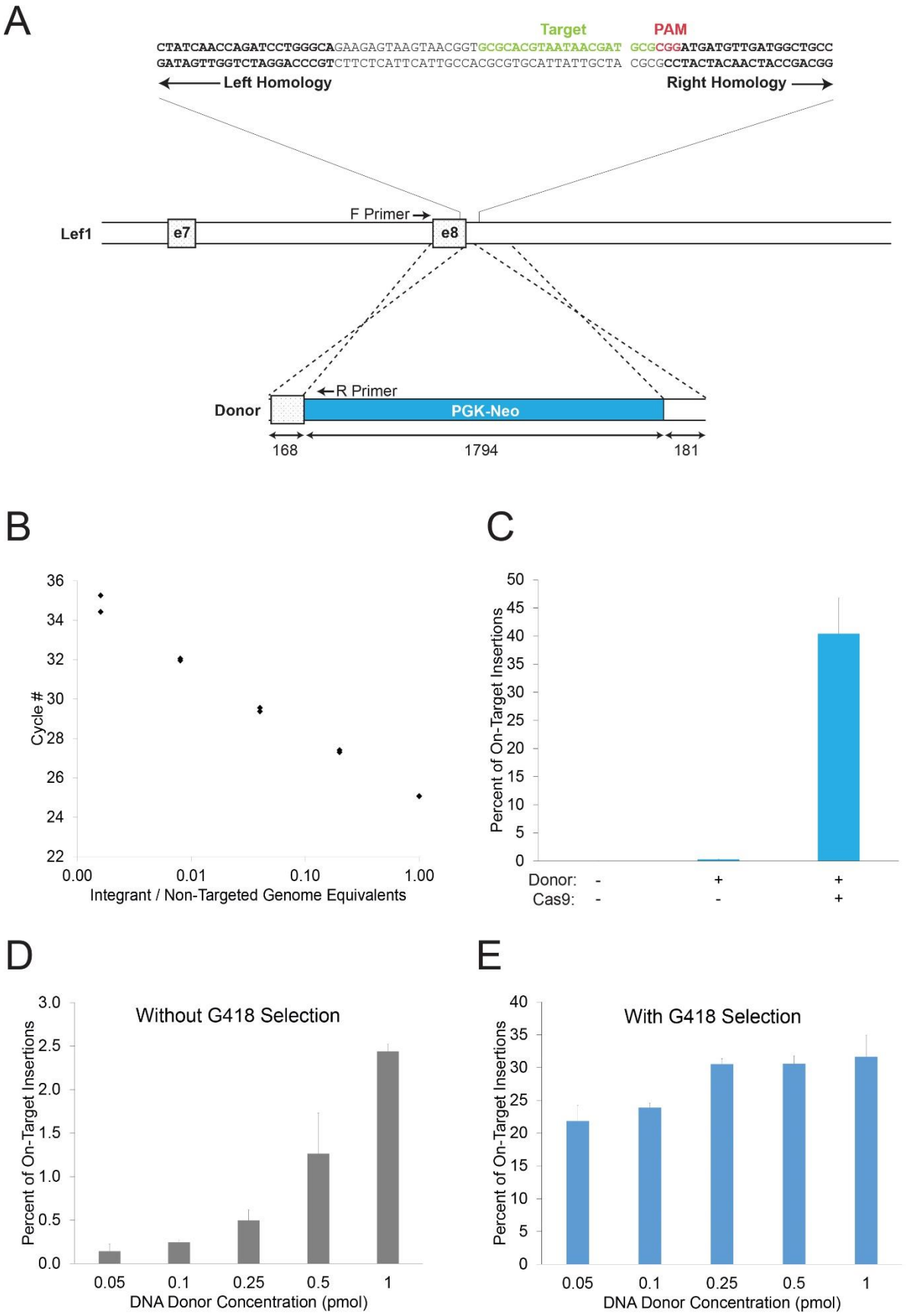
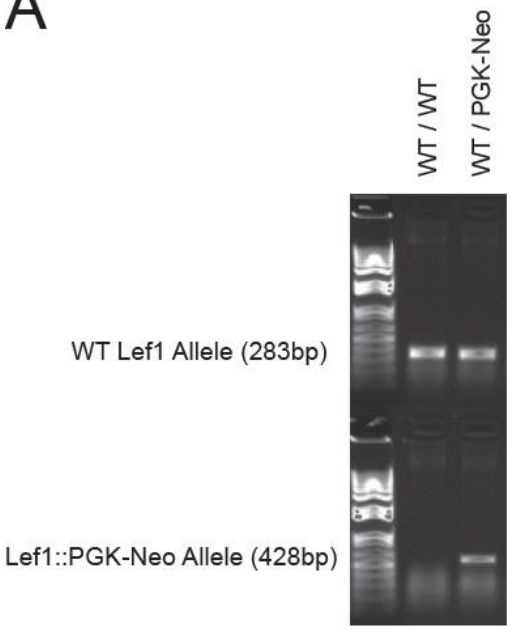

A
Target
PAM
CTATCAACCAGATCCTGGGCAGAAGAGTAAGTAACGGTGCGCACGTAATAACGAT GCGCGGATGATGTTGATGGCTGCC
GATAGTTGGTCTAGGACCCGTCTTCTCATTCATTGCCACGCGTGCATTATTGCTA CGCGCCTACTACAACTACCGACGG
Left Homology
Right Homology
F Primer
Lef1
e7
e8
R Primer
Donor
PGK-Neo
168
1794
181
B
Cycle #
Integrant / Non-Targeted Genome Equivalents
C
Percent of On-Target Insertions
Donor:
Cas9:
D
Without G418 Selection
Percent of On-Target Insertions
DNA Donor Concentration (pmol)
E
With G418 Selection
Percent of On-Target Insertions
DNA Donor Concentration (pmol)

Figure 4.2 PCR genotyping for wild-type and heterozygous Lef1::PGK-Neo clone

(A) PCR genotyping for wild-type and heterozygous Lef1::PGK-Neo clone used in the generation of the quantitative PCR ladder illustrated in figure 4.1B.

A


WT / WT
WT / PGK-Neo
WT Lef1 Allele (283bp)
Lef1::PGK-Neo Allele (428bp)

saturation, but was limited by the amount of DNA that could be used without causing toxicity. By contrast, the ratio of on-target to off-target insertions did not increase as substantially when more donor DNA was added (Fig 4.1D). Thus, increase in donor DNA enabled a greater number of transfected cells to insert the donor DNA into their genome, but had only a minor effect on where the insertion occurred.

Homology Arm Length and Insert Size Contribute Opposing Effects on Usage of Donor DNA for HDR-mediated Insertion at Cas9-generated DSB

Although decades of gene targeting experiments without RNA guided nucleases has led to the general consensus that longer arms of homology and smaller sized inserts between those arms are beneficial for gene targeting, the relative benefits of each of these donor DNA characteristics has not been clearly defined using Cas9 in ES cells. To determine their effects on on HDR-mediated insertion, we focused on events at one Cas9 site and tested a wide range of donor DNAs. For donors with left homology arms of 168bp or smaller, qPCR was used to determine insertion frequencies (Fig 4.1A,B).

Keeping homology arms constant at 168/181bp, the distance between homology arms was varied from 1794bp to 21bp (Fig 4.4A,B). Insertion frequency did not significantly change as insert size was reduced from 1794bp to 500bp, but the frequency logarithmically increased as insert size decreased below 500bp (Fig 4.3B). Combined with Cas9 and sgRNA expression, it was possible to obtain insertions of 1794bp with as little as 25bp homology arms (Fig 4.3C,D). Increasing the length of homology arms led to substantial increases in targeting efficiency, with a notable jump at arm lengths greater than 100bp (Fig 4.3C). Unlike experiments with varying donor DNA concentration (Fig 4.1D,E), increasing homology arm length did not lead to higher off-target integration frequency, resulting in an overall rise in ratio of on:off-target insertions in

transfected cells (Fig 4.3D). Presumably, total number of off-target integrations is affected by the amount of donor DNA and not homology arm length.

The HDR-mediated insertion of donors with large homology arms was incompatible with qPCR analysis. For experiments assessing longer arms of homology, PCR analysis of isolated clones transfected with donor DNA harboring a PGK-Neo cassette was used to determine ratio of on:off-target insertions (Fig 4.1A, 4.3E,F). Experiments also used flow cytometry of populations of cells transfected with a donor DNA harboring a GFP expression cassette to determine frequencies of overall insertion events (Fig 4.3G,H). The on:off-target ratio (Fig 4.3F) for each homology arm length was applied to its overall insertion frequency (Fig 4.3H) to generate separate on-, off-, and mis-targeted insertion frequencies (Fig 4.3I). Similar to the trend exhibited by shorter arms, increasing the arm length up to 1443/1471bp led to more frequent on-target insertions with no change in off-target insertions (Fig 4.3I). This is consistent with off-target insertions occurring primarily through homology-independent repair.

Frequent Co-occurrence of Independent HDR-mediated Insertions at Cas9-generated DSB

Analysis of individual clones revealed an unexpected relationship between the frequencies of monoallelic and biallelic Lef1::Neo insertions (Fig 4.3E). A priori, one would expect the two alleles to be targeted independently, and thus, biallelic insertions would be expected to occur with a frequency equal to the product of two monoallelic insertions. In contrast, biallelic insertions were recovered at 7- to 14-fold higher than their predicted frequency (Fig 4.3J). Interestingly, other groups have noted exceptionally high frequencies of biallelic events relative to monoallelic events. Byrne et al reported a high frequency of biallelic insertions at Cas9-generated DSB using donor DNAs with large arms of homology (>2kb) in human induced

Figure 4.3 Effect of homology arm length on frequency of HDR-mediated donor DNA insertion

(A) Schematic of the parental Lef1 locus surrounding exon 7/8 (left) and Lef1::Neo allele after on-target insertion of Lef1-Neo DNA donor (right). PCR primer locations and corresponding PCR products are illustrated. Blue color denotes specificity for the Lef1::Neo allele.

(B) Effect of insert size was determined by qPCR for inserts from 21bp up to 1794bp with a constant homology arm length (168/181bp). DNA donors are shown in Figures 4.4A,B.

(C,D) Effect of homology arm length on targeting efficiency for nearly symmetrical arms from 25bp to 202bp was determined by qPCR for all transfected cells (C) or only those surviving G418 selection (D). Bars represent mean +/- SD for biological and technical duplicates. DNA donors are illustrated in Figure 4.4C and PCR products are shown in Figure 4.4D.

(E) PCR genotyping for 48 G418-resistant clones obtained using donor DNA with indicated homology arm lengths. A clone was identified as having a monoallelic on-target insertion if F1/R1 and F2/R2 primer pairs generated 1.0kb and 1.7kb bands, respectively. Biallelic clones were identified by F1/R1 primers producing a 2.8kb band only and F2/R2 generating a 1.7kb band. Clones with a single band from F2/R2 primers size were counted as a mis-target when the band was not 1.7kb (#2,31). DNA donors are shown in Figure 4.4C,D.

(F) Off-target insertion was calculated as the percentage of clones lacking either on- or mis-target bands from panel (E).

(G,H) Insertion of Lef1-GFP donor DNA with the indicated homology arm lengths was measured by flow cytometry (bottom). Control lacking donor DNA is shown in Figure 4.4E (0.04%). Representative images of cells in culture are shown by DIC (top row) and green fluorescence (middle row). Lef1-GFP insertion frequency relative to all transfected cells is shown in (H) as the mean +/- SD of 6 separate measurements for each homology arm length (biological duplicates from 3 independent experiments). Experiments are shown individually in Figure 4.4F.

(I) Calculated on-, mis-, and off-target rates for donor homology arm lengths of 208/202, 459/525, 1001/1035, and 1443/1471 bp. The total GFP+ cells from panel (H) were divided into subsections based on the ratios shown in panel (F).

(J) Discrepancy between predicted and observed frequencies of biallelic insertions. The expected frequency of biallelic is the observed frequency of on-target insertions in cells not subjected to selection from panel (I). Fold increase is calculated by dividing observed frequency by predicted frequency of the second Lef1::Neo allele.

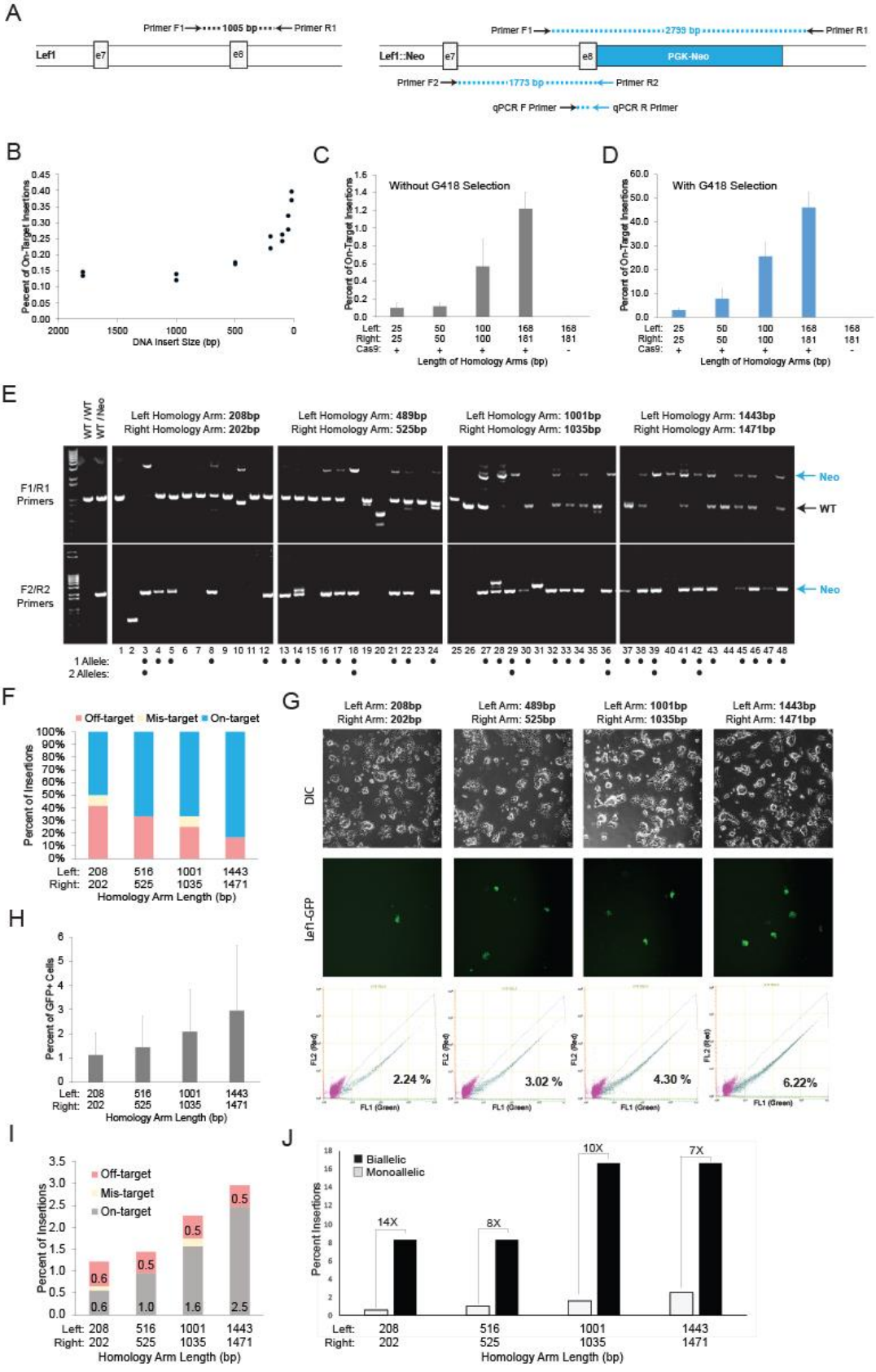

A
Primer F1
1005 bp
Primer R1
Lef1
e7
e8
Lef1::Neo
2739 bp
PGK-Neo
Primer F2
1773 bp
Primer R2
qPCR F Primer
qPCR R Primer
B
Percent of On-Target Insertions
DNA Insert Size (bp)
C
Without G418 Selection
Percent of On-Target Insertions
Left: 25 50 100 168 168
Right: 25 50 100 181 181
Cas9: + + + + -
Length of Homology Arms (bp)
D
With G418 Selection
Percent of On-Target Insertions
Left: 25 50 100 168 168
Right: 25 50 100 181 181
Cas9: + + + + -
Length of Homology Arms (bp)
E
WT /WT
WT /Neo
Left Homology Arm: 208bp
Right Homology Arm: 202bp
Left Homology Arm: 489bp
Right Homology Arm: 525bp
Left Homology Arm: 1001bp
Right Homology Arm: 1035bp
Left Homology Arm: 1443bp
Right Homology Arm: 1471bp
F1/R1 Primers
F2/R2 Primers
Neo
WT
Neo
1 Allele:
2 Alleles:
F
Off-target
Mis-target
On-target
Percent of Insertions
Left: 208 516 1001 1443
Right: 202 525 1035 1471
Homology Arm Length (bp)
G
Left Arm: 208bp
Right Arm: 202bp
Left Arm: 489bp
Right Arm: 525bp
Left Arm: 1001bp
Right Arm: 1035bp
Left Arm: 1443bp
Right Arm: 1471bp
DIC
Lef1-GFP
2.24 %
3.02 %
4.30 %
6.22%
FL1 (Green)
H
Percent of GFP+ Cells
Left: 208 516 1001 1443
Right: 202 525 1035 1471
Homology Arm Length (bp)
I
Off-target
Mis-target
On-target
Percent of Insertions
0.6
0.5
0.5
0.5
0.6
1.0
1.6
2.5
Left: 208 516 1001 1443
Right: 202 525 1035 1471
Homology Arm Length (bp)
J
Biallelic
Monoallelic
Percent Insertions
14X
8X
10X
7X
Left: 208 516 1001 1443
Right: 202 525 1035 1471
Homology Arm Length (bp)

Figure 4.4 Generation of linear dsDNA donors for HDR

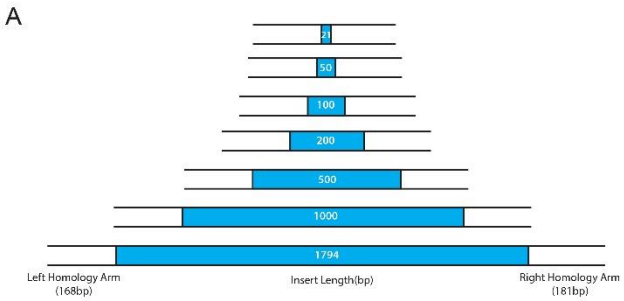
(A and B) DNA donor templates used in figure 4.3B. Homology arm length was held constant at 168 bp on the left and 202 bp on the right. Insert length was varied from 21 bp up to 1794 bp. Schematic (A) and PCR products (B) are shown.

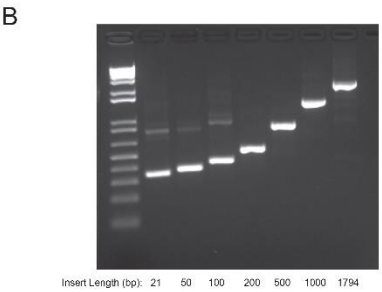
(C and D) DNA donor templates used in figure 4.3C-E. Homology arm length was varied while holding insert length constant. Arm lengths are illustrated in the schematic (C) and PCR products are shown (D).

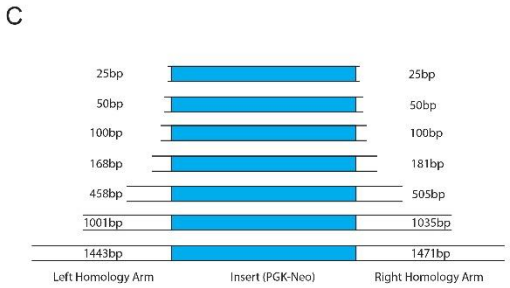
(E) Flow cytometry of no donor control for figure 4.3G. Shows background level of GFP+ cells (0.04%).

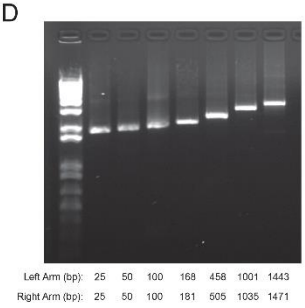
(F) Flow cytometry measurements for 3 independent experiments varying Lef1-CMV-GFP donor homology arms. Converted to mean in figure 4.3H. Pairs represent biological duplicates.

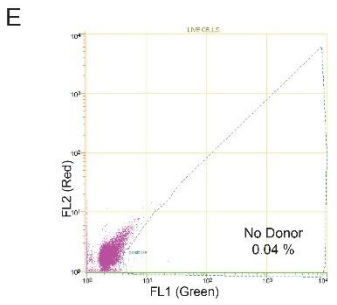
(G) DNA donor templates used in figure 4.3F-H. Sixteen combinations of homology arms targeting the Lef1 locus were generated flanking a CMV-GFP insert.

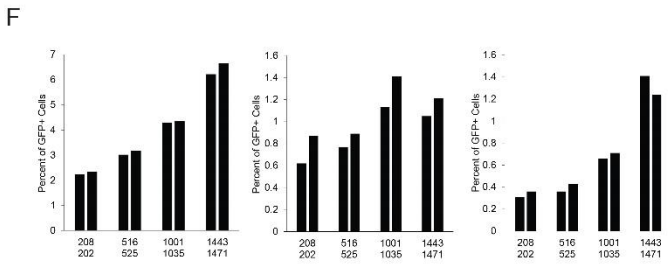

A
21
50
100
200
500
1000
1794
Left Homology Arm
(168bp)
Insert Length(bp)
Right Homology Arm
(181bp)

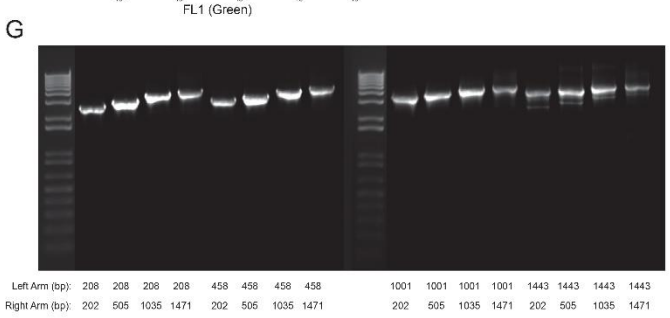

B
Insert Length (bp): 21 50 100 200 500 1000 1794

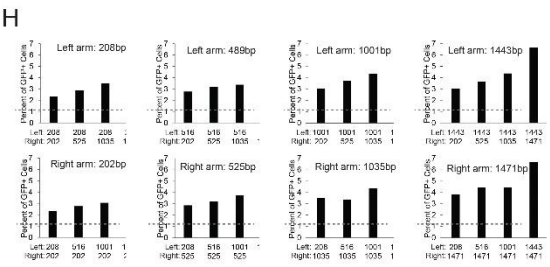

C
25bp
50bp
100bp
168bp
458bp
1001bp
1443bp
25bp
50bp
100bp
181bp
505bp
1035bp
1471bp
Left Homology Arm
Insert (PGK-Neo)
Right Homology Arm


D
Left Arm (bp): 25 50 100 168 458 1001 1443
Right Arm (bp): 25 50 100 181 505 1035 1471


E
FL2 (Red)
FL1 (Green)
No Donor
0.04 %


F
Percent of GFP+ Cells
208
202
516
525
1001
1035
1443
1471


G
Left Arm (bp): 208 208 208 208 458 458 458 458 1001 1001 1001 1001 1443 1443 1443 1443
Right Arm (bp): 202 505 1035 1471 202 505 1035 1471 202 505 1035 1471 202 505 1035 1471


H
Left arm: 208bp
Left arm: 489bp
Left arm: 1001bp
Left arm: 1443bp
Right arm: 202bp
Right arm: 525bp
Right arm: 1035bp
Right arm: 1471bp

pluripotent stem cells [230]. Canver et al reported frequent biallelic deletions between two Cas9-generated DSB in murine erythroleukemia (MEL) cells [231]. Given the occurrence of biallelic events in diverse experimental systems, we reasoned that the phenomena could be caused by a common mechanism, and that elucidating the mechanism could lead to a significant advance in genome engineering.

Expectations for biallelic frequencies are based on the two events occurring independently; however, it is possible that an event at one allele provides a substrate for gene conversion to generate the second allele. To test this possibility, we transfected cells with two different donor DNA; one expressing GFP used for detection of HDR-mediated insertion (detection donor), and one expressing G418 resistance used for positive selection of cells (selection donor). Each donor DNA had the 208/202bp arms of homology for the Lef1 site characterized in Fig 4.3. By performing G418 selection for insertion of one of the donor DNA, insertion of the second GFP donor was increased from 0.49% to 15.44%, which represents a 32-fold enrichment (Fig 4.5A). Since the high frequency of insertions at both alleles occurred even when two different sequences were inserted, we conclude that the high frequency was not caused by gene conversion.

We posited that the high biallelic frequencies were caused by independent but coincidental insertion (COIN) of each donor DNA after Cas9-mediated DSB. To test whether the COIN effect occurred through independent events, we determined if it would work for combinations of genes on different chromosomes and using different sgRNA for the Cas9-mediated DSB. Combinations of insertions targeting Lef1 and Tcf1 genes were detected by insertion of GFP. These detection insertions were performed either with or without selection for G418-resitance by insertion at Lef1 and Rosa26 genes (Fig 4.5B-E). Each combination of selection and detection donors generated a substantial COIN effect, ranging from 10-fold for Rosa26-Neo/Tcf1-GFP to 22-fold for Lef1-Neo/Tcf1-GFP combination (Fig 4.5C,E). These

results demonstrate that COIN is effective even when distinct, distant genomic sites are combined together in selection/detection combinations.

Effect of Homology Arm Length on COIN Outcomes

We interpret COIN working for both distinct, unlinked genomic sites and for different alleles at a single genomic site to suggest that COIN functions through the co-occurrence of independent events in cells. This co-occurrence allows for the positive selection of one event (e.g Rosa26::Neo) to enrich for cells more likely to have completed the second, independent event (e.g. Tcf1::GFP). In order for co-occurrence to provide enrichment for these similar but independent events, the cells must be heterogeneous in their ability to undergo HDR. This hypothesis predicts that selection donor DNAs that are highly effective at HDR insertion will produce the weakest COIN effect, because they provide the least stringent selection for HDR-proficiency. Conversely, relatively poorly integrated selection DNA donors should produce the greatest COIN effect. We tested this possibility by comparing the COIN effect on Lef1::GFP insertion with positive selection provided by Rosa26-Neo donor DNAs with various homology arm lengths (Fig 4.7A). The weakest COIN effect came from the selection donor with the longest homology arms (1087/4039bp), as it provided only 5-fold enrichment of Lef1::GFP insertions (Fig 4.5A). Selection donors with the shortest homology arms (219/235bp) provided the greatest COIN effect with a 16-fold enrichment of Lef1::GFP insertions (Fig 4.5A). Thus, the COIN enrichment effect is indirectly proportional to the effectiveness of the selection donor DNA. These results suggest that COIN effects exist because transfected cells exhibit a range of HDR-proficiency, and higher stringency of selection effectively enriches for the most HDR-proficient cells.

To determine if the COIN effect can enrich for both relatively frequent and infrequent HDR events, the homology arm length of the detection donor DNA (Lef1-GFP) was varied. The homology arm length of the selection donor DNA (Rosa26-Neo) was kept constant at 516/495bp, a length that provided an intermediate COIN enrichment effect (Fig 4.7A). As expected, the addition of longer arms of homology increased baseline insertion frequency without COIN, from 0.78% for 208/202bp arms to 1.91% for 1443/1471 arms (Fig 4.7B). COIN increased insertion frequencies for all Lef1-GFP donors (from 3.3% to 21.47%), with the COIN effects rising with increasing baseline detection insertion rates. These results are consistent with COIN effectively enriching for HDR-proficient cells over a broad range of baseline efficiencies, indicating a potential to provide benefits to diverse genome engineering applications.

Figure 4.5 High frequency of coincidental insertion of distinct donor DNA targeting the same gene and independent genes

(A) Experimental approach showing transfection of donor DNA for selection (Lef1-Neo) and detection (Lef1-GFP), splitting cells into G418-containing media or non-selective media, and detection of Lef1::GFP frequency by microscopy and flow cytometry to determine percentage of GFP-positive cells. Each donor DNA had 208/202bp homology arms. Fold increase is calculated as percentage of GFP+ cells after selection divided by GFP+ percentage for all transfected cells. Insertion % is measured by flow cytometry (shown in Figure 4.6A).

(B-E) Similar to (A); however sgRNA and donor DNA target insertion at distinct genes (Lef1, Tcf1, Rosa26) (C-E). The combination of selection and detection donor DNA are noted at the top of each column. Insertion % is measured by flow cytometry (shown in Figure 4.6C-D).

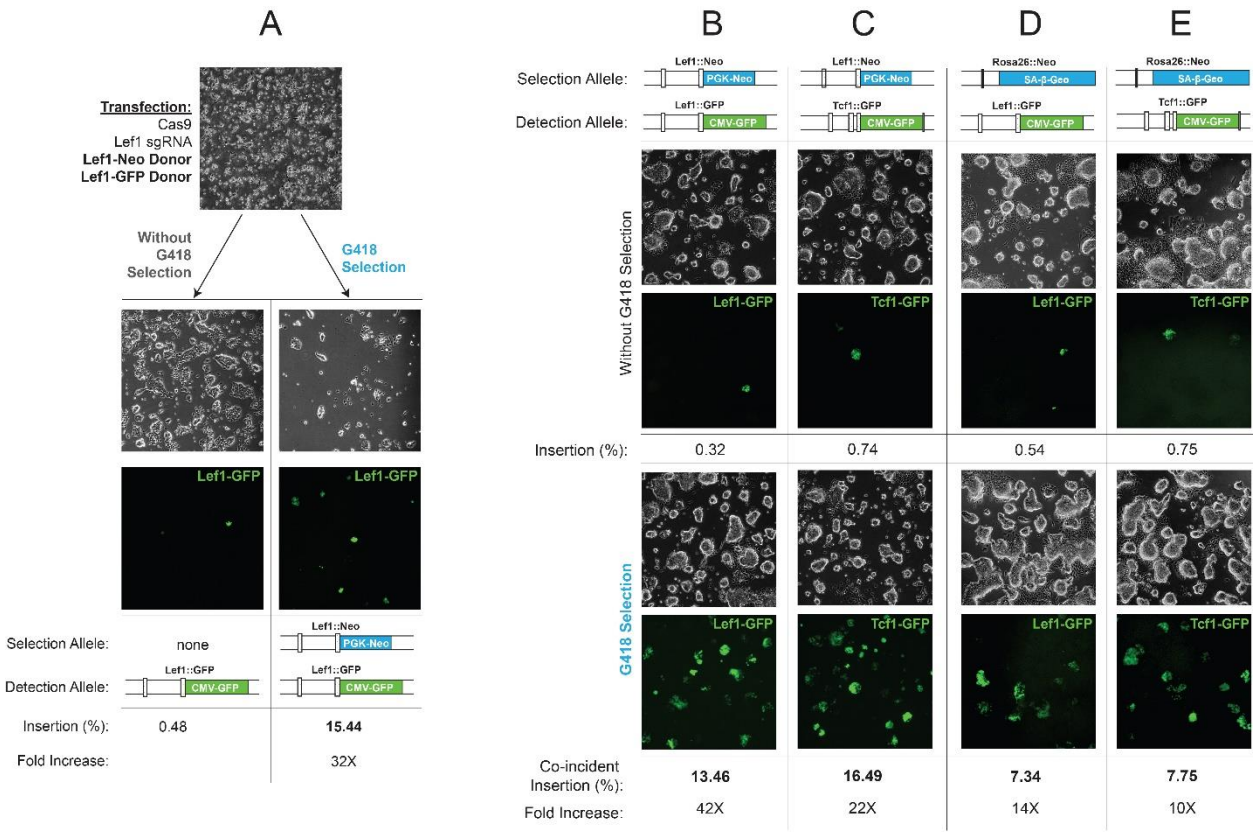

A
Transfection:
Cas9
Lef1 sgRNA
Lef1-Neo Donor
Lef1-GFP Donor
Without G418 Selection
G418 Selection
Lef1-GFP
Lef1-GFP
Selection Allele:
none
Lef1::Neo
PGK-Neo
Detection Allele:
Lef1::GFP
CMV-GFP
Lef1::GFP
CMV-GFP
Insertion (%):
0.48
15.44
Fold Increase:
32X
B
C
D
E
Selection Allele:
Lef1::Neo
PGK-Neo
Lef1::Neo
PGK-Neo
Rosa26::Neo
SA-β-Geo
Rosa26::Neo
SA-β-Geo
Detection Allele:
Lef1::GFP
CMV-GFP
Tcf1::GFP
CMV-GFP
Lef1::GFP
CMV-GFP
Tcf1::GFP
CMV-GFP
Without G418 Selection
Lef1-GFP
Tcf1-GFP
Lef1-GFP
Tcf1-GFP
Insertion (%):
0.32
0.74
0.54
0.75
G418 Selection
Lef1-GFP
Tcf1-GFP
Lef1-GFP
Tcf1-GFP
Co-incident Insertion (%):
13.46
16.49
7.34
7.75
Fold Increase:
42X
22X
14X
10X

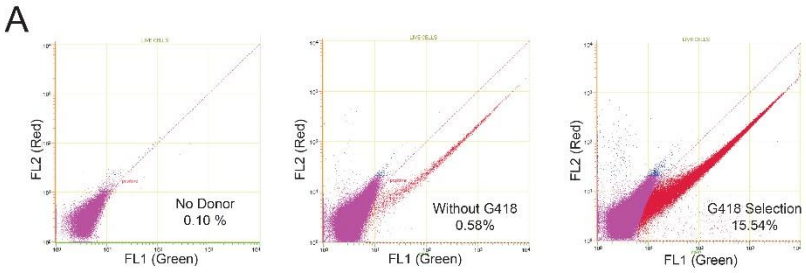
Figure 4.6 Flow cytometry data for COIN

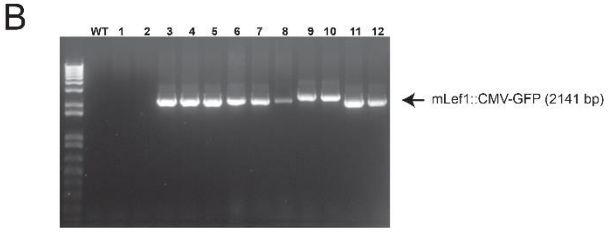
(A) Flow cytometry measurements for figure 4.3A.

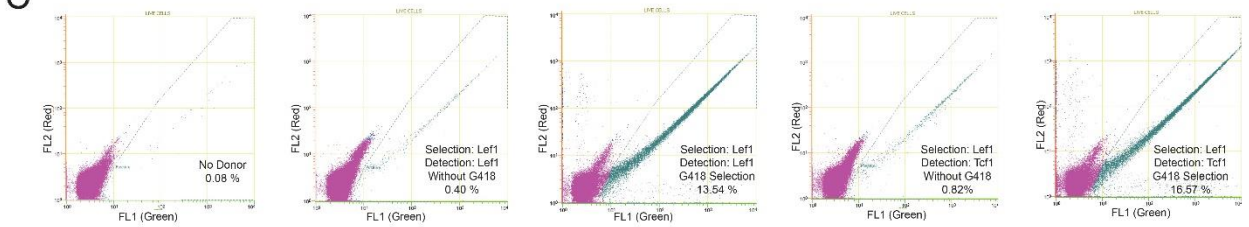
(B) PCR genotyping analysis of 12 clones exhibiting both Neomycin resistance and GFP expression. Length of left and right donor homology arms were 208bp and 202bp, respectively. Primers are specific for CMV-GFP insertion into the Lef1 locus. Results show 10/12 clones with CMV-GFP insertion at the correct location. Compare to 6/12 when only one is allele selected as in figure 4.3E.

(C) Flow cytometry measurements for figure 4.5B-C.

(D) Flow cytometry measurements for figure 4.5D-E.


A
FL2 (Red)
FL1 (Green)
No Donor
0.10 %
Without G418
0.58%
G418 Selection
15.54%


B
WT 1 2 3 4 5 6 7 8 9 10 11 12
mLef1::CMV-GFP (2141 bp)


C
FL2 (Red)
FL1 (Green)
No Donor
0.08 %
Selection: Lef1
Detection: Lef1
Without G418
0.40 %
Selection: Lef1
Detection: Lef1
G418 Selection
13.54 %
Selection: Lef1
Detection: Tcf1
Without G418
0.82%
Selection: Lef1
Detection: Tcf1
G418 Selection
16.57 %

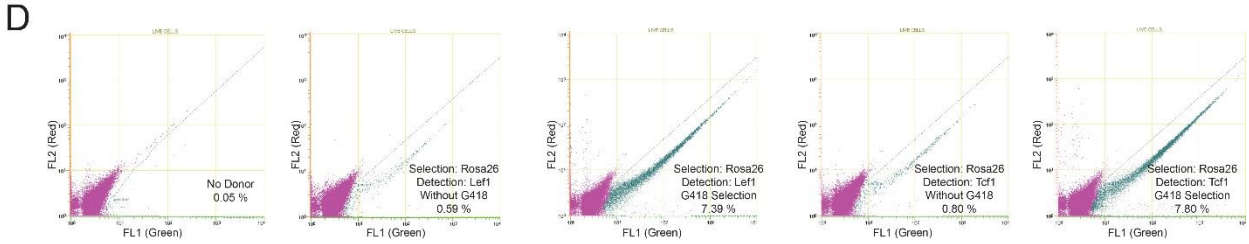
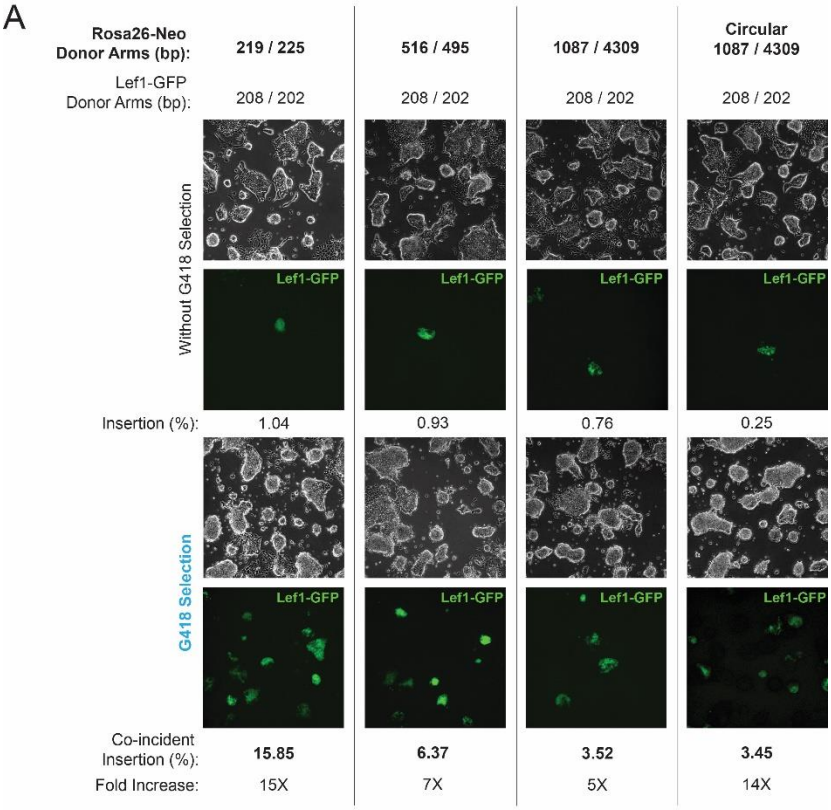

D
FL2 (Red)
FL1 (Green)
No Donor
0.05 %
Selection: Rosa26
Detection: Lef1
Without G418
0.59 %
Selection: Rosa26
Detection: Lef1
G418 Selection
7.39 %
Selection: Rosa26
Detection: Tcf1
Without G418
0.80 %
Selection: Rosa26
Detection: Tcf1
G418 Selection
7.80 %

Figure 4.7 Homology arm length alters the frequency of co-incident insertion.

(A) Flow cytometry analysis and representative DIC and fluorescent microscopy images of using Lef1-GFP as the donor DNA for detection and Rosa26-Neo as the donor DNA for selection. Percentage of GFP+ cells was determined for all transfected cells (top) and for cells following G418 selection (bottom). As indicated at the top of each column, homology arm length was varied for the Rosa26-Neo donor DNA and kept constant at 208/202bp for Lef1-GFP. Insertion % is measured by flow cytometry (shown in Figure 4.8A).

(B) Similar to (A), except homology arm length was varied for Lef1-GFP and kept constant for Rosa26-Neo at 516/495bp. Insertion % is measured by flow cytometry (shown in Figure 4.8B).

A


Rosa26-Neo Donor Arms (bp):
219 / 225
516 / 495
1087 / 4309
Circular 1087 / 4309
Lef1-GFP Donor Arms (bp):
208 / 202
208 / 202
208 / 202
208 / 202
Without G418 Selection
Lef1-GFP
Insertion (%):
1.04
0.93
0.76
0.25
G418 Selection
Co-incident Insertion (%):
15.85
6.37
3.52
3.45
Fold Increase:
15X
7X
5X
14X

B

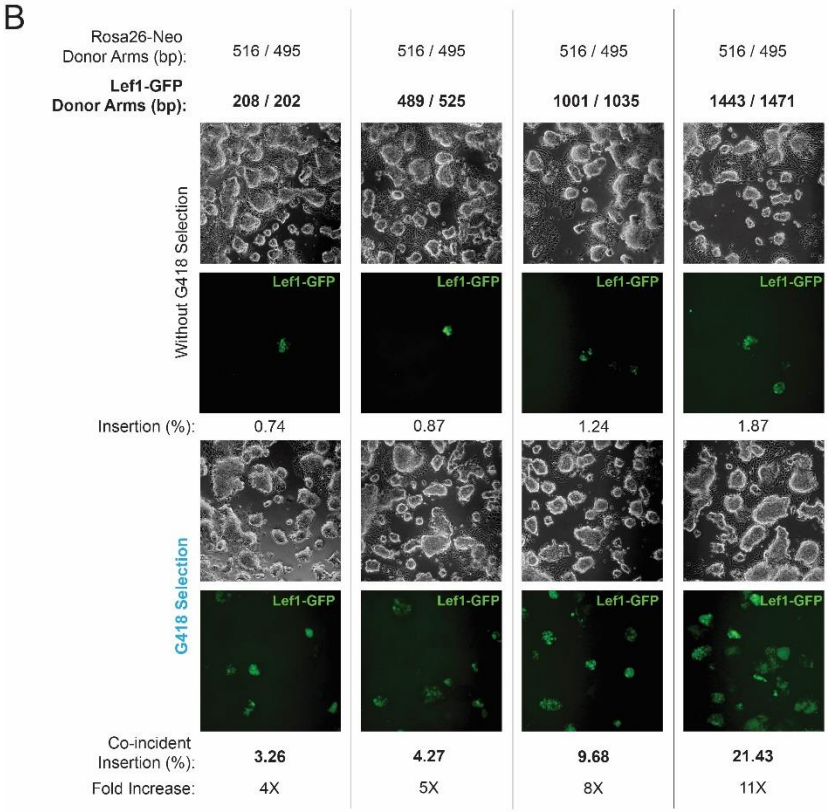
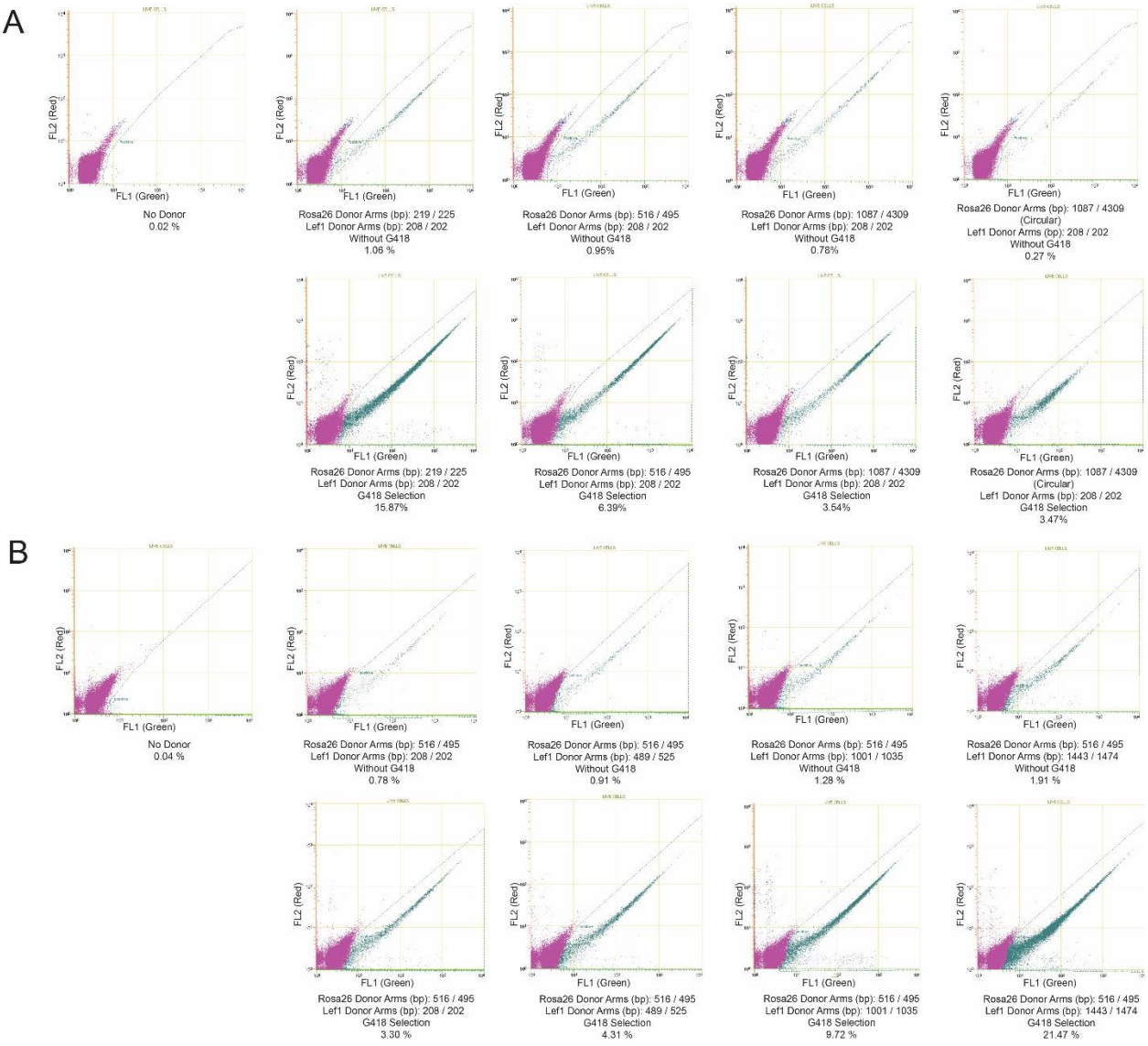

Rosa26-Neo Donor Arms (bp):
516 / 495
516 / 495
516 / 495
516 / 495
Lef1-GFP Donor Arms (bp):
208 / 202
489 / 525
1001 / 1035
1443 / 1471
Without G418 Selection
Lef1-GFP
Insertion (%):
0.74
0.87
1.24
1.87
G418 Selection
Co-incident Insertion (%):
3.26
4.27
9.68
21.43
Fold Increase:
4X
5X
8X
11X

Figure 4.8 Flow cytometry data for COIN

(A) Flow cytometry measurements for figure 4.7A.

(B) Flow cytometry measurements for figure 4.7B.


A
FL2 (Red)
FL1 (Green)
No Donor
0.02 %
Rosa26 Donor Arms (bp): 219 / 225
Lef1 Donor Arms (bp): 208 / 202
Without G418
1.06 %
Rosa26 Donor Arms (bp): 516 / 495
Lef1 Donor Arms (bp): 208 / 202
Without G418
0.95%
Rosa26 Donor Arms (bp): 1087 / 4309
Lef1 Donor Arms (bp): 208 / 202
Without G418
0.78%
Rosa26 Donor Arms (bp): 1087 / 4309
(Circular)
Lef1 Donor Arms (bp): 208 / 202
Without G418
0.27 %
Rosa26 Donor Arms (bp): 219 / 225
Lef1 Donor Arms (bp): 208 / 202
G418 Selection
15.87%
Rosa26 Donor Arms (bp): 516 / 495
Lef1 Donor Arms (bp): 208 / 202
G418 Selection
6.39%
Rosa26 Donor Arms (bp): 1087 / 4309
Lef1 Donor Arms (bp): 208 / 202
G418 Selection
3.54%
Rosa26 Donor Arms (bp): 1087 / 4309
(Circular)
Lef1 Donor Arms (bp): 208 / 202
G418 Selection
3.47%
B
FL2 (Red)
FL1 (Green)
No Donor
0.04 %
Rosa26 Donor Arms (bp): 516 / 495
Lef1 Donor Arms (bp): 208 / 202
Without G418
0.78 %
Rosa26 Donor Arms (bp): 516 / 495
Lef1 Donor Arms (bp): 489 / 525
Without G418
0.91 %
Rosa26 Donor Arms (bp): 516 / 495
Lef1 Donor Arms (bp): 1001 / 1035
Without G418
1.28 %
Rosa26 Donor Arms (bp): 516 / 495
Lef1 Donor Arms (bp): 1443 / 1474
Without G418
1.91 %
Rosa26 Donor Arms (bp): 516 / 495
Lef1 Donor Arms (bp): 208 / 202
G418 Selection
3.30 %
Rosa26 Donor Arms (bp): 516 / 495
Lef1 Donor Arms (bp): 489 / 525
G418 Selection
4.31 %
Rosa26 Donor Arms (bp): 516 / 495
Lef1 Donor Arms (bp): 1001 / 1035
G418 Selection
9.72 %
Rosa26 Donor Arms (bp): 516 / 495
Lef1 Donor Arms (bp): 1443 / 1474
G418 Selection
21.47 %

Discussion:

The adaptation of the CRISPR/Cas9 system for use in mammalian cells provided a simple and powerful method to edit stem cell genomes [111,167,168]. Currently, the high frequency of mutagenic NHEJ-driven repair relative to sparse HDR processing of Cas9-mediated DSB constitutes a barrier to engineering large and precise changes [111,228]. To optimize frequency of precise insertions into the ES cell genome, we first focused on the donor DNA used as a substrate for HDR. Characterization of homology arm length demonstrated the ability of 200bp arms to stimulate consistent on-target insertions. Interestingly, longer homology arms increased frequency of on-target insertions but did not affect that of off-target insertions, consistent with the notion that on- and off-target insertions are mediated by distinct mechanisms. Recent rational approaches to the off-target insertion problem aim to either enhance on-target rates by stimulating HDR activity, or to reduce off-target integration by inhibiting NHEJ or designing donor DNA to be a poor substrate for end-ligation reactions [232,233]. To date, attempts at manipulating repair activity have yielded only minor improvements relative to the dramatic benefits of localizing repair activity to a genomic site by a Cas9-mediated DSB.

The COIN effect substantially boosts efficiency of current genome engineering practices. Recent advances from a small molecule screen identified two compounds that increased insertion of a Nanog-GFP donor DNA by 2- to 3-fold [229]. In addition, electroporation of active Cas9 RNPs into G2/M synchronized cells similarly increased frequency of a 12bp insertion from a 250bp dsDNA donor [169]. By contrast, COIN increased the frequency of HDR-mediated insertion by up to 42-fold when selection and detection dsDNA donors targeted a single site, and up to 22-fold when the two donors were targeted to distinct sites. Relative to HDR without a second selection donor to enrich for HDR events, COIN reduces the number of clones one must screen by 90-95%. This increased efficiency facilitates ease in design of effective donor DNAs by reducing the length of homology arms needed and removing the need for a selection marker

cassette within the donor DNA. Finally, by using dsDNA as donors, COIN allows large changes greater than 1kb to be inserted at a frequency similar to that of smaller changes generated using ssODN with current methods.

Potential limitations of COIN can be largely mitigated by some simple enhancements. The primary concern with using COIN is that the second sgRNA necessary for the effect could contribute to the number of Cas9-mediated DSB at off-target sites. For new stem cell lines where off-target effects are a concern, it will be necessary to use an sgRNA that has been carefully assessed for off-target cutting using an empirical method, such as GUIDE-Seq [234]. Donor DNA used for selection should also be optimized for a high ratio of on:off-target insertion events. Here we use a splice acceptor Neomycin resistance cassette (SA-β-Geo) to target the Rosa26 locus. However, a fusion protein requiring a precise in-frame insertion or the use of negative selection outside of homology arms could provide substantially higher specificity [147].

Finally, while we use COIN specifically for the promotion of HDR in mouse ES cells, we expect the principles to be broadly applicable. High levels of co-occurrence in terms of bialleleic events have been reported for genome editing procedures in a variety of mammalian systems [228,230,231,235]. A recent study extends the co-occurrence of HDR events to *S. cerivisiae* [236]. We suggest that COIN is an example of an important and broadly applicable principle to consider when performing genome engineering experiments. The principle is based on the co-occurrence of independent events wherein epigenetic heterogeneity exists among genetically homogenous cells. The rare cells that are proficient at completing one genomic manipulation have an increased probability of completing a second, independent genomic manipulation, provided the two are sufficiently similar. Using positive selection to enrich for one manipulation also enriches for the second. In this way, manipulations incompatible with direct selection or screening can still be enriched.

Experimental Procedures:

ES Cell Culture.

$2 \times 10^5 – 2 \times 10^6$ C57BL/6 mouse embryonic stem cells (ESC) were plated on 0.1% gelatin (Millipore #ES-006-B) on 6 well plates (Falcon #353046). Cells were grown in Knockout DMEM (GIBCO #10829-018) supplemented with the following: 15% Fetal Bovine Serum (GIBCO #10437-028), 2 mM L-Glutamine (GIBCO #25030-081), 1000 U/mL Pen Strep (GIBCO #15140), 1 mM HEPES (Thermo Scientific #SH30237.01), 1X MEM NEAA (GIBCO #11140), 55 uM 2-Mercaptoethanol (GIBCO #21985-023), 100 U/mL LIF (Millipore #ESG1106), and 3uM CHIR99021 (Sigma #SML1046). Cells were split 1:10 with 0.25% Trypsin-EDTA (GIBCO #25200-072) every 2-3 days.

Quantitative PCR (qPCR) Assay.

A set of concentration standards for qPCR genotyping the Lef1::Neo insertion was generated by mixing genomic DNA (gDNA) obtained from a heterozygous Lef1::Neo line with WT gDNA. Original clones are shown in Fig S1A. gDNA was mixed to generate the following percentages of Lef1::Neo: 50%, 25%, 10%, 2%, 0.5%, and 0.1%. Sample DNA was isolated from ~2×10^6 cells by lysis with 0.5% SDS followed by ethanol precipitation. 50ng of gDNA from each sample was combined with Perfecta SYBR Green Supermix (Quanta #95053) and primers specific for either the Lef1::Neo insert (Fig 1A) or a WT Lef1 (exon12) control ~30kb downstream. Primers used are: Lef1::Neo (F Primer): CTGCCCCTTTCCCTAACTG, Lef1::Neo (R Primer): CCGGATCCACTTTCATCATC, Lef1(exon12) Forward: CTGCCCTGTGAAGTGTCTGA, Lef1(exon12) Reverse: AATGAACTGCAAACGGGTTC. qPCR was performed on a C1000 thermal cycler and CFX96 Real Time System (Bio-Rad) with the following parameters: 95° for 2m, then 40 cycles of 95° for 30s and 60° for 30s. Quantities were normalized to the Lef1 exon 12 control and cycle number was compared against the ladder for Lef1::Neo to quantify insert percentage.

Acknowledgements:

We thank Matt MacDougall, Jenny Zhang, and Maureen Regan for helpful discussions associated with the manuscript. The work was funded by grants from the National Institutes of Health (R01-HD081534, B.J.M and T32-HL007829, B.R.S) and B.R.S was partially funded by the Chicago Biomedical Consortium with support from the Searle Funds at The Chicago Community Trust (SCH-032).

Chapter 5: Concluding Remarks

Part I: TCF/LEFs, WNT signaling, and the differentiation of mouse embryonic stem cells

Mechanism of TCF3 requirement for EpiLC differentiation

TCF3 is required both *in vivo* to prepare cells for gastrulation[89,237] and *in vitro* for EpiLC differentiation. As described in Chapter 2, cells without TCF3 are unable to turn off ESC genes, were deficient in activating EpiLC genes, and failed functional tests of differentiation. Conversely, cells with only TCF3 and none of the other TCF/LEFs differentiated appropriately to EpiLC. This identifies TCF3 as unique among the TCF/LEFs in its ability to drive differentiation but does not describe the mechanism.

One clue may be in the disappearance of TCF3 bound DNAse hypersensitivity sites (DHS). DNAse I preferentially digests regions of genomic DNA that lack nucleosomes and thus DHS represent regions of open chromatin. Although there is a group unique transcription factors which can bind to condensed chromatin, so-called “pioneer” factors[238], the majority of transcription factors can only bind efficiently to nucleosome-free DNA within these DHS. DHS are thus central points of regulation for both gene expression and nuclear architecture. The pattern of DHS throughout the genome is the epigenetic modification that best correlates with the differentiation status of a cell and may be a major determinant of a cell’s properties through activating and inactivating regions of genomic DNA[47-49,239,240].

A particular cell fate is frequently maintained by a core set of transcription factors specific for that fate. These core factors regulate the expression of one another through cooperative binding at large clusters of open chromatin called “super” or “stretch” enhancers (SE)[47,49,240]. These SE provide an epigenetic signature of a particular cell fate and a convergence point for transcriptional regulation by many different signaling pathways[47-49,240]. They are preferentially lost during differentiation as circuits responsible for one cell fate are shut down and new circuits are activated[47-49,178,240].

Like other cell types, ESC contain a core set of transcription factors including OCT4, SOX2, NANOG, ESRRB, TFCP2L1, KLF2, and KLF4[21,36-42]. These form a feed forward circuit and contain SE with a high frequency of cooperative binding[44,46,49]. TCF3 appears to act in opposition to these factors. It binds to the majority of the same enhancers and SE[43,44,46], but rather than promoting ESC self-renewal, TCF3 inhibits self-renewal and stimulates EpiLC differentiation. Thus, TCF3 is perfectly positioned in regions of open chromatin that are important for maintenance of the ESC state and closed during differentiation. TCF3 is also functionally required for this transition. This suggests TCF3 may be required for shutting down these important regulatory regions. This hypothesis could be tested by analyzing whether these sites are lost in TCF3 -/- cell lines.

If ESC specific DHS sites are inappropriately maintained in the absence of TCF3, the next question would be whether this is a direct effect of TCF3 at the local level, or a more general effect of TCF3 based repression. TCF3 could possess a unique biochemical activity or more likely recruit other proteins to stimulate chromatin condensation. An example would be recruitment of chromatin remodeling complexes such as the SWI/SNF or nucleosome remodeling and deacetlyase (NuRD) complexes[241]. Intriguingly, numerous members of these complexes were identified in a mass spectrometry analysis of proteins which co-purified with TCF3 (Laura Pereira, unpublished data). To test the possibility of a local requirement for TCF3, TCF3 binding sites could be removed from specific enhancers which close during differentiation but are not required for either the ESC or EpiLC state. This could be accomplished using CRISPR/Cas9 based tools to delete or replace small stretches of genomic DNA within enhancers. Enhancers for Zinc finger protein 42 (Zfp42, also called Rex1) would be a good choice because its expression is highly correlated with the naïve ESC state[94], it contains enhancers which are lost during differentiation, these enhancers are bound by TCF3 and other

core transcription factors, and its expression is not required for maintenance of the ESC state or differentiation to EpiLC.

Alternatively, TCF3 may function indirectly through repression of factors which prevent enhancer decommissioning. For example, it has been proposed that the cooperative binding of transcription factors may hold back a constitutive drive towards chromatin condensation. During EpiLC differentiation, loss of SEs might therefore be mediated just by a reduction in the binding of core transcription factors. TCF3 represses *Nanog*, *Esrrb*, and *Tfcp2l1* directly and may directly or indirectly reduce expression of other core factors[5,21,37]. By reducing their expression and thus lowering their presence on chromatin, TCF3 may indirectly cause the closure of these sites. To test this alternative, one could knock down *Nanog*, *Esrrb*, and *Tcfcp2l1*, individually or in combination, in a *Tcf3* -/- background. If the role of TCF3 is simply to reduce expression of these genes, then reducing their levels should allow closure of sites in the absence of TCF3.

Mechanism of TCF3 Inactivation

Dysregulation of the WNT pathway is present in a diverse array of cancer types. Mutations which stabilize β-Catenin and epigenetic effects that stimulate WNT signaling cause cancer of the colon, liver, brain, and breast[242-246]. Several studies have shown a correlation between β-Catenin activity and poor differentiation status of the tumor[245]. “Triple-negative” or “basal” breast cancers are an example of poorly differentiated tumors with frequent activation of WNT, and are notable for their aggressive nature, lack of effective treatments, and poor clinical outcomes[201,208,209]. They have additionally been associated with an ESC-like gene expression signature, and like ESCs, they express high levels of *Tcf3* mRNA[201].

Blocking the WNT pathway in triple negative breast cancer cell lines has shown promising results: reduced cell proliferation, invasiveness, and colony formation *in vitro*, and

fully arrested tumor growth in xenografts[247]. Considerable effort has been expended to develop drugs which can similarly inhibit WNT activation; however, they have met little success. One barrier is the lack of druggable target proteins necessary for WNT signaling specifically in cancer cells. Our results *in vitro* indicate that Wnt stimulates TCF3 inactivation in breast cancer cell lines similarly to ESC. CHIR reduces TCF3 protein levels, stimulates nuclear export, and TCF3 overexpression blocks target gene activation. While high *Tcf3* mRNA is correlated with basal type breast cancers, analysis of tumor samples indicate low and cytoplasmic TCF3 protein. This suggests that *Tcf3* mRNA is being expressed but subsequently inactivated at the protein level. If inactivation of TCF3 is required to prevent differentiation in breast cancers, as it is in ESC, then targeting the enzymes responsible could provide a new therapeutic approach to attack currently intractable cancers.

We have demonstrated that in response to β-Catenin binding, TCF3 is removed from chromatin, exported from the nucleus, and subsequently degraded. However, the biochemical mechanism that mediates this response is still unclear. One clue to the underlying mechanism of TCF3 inactivation may be the post-translational modifications (PTM) of TCF3. TCF3 is phosphorylated both in the presence and absence of WNT stimulation at unknown sites. WNTs have been shown to induce phosphorylation of the TCF/LEF transcription factors in a number of contexts (see Table 5.1) with the most well established roles through Homeodomain-interacting Protein Kinase 2 (HIPK2) and Nemo-like Kinase (NLK).

HIPK2 was identified in a yeast 2-hybrid screen for TCF3 interacting proteins[88]. In *Xenopus laevis*, it was shown to stimulate phosphorylation of TCF3 in a β-Catenin dependent matter, leading to decreased chromatin occupancy at a WNT responsive promoter[88,212]. *Hipk2* and *Tcf3* morpholinos showed opposing phenotypes in the development of the antero-posterior axis with phenotype reversal by co-injection[88]. HIPK2 phosphorylation of TCF3 was also observed in 293 cells following treatment with WNT3a, suggesting conservation in mammalian

Table 5.1 TCF/LEF phosphorylation sites and associated kinases

Kinase	Effect	Substrate(s)	Wnt pathway regulation	Reference
LIT-1 (NLK)	phosphorylation	POP1	positive	Rocheleau 1999
NLK	phosphorylation	LEF1 TCF3 TCF4	negative	Ishitani 1999, 2003
	reduced DNA binding	TCF4	negative	Ishitani 1999, 2003
	synergizes with Tcf3 MO	TCF3	positive	Thorpe 2004
	ubiquitylation	LEF1 TCF4	negative	Yamada 2006, Li 2010
	degradation	LEF1 TCF4	negative	Yamada 2006
HIPK2	phosphorylation	LEF1 TCF3 TCF4 (not TCF1)	positive	Hikasa 2010, 2011
	reduced chromatin occupancy	LEF1 TCF3 TCF4	positive	Hikasa 2010, 2011
	increased chromatin occupancy	TCF1	positive	Hikasa 2010, 2011
GSK3	phosphorylation	TCF3		Lee 2001
CK1	phosphorylation	TCF3		Lee 2001
	phosphorylation	LEF1	negative	Hammerlein 2005, Wang 2006
	disrupts beta-catenin interaction	LEF1	negative	Hammerlein 2005, Wang 2006
CK2	phosphorylation	LEF1 TCF4	positive	Hammerlein 2005, Gao 2006, Miravet 2002, Wang 2006
	promotes beta-catenin interaction	LEF1	positive	Wang 2006

cells[88,212]. The TCF3 phosphorylation sites identified in *Xenopus* are conserved in mouse and human, making HIPK2 a prime candidate for the regulation of TCF3. Unfortunately, mutation of these sites in mouse ESC failed to inhibit either WNTs inactivation of TCF3, or TCF3 phosphorylation. This indicates that either HIPK2 does not mediate TCF3 inactivation in mammals, or it is through a unique set of amino acids. A knockdown or knockout of HIPK2 may be able to address this question, however it is complicated by potential redundancies with HIPK1 and HIPK3. A triple knockout in mESC could be generated with CRISPR/Cas9 based tools to definitively address this question.

The activity of NLK on TCF/LEFs is better studied, yet its function is less clear. It was originally suggested from work in *C. elegans*, which identified the stimulation of its homolog, LIT-1, in response to WNT. This led to the phosphorylation and inactivation of the TCF homolog, POP-1, in a WRM-1 (β-Catenin homolog) dependent manner[248]. Like TCF3, POP-1 acts as a constitutive repressor, and this mechanism suggested NLK was a positive regulator of WNT activation[248]. This was supported by work in *Xenopus* embryos, which identified NLK as a positive regulator of WNT activity which counteracts the role of TCF3 in neurectoderm patterning[215]. Subsequent work confirmed that NLK phosphorylated mammalian LEF1, TCF3, and TCF4, and it was further shown that phosphorylation of LEF1 and TCF4 led to reduced DNA binding, ubiquitination, and degradation (effects on TCF3 were not reported). However, this work also suggested that NLK acts as a negative regulator of the WNT pathway because target gene expression was reduced by NLK[213,214,249,250]. Thus, it is known that mammalian TCF3 is phosphorylated by NLK but the position and effect of this phosphorylation is unclear. A knockdown or knockout of NLK in mESC could address its requirement.

Other reports indicate that CK1, CK2, and GSK3 can also phosphorylate TCF/LEF transcription factors to effect WNT activation[251-253]. These kinases may also be knocked out to identify their requirement but may be more difficult to analyze because of their additional roles in

the β-Catenin destruction complex. Inhibition of GSK3 with CHIR, for example, is the method we typically use to stimulate TCF3 inactivation. In this case, it may be simpler to generate point mutations in TCF3 at potential target sites. Many of these sites have been identified for other TCF/LEFs and are variably conserved in TCF3.

In addition to phosphorylation, TCF3 undergoes a second PTM in response to WNT stimulation that is not removed by incubation with phosphatase. The identity of this PTM is unknown, but its small size (<5 kDa) makes it unlikely to be mono or poly-ubiquitin, as might be expected from the subsequent degradation of TCF3. Although SUMOylation has been implicated for other TCF/LEFs[254,255], the large size of the SUMO proteins (~12.5 kDa) also rules these out. Acetylation of lysine residues is one possibility that remains and intriguingly, acetylation of POP-1 regulates its nuclear localization in *C. elegans*[256]. Acetylation of TCF4 regulates DNA binding in human cells and requires β-Catenin interaction. The reaction is mediated by CREB-binding protein (CBP) and the closely related p300[257]. A knockdown or knockout in ESC could be informative.

Due to the long list of potential enzymes involved, a more top-down approach beginning with the identification of the PTM may be more appropriate. One option would be to analyze TCF3 by mass spectrometry before and after WNT stimulation. Both phosphorylation and acetylation sites can be identified by mass spectrometry although sensitivity can be low, particularly if it is a low abundance modification[258]. However, this unsupervised analysis has the potential to identify unpredicted static and WNT induced PTM.

Alternatively, a number of phosphorylation sites and one ubiquitination site have been identified in high-throughput screens using modification specific antibodies. The position and number of hits for each modification have been aggregated in an online database at www.phosphosite.org[259]. Three hits that show numerous identifications are phosphorylation at

Y392, phosphorylation at S428, and ubiquitination at K390. These sites are all within or just downstream of the HMG binding domain.

Finally, a series of truncations could be generated to narrow down regions of PTM and modification specific antibodies can be used to identify specific PTMs. Once identified, a series of point mutations could be made to determine their importance, and potential enzymes may be inferred from known recognition motifs.

Part II: Tools for genome editing

Optimization of COIN

Cas9 induced DSBs greatly enhance the frequency of HDR and yet, in many circumstances, it is still an infrequent event. Certain regions of the genome and many cell types are refractory to HDR[260-263]. For example, human and other primate cell lines are up to 100-fold less likely to integrate exogenous DNA than rodent cell lines[260-263]. In addition, higher efficiencies typically require the use of large homology arms totaling 2-14kb. De novo DNA synthesis rates have plummeted in recent years but are limited in size[264]. Inexpensive ssDNA synthesis, which can be used to quickly generate donor arms by PCR, is currently limited to 200bp. Inexpensive dsDNA can be generated up to 2kb, but there is a long list of exclusion criteria that make it nearly impossible to generate the specific sequences required for homology arms. Larger sequences also exhibit a higher error rate that may introduce undesired mutations into the genome. This means that expensive and time consuming multi-step cloning procedures must still be utilized to generate donor constructs.

To improve the efficiency of genome editing, we have developed a method of co-insertion (COIN) which increases the recovery of desired mutations by as much as 30-fold. In practice, this means that instead of screening hundreds of clones, a user may screen less than

ten. For genome editing experiments, this allows better coverage of the genome, the generation of larger changes, and the use of smaller homology arms. Using COIN, we have inserted changes greater than 2kb using less than 200bp arms at frequencies of higher than 16% of cells. This is well within the range of inexpensive de novo synthesis options, allowing quick and easy generation of donor constructs.

COIN was developed from the observation that a cell which underwent one HDR event was more likely to undergo a second. This phenomenon is true for two events at a single locus as well as two separate locations in the genome. This allowed us to develop a set of reagents which target an HDR event to a defined “safe-harbor” in the genome. For mouse cells, we use the *Rosa26* locus and for human cells we chose the *Aavs1* locus. Inserting a selectable marker at this location, such as Neomycin resistance or GFP expression, allows for the isolation of cells which undergo one HDR event. Selecting for this first event greatly improves the likelihood of obtaining a second desired mutation.

Targeting a single site for each organism allows for a high level of optimization in reagents. Once optimized, these reagents can be universally applied for all genome editing experiments in that organism. For example, a single plasmid could be generated that contains Cas9, a gRNA targeting *Rosa26*, and a selection donor targeting the *Rosa26* locus. A small amount of this plasmid could then be included in a standard transfection or electroporation to increase the frequency of any second mutation in mouse cells. Optimizing the system for each species comes down to identifying a good safe-harbor site to introduce the selectable marker, choosing a gRNA with a low rate of off-target cutting, and choosing a donor design that facilitates easy selection and removal.

Safe-harbor sites are well established in most organisms. The first requirement is that these locations are not important for any activity of the cell. Making a mutation at the site should not alter the growth or characteristics of the cell in any way. Second, it is useful if the site

contains an endogenous promoter that is constitutively expressed at a reasonable level in all cell types. This allows the use of the endogenous promoter to drive expression of the selectable marker. An exogenous promoter can be introduced in the donor construct, but it will be expressed regardless of where it is inserted in the genome. This substantially increases the recovery of off-target integrations during selection. In contrast, if the endogenous promoter is used, the selectable marker is only expressed at the correct location, particularly if an in-frame fusion is required.

gRNAs exhibit a range of off-target digestion that can be as high as the on-target site[171]. This leads to the generation of unwanted DSB and indels throughout the genome. Some observations have been made that help reduce the frequency of these events. First, the off-target sites are more frequently found in degenerate sequences throughout the genome. A standard gRNA targets a 20bp sequence, and if other locations in the genome are different by only 1-2bp, they are also likely to be cut[168,182]. It appears that a divergence of 3bp or more greatly reduces the likelihood that a site will be cut. In addition, the sequence closest to the 3' end of the gRNA is more important because it initiates the binding. Thus, differences near the 3' end are more likely to reduce the frequency of digestion. A number of algorithms have been generated which incorporate these rules and allow the quick identification of good target sites[168,182].

In addition, a number of gRNA modifications have been developed that reduce the frequency of off-target cutting. One of the simplest methods is a reduction in the size of the gRNA. It appears that there is redundancy built in to the 20bp recognition sequence of gRNAs. Reducing the length to 18bp does not reduce the frequency of on-target digestion for all examined gRNAs[170]. This may be an optimization that is useful for bacteria, *in vivo*, to accommodate the rapid mutation of viral genomes. For gRNA design, the first two bases on the 5' end of the gRNA can be removed without reducing the on-target cutting efficiency. However,

the frequency of off-target digestion and indel formation is substantially reduced[170,171]. Off-target locations already have one or more mismatched bases and therefore more sensitive to additional perturbations.

Finally, a series of well-designed gRNAs can be tested empirically using a method called GUIDE-seq[171]. This relies on the observation that small pieces of exogenous DNA are frequently incorporated at a DSB, even in the absence of explicit homology arms. Thus, a dsDNA can be added along with Cas9/gRNA, and will be incorporated at an increased frequency at sites where a DSB is generated. If a specific primer sequence is included in the integrated dsDNA, populations of cells can be analyzed by deep sequencing to identify the location and frequency of breaks. These can then be compared between potential gRNAs to identify those with the fewest off-target DSB.

Optimization of COIN donors and Self-Destruct Sequences (SDS)

One potential downside of COIN is that a selection cassette is left behind in the genome. This occurs at a safe-harbor site and does not necessarily need to be removed. However, it may be desirable to do so for more rigorous applications, if the selectable marker may interfere with later experiments, or if a second round of COIN is to be used to generate additional mutations. To provide a convenient method to remove the selection cassette, we have generated a set of specialized donors referred to as self-destruct sequences (SDS) (See Appendix 6.10).

SDS contain selectable markers for use with COIN as well as all of the components required to stimulate their own removal. In current constructs, the entire selection cassette is flanked by inverted terminal repeats (ITR) recognized by the PiggyBac transposase (PBase). When PBase is expressed, any sequence in between the ITR undergoes a scarless excision. No residual sequence is left in the genome and it is restored to WT genomic DNA. Rather than

requiring a separate transfection of PBase expression vectors after the isolation of clones, we have included a PBase expression cassette within the SDS. To prevent the expressed PBase from being constitutively active, it is fused to a modified estrogen receptor (ERT2) that is preferentially stimulated by 4-Hydroxy Tamoxifen (4-OHT)[265]. In the absence of 4-OHT, PBase-ERT2 is sequestered away from the nucleus. After COIN selection, when the user wishes to excise the SDS, 4-OHT is added to the media to activate PBase and stimulate removal.

Finally, because PBase is fairly inefficient at excision, a marker is needed to identify those cells which have successfully removed the construct. If a visible marker like GFP is used for selection, then removal can be detected simply by loss of GFP. However, for constructs which use antibiotic resistance for the initial selection, we have included a negative selection construct that can be used to isolate those cells which have the sequence removed. We have chosen to use a variant of the Thymidine Kinase enzyme from Herpes Simplex virus (ΔTK)[266]. ΔTK can convert normally inert nucleoside analogues, including Fialuridine (FIAU), into toxic metabolites. Thus when FIAU is added to cell media, it will preferentially kill cells still expressing ΔTK (those with the selection cassette) and enrich for cells that no longer have ΔTK (those with the selection cassette removed).

For our first generation, we generated a pair of SDS donors that targeted the Rosa26 locus in mice and contained either GFP as a selectable marker or a Puro-ΔTK fusion protein that provides the ability for both positive (puromycin resistant) and negative (FIAU sensitive) selection. As with other COIN donors, these successfully enrich for secondary mutations. After selection, clones were isolated and then incubated with 4-OHT for 3 days. For clones using a GFP SDS, GFP was lost from approximately 1-2% of cells. Likewise for Puro-ΔTK, 1-3% of clones were resistant to FIAU selection versus around .05% in the absence of 4-OHT. Thus, SDS can be used with COIN and provide a convenient method for subsequent removal that does not require any additional transfections or clonal isolation.

Although the current generation of SDS are functional, a number of improvements can be made to further increase their usefulness. The first relates to the efficiency of excision. The natural PBase enzyme stimulates excision at a fairly low frequency, however, a number of amino acid subsitutions have been identified which greatly increase its activity. By combining these mutations, a hyperactive PBase has been developed which improves excision frequency by up to 17-fold[267]. Although the current SDS excision rates of 1-3% easily accommodate FACS-based isolation, it is tedious when screening for GFP- colonies by eye. In addition, for PuroTK variants an improved excision rate would reduce the probability of picking a spontaneous FIAU resistant clone, which occurs in about .05% of cells. Thus, one simple improvement would be to convert to a hyperactive PBase variant. If even higher excision rates are desired, the system could be converted to use Cre Recombinase and LoxP sites. For a sequence of this length, the excision frequency using Cre-Lox would be near 100%[268]. The downside is that the Cre-Lox system leaves a single LoxP site of 27bp behind in the genome, which may not be desirable for rigorous applications or when Cre is to be used in later experiments. However, the remaining LoxP site would not prevent the sequential use of COIN. SDS donors would already include LoxP sites in both homology arms and would therefore be compatible with the modified sequence.

In addition to stimulating excision, PBase is also stimulates re-integration at compatible sequences throughout the genome, albeit at a lower frequency. This could obviously inhibit the successful removal of SDS from the genome and could alter the expression or regulation of genes in unpredictable ways. Cells may be analyzed by PCR after isolation to confirm that no re-integration has occurred but this adds an additional step to the process. It is preferable to reduce the probability of re-integration directly. This may be accomplished in two ways. First, like the hyperactive PBase mutations, a series of amino acid changes have been identified which reduce the integration efficiency, and these can be combined to produce an integration

deficient variant of PBase[269]. Additionally, combining the integration deficient mutations with the hyperactive excision mutations does not ameliorate either effect. It would be beneficial to generate an ERT2-fused, excision-hyperactive, integration-deficient PBase variant for use with the SDS.

To reduce the probability of re-integration further, the negative selection cassette should be removed from the endogenous promoter, as with the current Puro-ΔTK variant, and placed under an exogenous promoter instead. Placement under an exogenous promoter allows its expression regardless of position in the genome, allowing for negative selection of re-integration products. A simple way to accomplish this would be to combine it with the PBase-ERT2 fusion, which is already driven by an exogenous promoter, using a 2A peptide. 2A peptides undergo self-cleavage following translation, leaving two independent proteins from a single initial peptide[270]. The final construct would then be PBase-ERT2-2A-ΔTK for FIAU selection. If fluorescence was desired for selection, as with the GFP-SDS variant, then a second red fluorescent protein could be used to mark re-integration. This could be accomplished with a PBase-ERT2-2A-RFP fusion.

Finally, it is possible that the benefits of COIN could be realized without integrating a selection cassette into the target genome. Evidence indicates that the homologous recombination machinery can also modify extra-chromosomal DNA. Pairs of incomplete plasmids can be digested and recombined outside of the genome with high efficiency[271,272] and these extra-chromosomal recombination events are linked to intra-chromosomal integrations[273]. Rather than inserting a fragment of DNA into the genome, one could instead select for extra-chromosomal recombination. The simplest method would be to split a selectable marker between two ends of a linear DNA and include some amount of overlapping homologous sequence. For expression of the selectable marker, the linear DNA would need to be circularized by HDR. Cells which performed this HDR mediated circularization could be isolated

by standard methods of selection. This would substantially simplify the process of developing COIN reagents and because no gRNA is required, would prevent any complications of off-target digestion. A mechanism for selecting against off-target integrations of the linear dsDNA would need to be developed but could be as simple as including a negative selection cassette in the middle of the linear DNA.

Reduction of off-target integrations

Much of current work in genome editing has focused on methods of inhibiting off-target DSBs, a source of potential mutagenesis throughout the genome. In contrast, relatively little attention has been given to the reduction of off-target integrations. Like DSB, off-target insertions can be highly mutagenic. Almost half occur within genes and most frequently within exons, thus, there is significant potential for loss-of-function mutations. In addition, donor DNAs often carry exogenous promoters which may inappropriately drive a gene's expression. Off-target integrations are also relatively common. The frequency has been estimated at 1 in 10^2-10^4 cells[6,147-149] and as high as 2% of cells obtain an off-target integration in our experiments. The reduction of off-target integrations will therefore be essential for the safety of genome editing techniques which utilize HDR. Understanding the mechanism of off-target integration will help identify interventions that selectively reduce its frequency without disrupting on-target integration.

It is commonly proposed that off-target integrations are mediated by NHEJ. Classic experiments for gene targeting indicate that linear DNAs are more frequently incorporated into the genome and the authors suggest that off-target integrations usually include the free ends of these linear molecules[144,147]. In agreement, the ratio of on:off-target integrations can be improved by including a negative selection cassette on the ends of linear molecules outside of

the homology arms[147]. This can be used to preferentially kill cells with off-target integrations, while preserving on-target integrations, which don't include regions outside of the homology arms. This method can improve the ratio of on:off-target integrations by as much as 10-fold in mouse ES cells[172], although it does not appear to work as well in other cell types[153].

After a DSB, the classic NHEJ pathway (cNHEJ) is initiated by the protection of DNA ends with heterodimers of KU70 and KU80[115-117,274]. In addition, the KU proteins act as a scaffold to recruit other required components such as DNA-PK, which may help bridge the gap between broken ends, and LIG4, which seals the break[118,120,128]. None of these proteins are required for HDR and therefore a selective reduction in off-target integrations could, in theory, be obtained by their inhibition. Several groups have additionally reported that a selective inhibitor of LIG4, called SCR7, increases HDR efficiencies in a variety of mammalian cell lines, and even in injected mouse embryos[275,276]. Presumably, inhibiting NHEJ stimulates the cell to use alternative pathways such as HDR. HDR was stimulated by up to 19-fold in a melanoma cell line[276] and up to 5-fold in HEK293 cells[276].

However, in mouse ESC, the cNHEJ pathway does not seem to be required for off-target integrations (See Appendix 6.7-6.9). We performed a knockdown of the KU70 protein, chemical inhibition of DNA-PK with NU7026, and chemically inhibition of DNA Ligase IV with SCR7. In addition, we attempted to alter the ends of donor DNAs to prevent their ligation. 5' hydroxyl groups were removed with shrimp alkaline phosphatase (SAP), dideoxynucleotides were added to the 3' ends with Terminal Deoxynucleotidyl Transferase (TdT), and the linkage of the last 3 nucleotides on each end were protected from nuclease digestion and subsequent removal by the use of phosphorothioate (PS) bonds. None of these provided any improvement in the ratio of on:off-target integrations or directly increased the frequency of on-target integration.

Altogether this suggests that in at least in mouse ESC, off-target integrations are not mediated by cNHEJ. A number of poorly understood end joining pathways have been

discovered in NHEJ deficient cells and are commonly referred to as alternative end joining (A-EJ). There is likely to be more than one alternative pathway, but as a group they are defined by not requiring the core proteins involved in cNHEJ, including KU70/80, DNA-PK, and LIG4[128,277-279]. One mechanism or set of mechanisms is loosely called micro-homology mediated end joining (MMEJ) because post-repair junctions typically exhibit small regions of homology from 1-30 bp. The details of this mechanism are still unclear but it has been shown in a number of organisms, including mouse and human, to require DNA Ligase III (LIG3)[280-284].

Micro-homologies may also be present in the cNHEJ. During processing, ends are frequently trimmed back or extended to produce small regions of complementary sequence that can stabilize ligation by LIG4[114,126]. cNHEJ has been described with 1-6bp micro-homologies in mammals, though these studies fail to account for potential contributions of MMEJ[285-287]. In yeast, cNHEJ rarely makes use of homologies greater than 4bp[113,114].

Off-target integrations appear to be evenly distributed at the chromosomal level[288] but within chromosomes may be pre-disposed to occur at certain hotspots[289,290]. Analysis of the mechanism of insertion is surprisingly limited, with just a few studies that analyze the sequence of DNA at the junction of the donor and host genome. Each of these reports only examines a handful of off-target integrations but, nonetheless, provide some interesting results. A study in Chinese Hamster Ovary (CHO) cell lines evaluated junctions for 23 sites after electroporation of a 1.3kb linear DNA[291]. The resulting integrations had an even distribution of 0-600bp removed from the ends indicating that loss of sequence from the donor ends is common. Note that this probably underestimates the extent of resection because the remaining 700bp was required for detection. Only 1/23 junctions between the donor DNA and host genome exhibited end joining without any detectable homology. 11/23 junctions were mediated by micro-homologies between 1-5bp, and the rest contained a range of insertions of up to 581bp. These were determined to derive from other regions of the host genome that also contained micro-homologies at the

junction. A more recent study analyzed the junctions at off-target integration sites in transgenic mice generated by injection of linear DNAs into single cell embryos[292]. They examined 38 junctions and also noted micro-homologies in 29/38 junctions. Interestingly, the length of micro-homology was longer in mice than in the CHO lines, ranging from 1-19bp with an average of 5.5bp. The remaining 9/38 sites contained a mixture of small deletions and insertions at the junction which are difficult to interpret.

The range of homology seen in CHO cells is thus consistent with either cNHEJ or MMEJ pathways, however the larger homologies seen in mice indicate MMEJ is likely to play a substantial role. The combined repertoire of off-target integrations may involve a combination of mechanisms but our results indicate that inhibition of cNHEJ alone is not sufficient to reduce off-target effects. A more thorough analysis of A-EJ requirements could prove extremely valuable. The rate of off-target integration can be easily quantified by introduction of a donor DNA with no homologous sequence to the host genome. This insert can drive expression of a fluorescent protein like GFP and the total number GFP+ cells can be quantified by flow cytometry. Similar experiments we have conducted produce off-target integrations in approximately 1% of cells. Within this background it would be relatively simple to inhibit or knockdown required components of MMEJ, cNHEJ, HDR and other A-EJ pathways individually or in combination, then look for decreases in the frequency of these off-target insertions. An excellent initial target would be LIG3, a factor required for MMEJ but not cNHEJ or HDR.

Chapter 6: Appendices

Appendix 6.1 TCF/LEF activity in developing forebrain

Defects in Wnt pathway regulation are implicated in a number of neurological disorders. Wnt dysregulation occurs in embryonal brain tumors such as medulloblastomas (estimated 15%[293]) and primitive neuroectodermal tumors (estimated 36%[294]). The Wnt pathway is also linked to Schizophrenia and related diseases through the Disrupted in Schizophrenia 1 (DISC1) gene. DISC1 mutations are risk factors for Schizophrenia, Autism spectrum disorders, major depression, and bipolar disorder[295,296], and mice with mutations in DISC1 display many schizophrenia-like behaviours[297]. DISC1 protein inhibits GSK3 activity, promotes β-catenin stability and stimulates Wnt target gene activation[298].

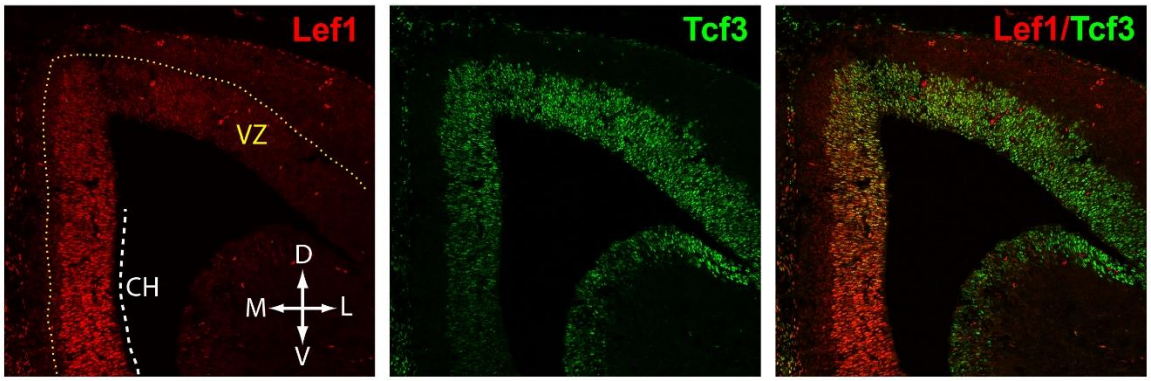
GSK3 and β-catenin both have a central role in normal brain development through control of neural progenitor cell (NPC) self-renewal and differentiation. Using Nestin driven Cre to target the NPC pool, both knocking out GSK3[299,300] and expression of stabilized β-catenin[301] led to a massive expansion of the neural progenitor pool and horizontal expansion of the cerebral cortex. Conversely, knocking out β-catenin decreased progenitor proliferation and increased neuronal differentiation[302].

Defects in the developing forebrain (telencephalon), specifically the cerebral cortex and hippocampus, have also been implicated in Autism spectrum disorders[303,304] and models of Schizophrenia[298,305]. In the telencephelon, the nuclei of proliferating neural stem cells (radial glia) are confined to a band surrounding the ventricles known as the ventricular zone (VZ)[306]. The proliferation and specification of these cells are directed to a large extent by extracellular morphogen gradients, including Wnts released from the dorso-medial signaling center known as the cortical hem (CH)[307]. Wnt3a released from the caudal region of the CH is essential for the proliferation of LEF1-expressing VZ cells near the CH and for formation of the hippocampus[308]. Phenotypes of LEF1 -/- mice show LEF1 is necessary for formation of the dentate gyrus region of the hippocampus[309].

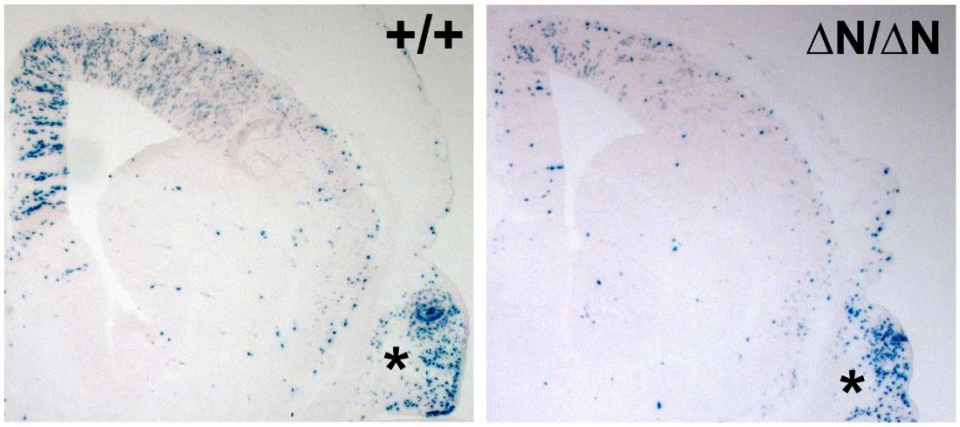
In addition to LEF1, TCF3 is also expressed in the developing telencephelon, whereas TCF1 and TCF4 are not expressed at high levels[309] (Fig. 6.1A). The role of TCF3 in the developing brain has not yet been examined; however, there is some genetic evidence of a LEF1-independent role for β-catenin[309]. We examined the expression of TCF3 and LEF1 in the telencephelon at e14.5, a period of hippocampal morphogenesis[309]. LEF1 was highly expressed in the VZ in a gradient with highest levels near the CH (Fig. 6.1A). TCF3 was also expressed at high levels in the VZ; however, it was found in an opposite gradient with low expression near the CH (Fig. 6.1A). The BAT-Gal transgenic mouse is an effective sensor of endogenous TCF/LEF-β-catenin transactivator activity in vivo[310]; BAT-Gal activity was high in cells near the CH and absent from cells far from the CH (Fig. 6.1B +/+). Thus, Wnt activity correlates with high LEF1 and low TCF3 expression in the telencephelon VZ. Our experiments using ESC provide a molecular framework for this observation. Wnt pathway stimulation increases LEF1 expression and inactivates TCF3, leading to its degradation.

We have generated a TCF3 knockin mutation (named TCF3ΔN) to disrupt TCF3-β-catenin interaction[311]. In ESC, Wnt pathway stimulation is unable to inactivate TCF3ΔN and stimulate Wnt target genes such as Axin2 and Cdx1[237]. Likewise, we found that the activation of BAT-Gal transgene in the brains of TCF3ΔN/ΔN mice was substantially diminished compared to TCF3+/+ littermates. In particular, BAT-Gal activation was markedly reduced in the telencephelon VZ of TCF3ΔN/ΔN embryos at e14.5 (Fig. 6.1B ΔN/ΔN). Since BAT-Gal was in the LEF1+ cells in WT mice, this result is consistent with the TCF3-LEF1 circuit being broken in TCF3ΔN/ΔN brains. Moreover, it highlights the potential importance of TCF3 in regulating Wnt/β-catenin responses in neural progenitor cells.

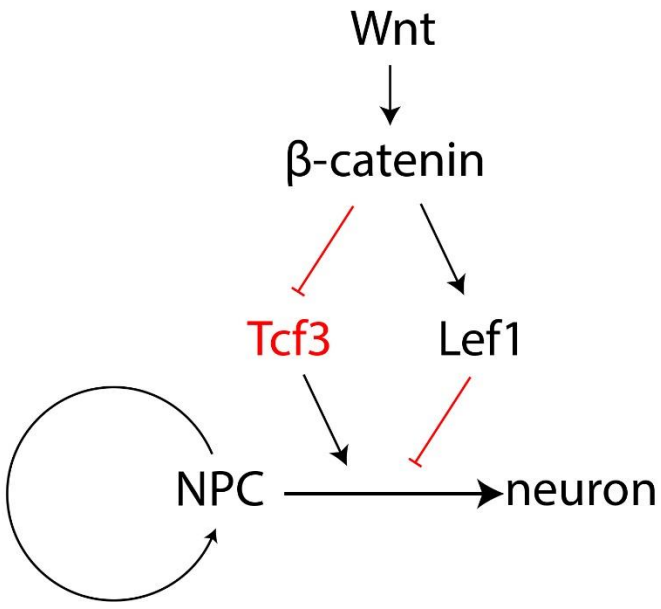
A


Lef1
VZ
CH
D
M
L
V
Tcf3
Lef1/Tcf3

B


+/+
*
ΔN/ΔN
*

C

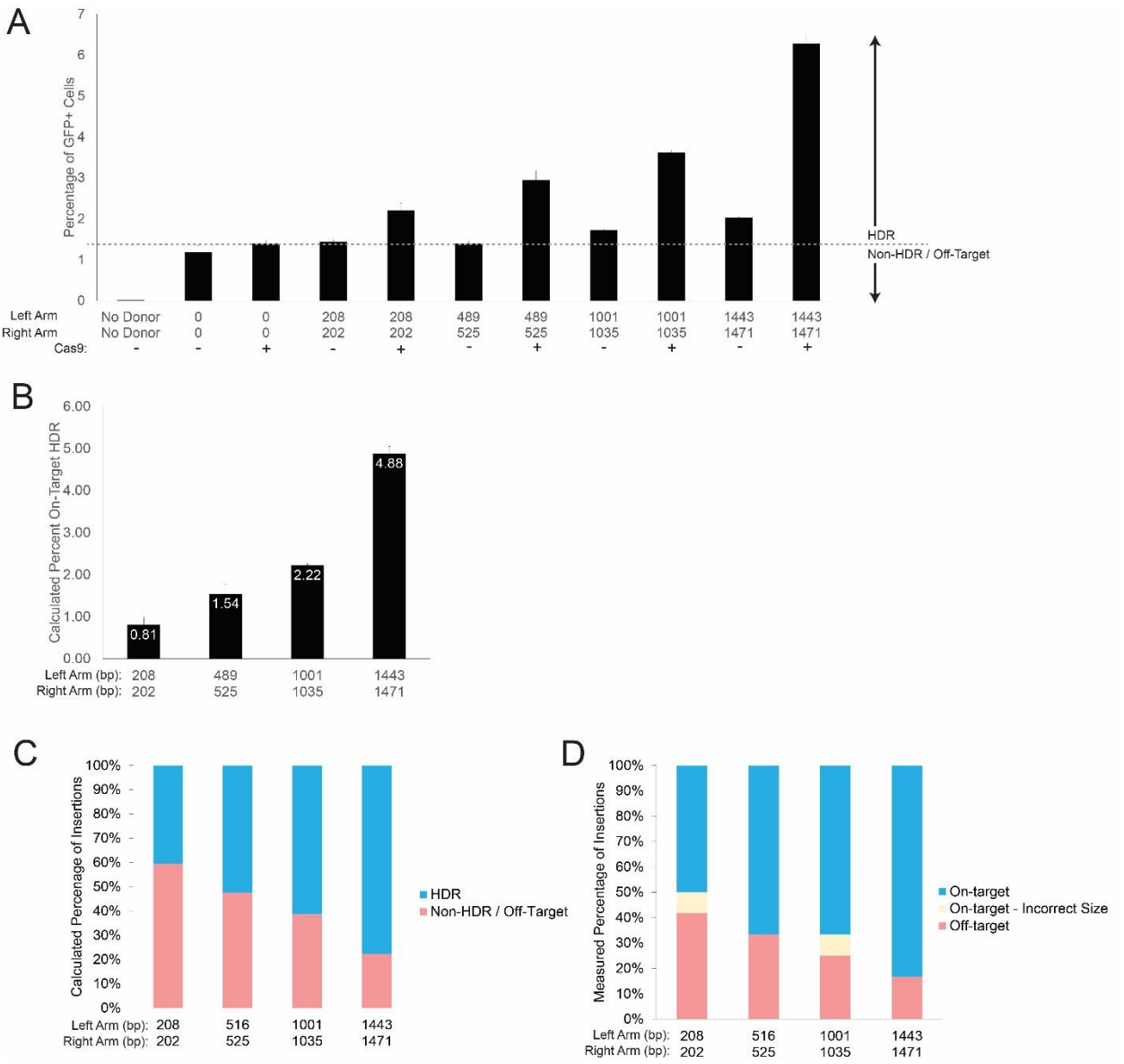

Wnt
β-catenin
Tcf3
Lef1
NPC
neuron

Appendix 6.2 On- and off-target integration with varied homology arm lengths

To evaluate on- and off-target insertion frequency, linear dsDNA donors where generated containing a GFP expression cassette flanked by different length homology arms targeting the *Lef1* gene. These were transfected into mouse ESC cells with or without Cas9 plus gRNA targeting *Lef1*. After 18 days, the percentage of GFP+ cells was quantified by flow cytometry to measure the number of stable integrations (Fig 6.2A). Biological duplicates were analyzed for each combination.

We first evaluated the percentage of GFP+ cells in the absence of homology arms. Because there is no homology this quantifies only the homology independent, or off-target, integrations. In the absence of Cas9/gRNA this was around 1% of cells (Fig 6.2A Column 2). The addition of Cas9 plus gRNA gave a slight but not significant increase, indicating that the vast majority of off-target integrations are not mediated by a Cas9 induced DSB (Fig 6.2A Column 3). In the presence of Cas9, adding homology arms of around 200bp to both the left and right side of the donor construct leads to an increase of around 0.81% (Fig 6.2A Column 5, Fig 6.2B Column 1). Increasing the length of the homology arms further progressively increases the rate of on-target insertion to nearly 5% of cells with 1443bp and 1471bp arms on the leftt and right, respectively (Fig 6.2A Columns 5-11, Fig 6.2B Columns 1-4).

From this data, the percent of on-target HDR was calculated for each homology arm length (Fig 6.2B) as well as the ratio of on-target (Fig 6.2C Blue columns) to off-target (Fig 6.2C Red columns) integrations. The latter was calculated by subtracting the percent integration of donor with no homology arms (Fig 6.2A Column 3 and dotted line). The predictions from this analysis match actual on- and off-target rates as quantified by PCR genotyping of 12 clones for each homology arm length (Fig 6.2D, data from Fig 4.2).

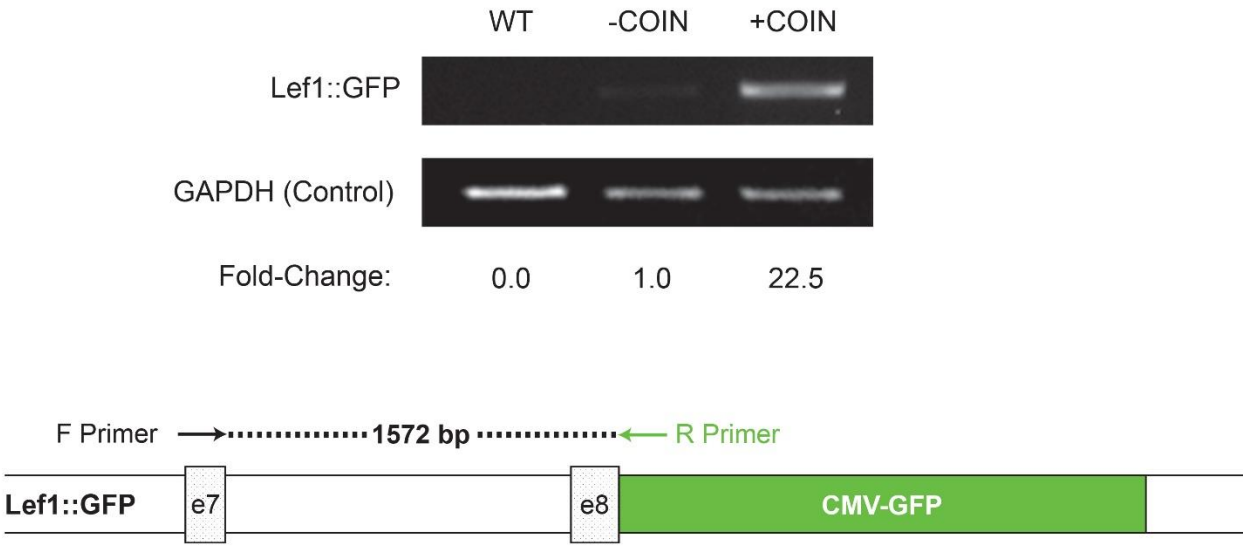

A
Percentage of GFP+ Cells
HDR
Non-HDR / Off-Target
Left Arm
Right Arm
Cas9:
No Donor
0 0 - 0 0 + 208 202 - 208 202 + 489 525 - 489 525 + 1001 1035 - 1001 1035 + 1443 1471 - 1443 1471 +
B
Calculated Percent On-Target HDR
0.81
1.54
2.22
4.88
Left Arm (bp): 208 489 1001 1443
Right Arm (bp): 202 525 1035 1471
C
Calculated Percenage of Insertions
HDR
Non-HDR / Off-Target
Left Arm (bp): 208 516 1001 1443
Right Arm (bp): 202 525 1035 1471
D
Measured Percentage of Insertions
On-target
On-target - Incorrect Size
Off-target
Left Arm (bp): 208 516 1001 1443
Right Arm (bp): 202 525 1035 1471

Appendix 6.3 Semi-quantitative PCR analysis of COIN for *Lef1* and *Tcf1* locus

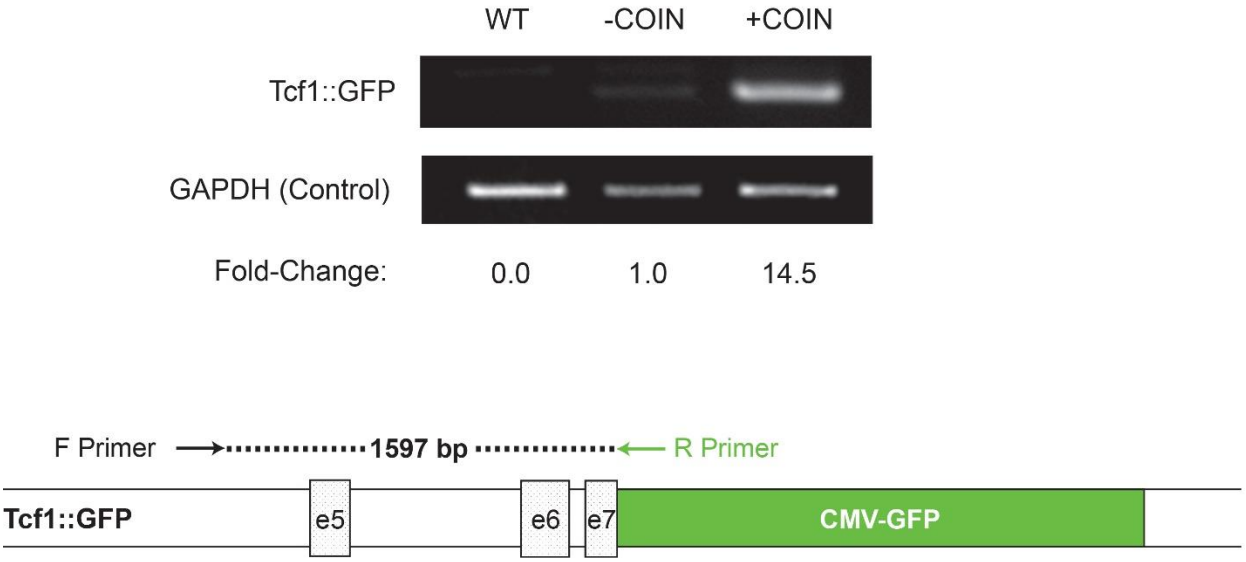
To evaluate the effects of COIN, donors were generated containing a CMV promoter driving expression of GFP with homology arms targeting either the *Lef1* or *Tcf1* gene (Figure 4.3, 4.4). These were transfected into mouse ESC cells along with Cas9 and gRNA targeting the appropriate gene. After 18 days, the percentage of GFP+ cells was quantified by flow cytometry to measure the number of stable integrations. This method shows a substantial enrichment of GFP+ cells when the COIN method is used, but does not eliminate the possibility that these effects could be caused by an increase in off-target, rather than on-target integrations.

Here we perform a semi-quantitative PCR to confirm a benefit specific to on-target integrations. Each PCR product is generated with one external primer that recognizes either the *Lef1*(A) or *Tcf1*(B) gene outside of the donor homology arms, and one internal primer that is specific for the insert. Densitometry of the PCR products is used to compare the relative frequency of integration in a population of cells with and without COIN. Each PCR product is specific for the on-target integration and confirms that COIN leads to substantial enrichment of on-target integrations at each site.

A


WT
-COIN
+COIN
Lef1::GFP
GAPDH (Control)
Fold-Change:
0.0
1.0
22.5
F Primer
1572 bp
R Primer
Lef1::GFP
e7
e8
CMV-GFP

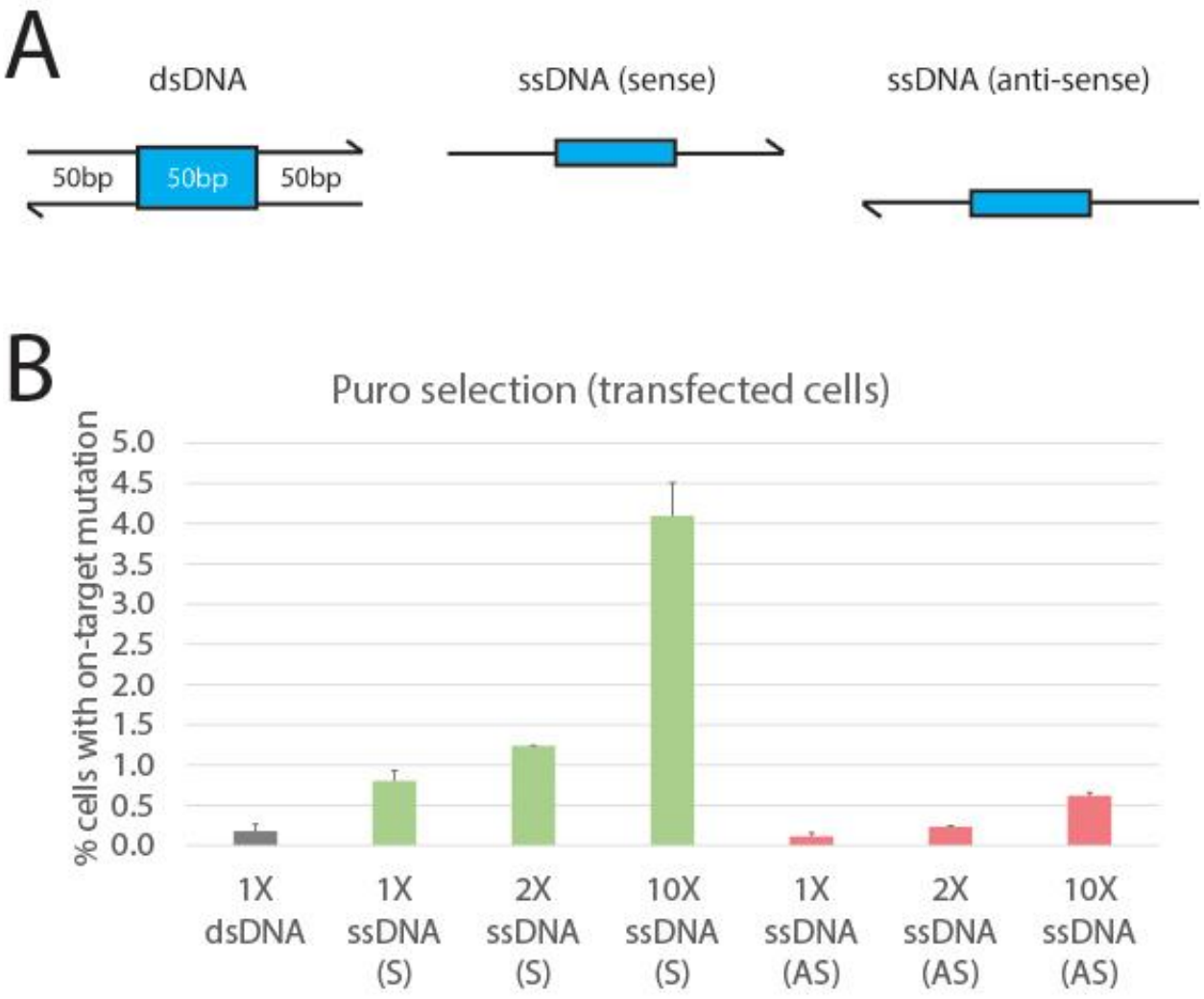
B


WT
-COIN
+COIN
Tcf1::GFP
GAPDH (Control)
Fold-Change:
0.0
1.0
14.5
F Primer
1597 bp
R Primer
Tcf1::GFP
e5
e6
e7
CMV-GFP

Appendix 6.4 Comparison of small dsDNA and ssDNA HDR donors

De novo synthesis of long DNAs is becoming increasingly inexpensive. This is particularly true for linear dsDNAs which can be routinely generated in lengths up to 2kb for a few hundred dollars. ssDNA synthesis is currently limited to lengths of 200bp, is more expensive for a typical genome editing project, and exhibits a lower frequency of full length product in the absence of purification. To evaluate whether ssDNA outperforms dsDNA, we first generated a dsDNA with a 50bp insert and 50bp homology arms on each side. We then ordered ssDNA oligomers corresponding to the sense or anti-sense strand of this dsDNA (Fig 6.4A).

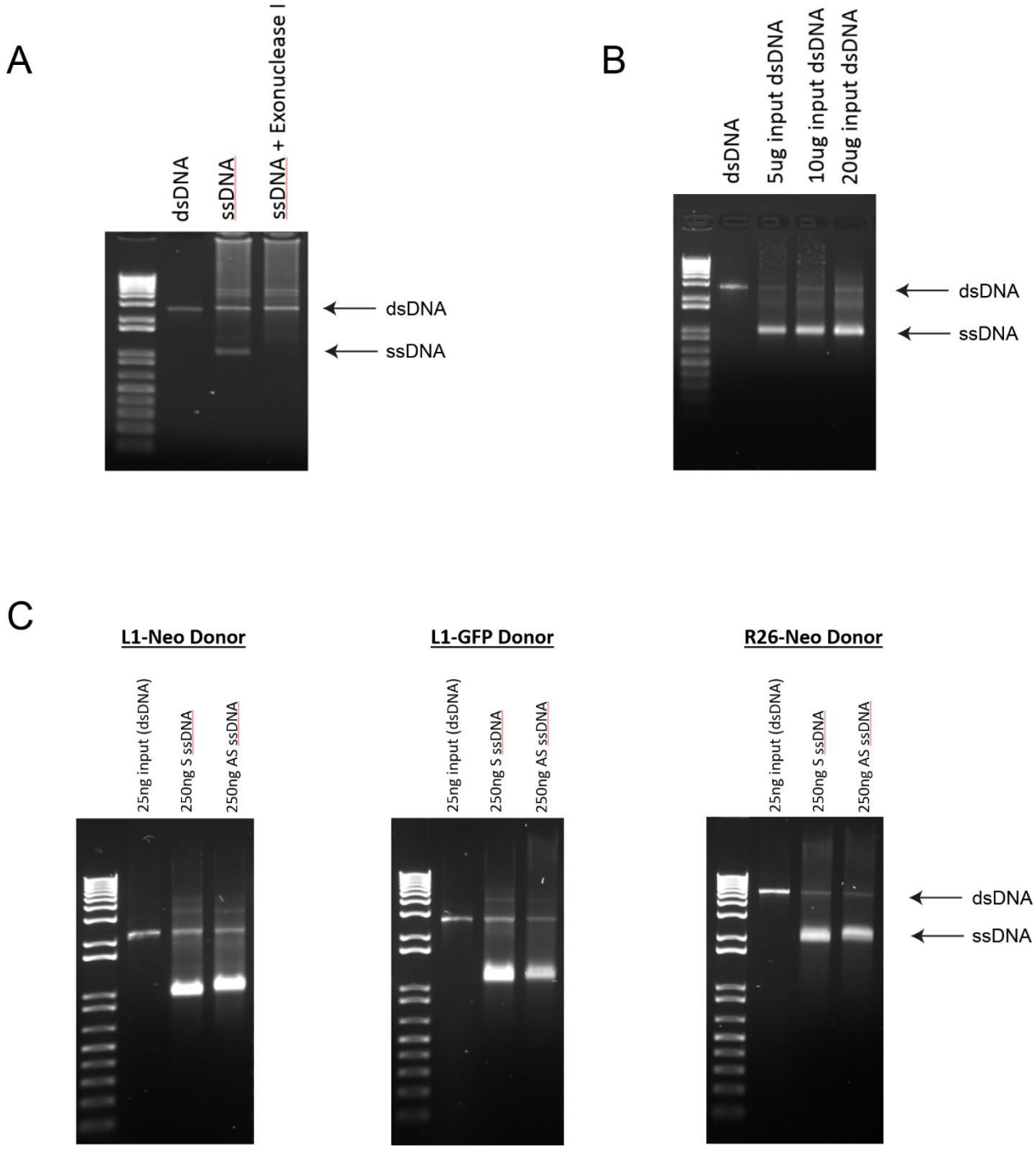
Donors were targeted to the *Lef1* locus in mouse ESC and were co-transfected with Cas9 plus gRNA. Our results indicate that a strand bias exists for this site and that the sense strand (S) ssDNA significantly outperforms equimolar amounts of dsDNA (Fig 6.4B). Conversely, the anti-sense ssDNA (AS) exhibits a lower targeting efficiency than the dsDNA. Increasing the concentration of both sense and anti-sense ssDNAs increases targeting efficiency with the 10X ssDNA mutating more than 4% of transfected cells.


A
dsDNA
ssDNA (sense)
ssDNA (anti-sense)
50bp
50bp
50bp
B
Puro selection (transfected cells)
% cells with on-target mutation
5.0
4.5
4.0
3.5
3.0
2.5
2.0
1.5
1.0
0.5
0.0
1X dsDNA
1X ssDNA (S)
2X ssDNA (S)
10X ssDNA (S)
1X ssDNA (AS)
2X ssDNA (AS)
10X ssDNA (AS)

Appendix 6.5 Generation of long ssDNA from biotinylated PCR products

Based on results indicating that short ssDNA are better HDR substrates than dsDNA (Appendix 6.4), long ssDNA may also be preferable. Current *de novo* synthesis of ssDNA is limited to 200bp. To generate long ssDNA products we have adapted a method used for the large scale production of ssDNA for oligonucleotide arrays[312]. Biotinylated primers are used to generate a PCR product with biotinylation on one strand only. The dsDNA is then denatured using a highly basic solution and the biotynlated strand is removed by strepatavidin-coupled magnetic beads. The remaining DNA can then be column purified out of solution. This process enriches for ssDNA from the non-biotinylated strand which can be identified by Exonuclease I digestion (Fig 6.5A)[313]. Substantial improvement in the ratio of ssDNA to residual dsDNA was obtained by converting to non-stick micro-centrifuge tubes (Fig 6.5B, Ambion #AM12450). The maximum capacity of MyOne Streptavidin beads (50uL) was examined and appears to be greater than 20ug for a 2.7kb dsDNA. There was no significant reduction in the ratio of ssDNA to residual dsDNA (Fig 6.5B).

The final protocol can generate highly enriched pools of ssDNA from either strand of the three different PCR products we tested (Fig 6.5C). These range in size from 2.3kb up to 6kb, indicating the protocol works even for very large constructs. In Figure 6.5C, 25ng of the input dsDNA is being compared to 250ng of purified sense (S) or anti-sense (AS) ssDNA. The first lane therefore represents 10% of the total amount of DNA in the second and third columns. Comparing the residual dsDNA for each reaction to this control, less than 10% of the dsDNA product remains, often substantially so. This represents a universal and inexpensive method for generating long ssDNA from PCR products.


A
dsDNA
ssDNA
ssDNA + Exonuclease I
dsDNA
ssDNA
B
dsDNA
5ug input dsDNA
10ug input dsDNA
20ug input dsDNA
dsDNA
ssDNA
C
L1-Neo Donor
L1-GFP Donor
R26-Neo Donor
25ng input (dsDNA)
250ng S ssDNA
250ng AS ssDNA
dsDNA
ssDNA

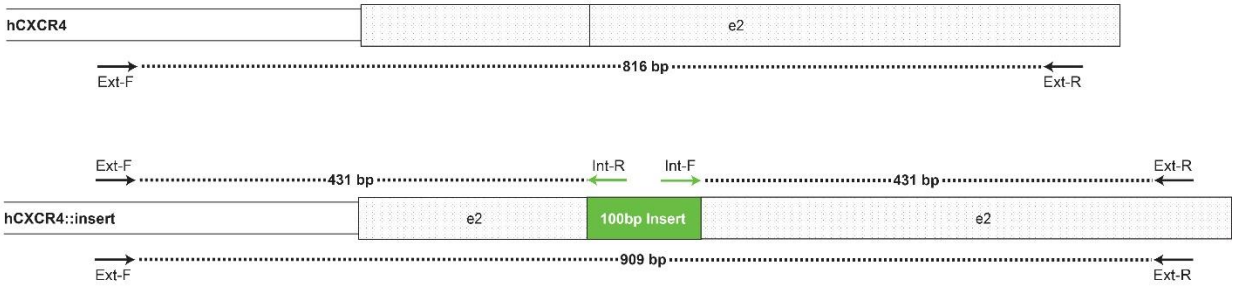
Appendix 6.6 COIN in human ESC

Tools for genome editing have the potential to revolutionize human medicine. Cells from patients can be obtained with relative ease, converted into induced pluripotent stem cells (iPSC), genetically modified in vitro, and then differentiated to relevant cell types. These cells may either be re-introduced into patients or used to study human disease in culture. However, human and other primate cell lines are up to 100-fold less likely to integrate exogenous DNA as compared to rodent cell lines[260-263]. Even generating very small insertions of less than 20bp, HDR efficiencies following a Cas9 induced DSB were 2-4% for human iPSC[111] and 1-2% for human embryonic stem cells (hESC)[169]. Thus, improvements of HDR efficiencies in human pluripotent cell lines would be a useful application of COIN.

Here we analyzed the effects of COIN for targeting a 100bp insert to the Cxcr4 locus in H9 hESC. Linear dsDNA donors along with plasmids encoding Cas9 plus gRNAs were transfected into cells using Lipofectamine 3000. The COIN selection gRNA and donor DNA were targeted to the AAVS1 locus. After integration, puromycin resistance is driven by the endogenous AAVS1 promoter.

Pools of cells were analyzed by semi-quantitative PCRs showing the relative frequency of integration. Primers external to the insert (Fig 6.6A Ext-F and Ext-R) and internal (Fig 6.6A Int-F and Int-R) were used to identify different products. PCR products produced from one external and one internal primer are specific for the on-target integration (Fig 6.6B Left side), thus no product is generated in WT cells. In the absence of COIN, there is a barely detectable band for one primer set and no detected band for the other, indicating very low levels of integration in the population. In contrast, when COIN selection is used, there is a substantial increase in the frequency of integration. Densitometry indicates a 26.9 fold increase for one primer set. The other primer set was not calculated (N.C.) because there was no detectable band in the absence of COIN. PCR products obtained from using both external primers show the WT product at 816 bp (Fig 6.6B Right side). Consistent with the increase in integration frequency, the 909 bp integration product is only visible in the COIN selected samples. Densitometry indicates the insert band is ~12.5% the density of the WT band in these samples, suggesting that 25% of cells contain a mutant allele. A PCR product generated from the human beta hemoglobin gene (hHBB) is used as a loading control.

A


hCXCR4
e2
Ext-F
816 bp
Ext-R
Ext-F
431 bp
Int-R
Int-F
431 bp
Ext-R
hCXCR4::insert
e2
100bp Insert
e2
Ext-F
909 bp
Ext-R

B


WT
-COIN
+COIN
Primers: Ext-F / Int-R
Densitometry:
0.0
1.0
26.9
Primers: Int-F / Ext-R
Densitometry:
0.0
0.0
N.C.
WT
-COIN
+COIN
Primers: Ext-F / Ext-R
909 bp (Insert)
816 bp (WT)
Primers: hHBB (control)

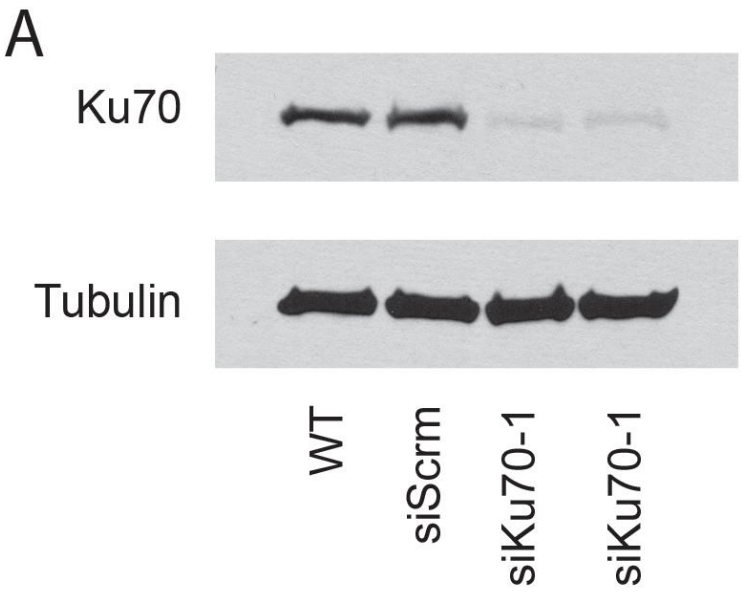
Appendix 6.7 Effects of Ku70 knockdown on HDR in mouse ESC

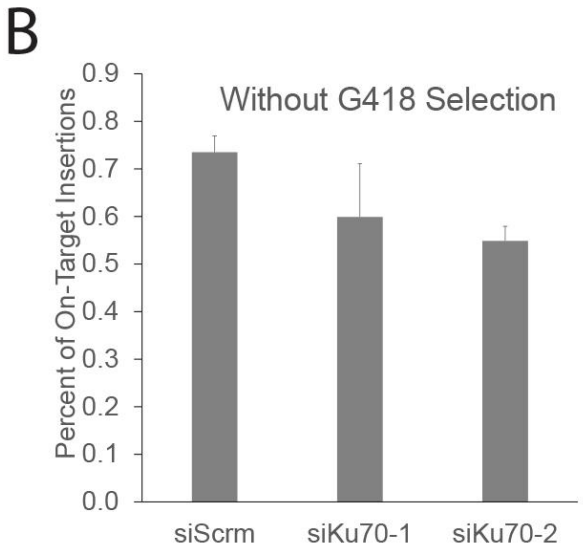
It is commonly proposed that off-target integrations are mediated by NHEJ[144,147]. In agreement with this, the ratio of on:off-target integrations can be improved by inclusion of a negative selection cassette outside of donor homology arms[147]. This preferentially kills cells which integrate the entire molecule rather than just the region within the homology arms. This method can improve the ratio of on:off-target integrations by as much as 10-fold in mouse ES cells[172], although it does not appear to work as well in other cell types[153].

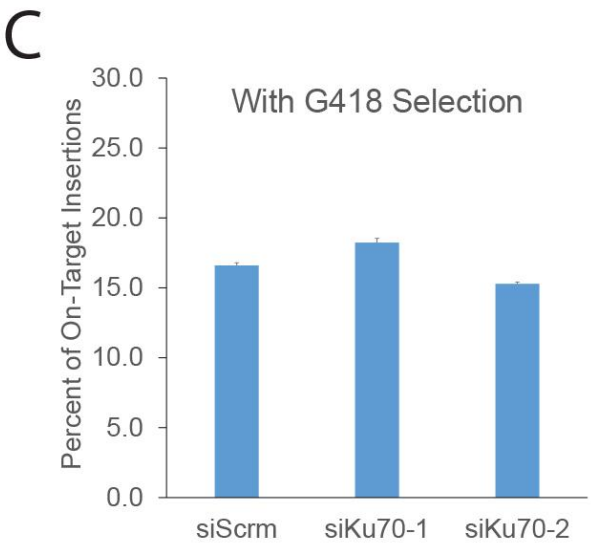
The classic NHEJ pathway (cNHEJ) is initiated by heterodimers of Ku70 and Ku80 which encircle the free ends of DNA surrounding a DSB and protect them from resection. In addition, they act as a scaffold to recruit a complex of proteins which mediate repair[115-117,274].

Several groups have recently reported that inhibition of NHEJ can stimulate increased HDR frequencies in mammalian cell lines[275,276]. One of these examined the effects of transient knockdown of either Ku70 or Ku80 in HEK293 cells. This was initiated at the time of transfection by including shRNAs on the Cas9 plasmid[276]. Results indicated a 2-3 fold increase in HDR efficiencies coupled with a separate 2-fold reduction in NHEJ. The combined effect should lead to a substantial improvement in the on:off-target ratio.

We performed a similar experiment in mouse ESC. Ku70 was knocked down at the time of transfection with two different siRNAs (Fig 6.7A). Each siRNA led to a reduction of greater than 95%. The frequency of on-target HDR (Fig 6.7B) and ratio of on:off-target insertions (Fig 6.7C) was measured by quantitative PCR as described in chapter 4. In contrast to results from HEK293 cells, knockdown of Ku70 had no effect on either parameter in mouse ESC[276].


A
Ku70
Tubulin
WT
siScrm
siKu70-1
siKu70-1


B
Without G418 Selection
Percent of On-Target Insertions
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0
siScrm
siKu70-1
siKu70-2


C
With G418 Selection
Percent of On-Target Insertions
30.0
25.0
20.0
15.0
10.0
5.0
0.0
siScrm
siKu70-1
siKu70-2

Appendix 6.8 Effects of DNA Ligase IV and DNA-PK inhibition on HDR in mouse ESC

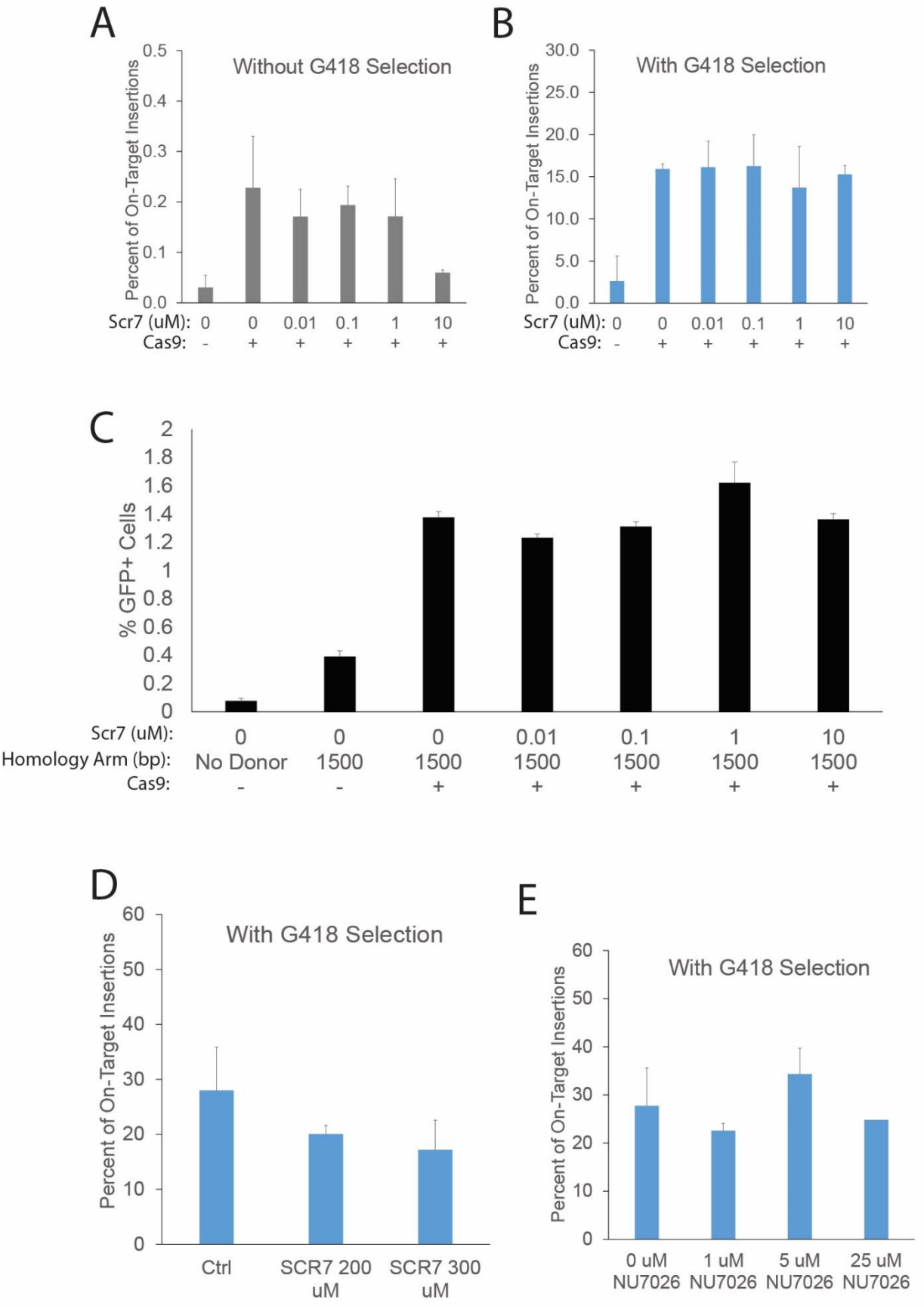
Following protection of free ends by the Ku proteins, a number of additional proteins are recruited to process the repair. This includes DNA-dependent protein kinase (DNA-PK) which is proposed to act as a scaffold, and DNA Ligase IV (LIG4) which seals the break. Both LIG4 and the kinase activity of DNA-PK are required for cNHEJ[118,120,128].

Two groups have recently reported that chemical inhibition of LIG4 with SCR7 can stimulate increased HDR frequencies of up to 3-fold in a human epithelial cell line, 19-fold in a melanoma cell line[275], and up to 5-fold in HEK293 cells[276]. The optimum SCR7 concentration was determined to be 0.01 uM, 1 uM, and 10 uM, respectively, although the effect of going 10-fold higher or lower was not substantial[263].

We tested SCR7 in mouse ESC at concentrations between 0.01 uM – 10 uM. The frequency of on-target HDR (Fig 6.8A) and ratio of on:off-target insertions (Fig 6.7B) was measured by quantitative PCR, as described in chapter 4, using a donor DNA with 200bp homology arms to the *Lef1* locus. The frequency of insertion of a GFP expression cassette using 1500bp homology arms was also tested and quantified by flow cytometry (Fig 6.7C). For all experiments, no improvement in either the on-target frequency or the on:off-target ratio was seen at any SCR7 concentration.

Previous work with SCR7 demonstrates a wide range of effective concentrations in various cancer cell lines[315]. It is therefore possible that mouse ESC might be less sensitive to SCR7 and require a higher concentration. We performed a kill curve to determine the maximum dose of SCR7 that was tolerated by ESC (data not shown). Approximately 50% of cells survived after 72 hours with 200uM SCR7, less than 30% survive with 300uM, and no cells survived at higher concentrations. We evaluated the on:off target ratio by quantitative PCR at these elevated concentrations and again observed no benefit (Fig 6.7D). If anything, the ratio of on:off target insertions was slightly lower.

We also examined the effect of NU7026, an inhibitor of DNA-PK that has been shown to effectively inhibit cNHEJ (Fig 6.7E)[316]. A range of concentrations from 0-100uM again provided no benefit to the on:off target ratio.


A
Without G418 Selection
Percent of On-Target Insertions
0.5
0.4
0.3
0.2
0.1
0.0
Scr7 (uM): 0 0 0.01 0.1 1 10
Cas9: - + + + + +
B
With G418 Selection
Percent of On-Target Insertions
30.0
25.0
20.0
15.0
10.0
5.0
0.0
Scr7 (uM): 0 0 0.01 0.1 1 10
Cas9: - + + + + +
C
% GFP+ Cells
2
1.8
1.6
1.4
1.2
1
0.8
0.6
0.4
0.2
0
Scr7 (uM): 0 0 0 0.01 0.1 1 10
Homology Arm (bp): No Donor 1500 1500 1500 1500 1500 1500
Cas9: - - + + + + +
D
With G418 Selection
Percent of On-Target Insertions
60
50
40
30
20
10
0
Ctrl
SCR7 200 uM
SCR7 300 uM
E
With G418 Selection
Percent of On-Target Insertions
60
50
40
30
20
10
0
0 uM NU7026
1 uM NU7026
5 uM NU7026
25 uM NU7026

Appendix 6.9 Modification of DNA ends to prevent end joining

cNHEJ involves the direct ligation of free ends produced by a DSB. Although not always necessary, the last few bases are frequently processed prior to ligation. This involves either small extensions of 1-4bp on the 3' ends by polymerases, or removal of 1-4bp from the ends by nucleases[274]. If off-target integrations are mediated by direct ligation of ends, then modifying the ends to prevent ligation could reduce the frequency of off-target insertion.

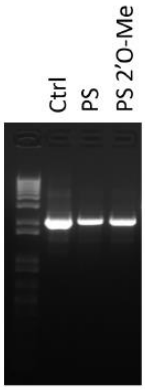
We have altered the donor DNA ends in order to interfere with cNHEJ. First, the phosphodiester linkage for the last 3 bases were modified with phosphorothioate bonds (PS)[317]. This replaces an oxygen atom with sulfur in the phosphate group and renders the bond resistant to endo- and exo-nuclease digestion in mammalian cells. It is commonly used to increase the lifespan of antisense oligonucleotides[317]. These modifications are extremely inexpensive and can be added to the ends of short primers during de novo synthesis. These modified primers can then be used to amplify a dsDNA by PCR. The PS bond has a slightly different rotation that can interfere with base-pairing and alter the Tm. The addition of a 2'-O-methyl group on the pentose ring (PS 2'O-Me) of the base corrects for this without compromising the resistance to digestion[317]. This can lead to better performance in PCR reactions or potentially in HDR reactions if base pairing is required.

We generated a set of dsDNA HDR donors by PCR using primers modified with either PS or PS 2'O-Me bonds for the last 3 bases on either side (Fig 6.9A). These PCR products were then further modified by two additional methods. First, the 5' ends were protected from ligation by removal of 5' phosphate groups with shrimp alkaline phosphatase (SAP). Second, the 3' ends were protected from ligation by addition of a single dideoxynucleotide (ddATP) to the 3' end using Terminal deoxy Transferase (TdT). Donors were transfected with Cas9 and a gRNA targeting the *Lef1* locus, then analyzed for on-target HDR and on:off-target ratio using the quantitative PCR assay described in Chapter 4.

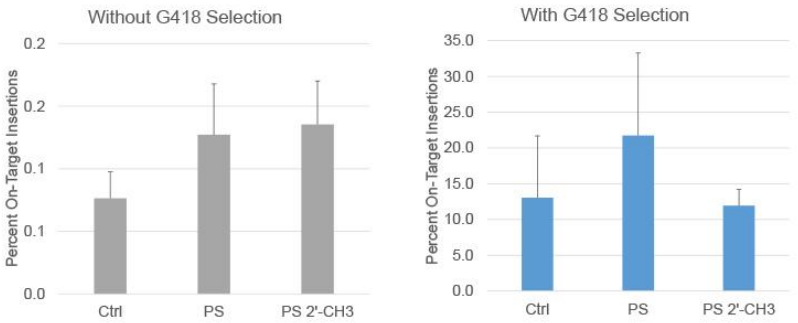
Neither of the phosphorothioate bonds provided a significant benefit to either the on-target HDR frequency or the on:off target ratio, however, the average on:off target ratio was slightly higher for the PS bond (Fig 6.9B). Similarly, neither method of end protection gave a significant increase to either the on-target HDR frequency or the on:off target ratio (Fig 6.9C). The average on:off target ratio was slightly higher with TdT treatment and slightly lower with SAP, although neither trend was significant.

Since TdT treatment and the PS bond exhibited a positive trend for the on:off target ratio, these were evaluated again both alone and in combination. In this second experiment, neither modification showed any benefit either alone or in combination (Fig 6.9D).

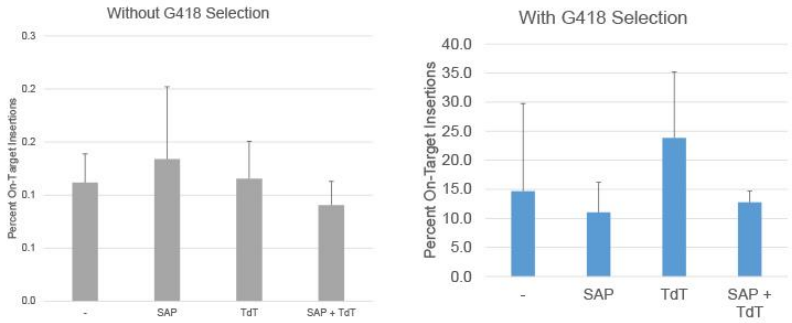
A


Ctrl
PS
PS 2'O-Me

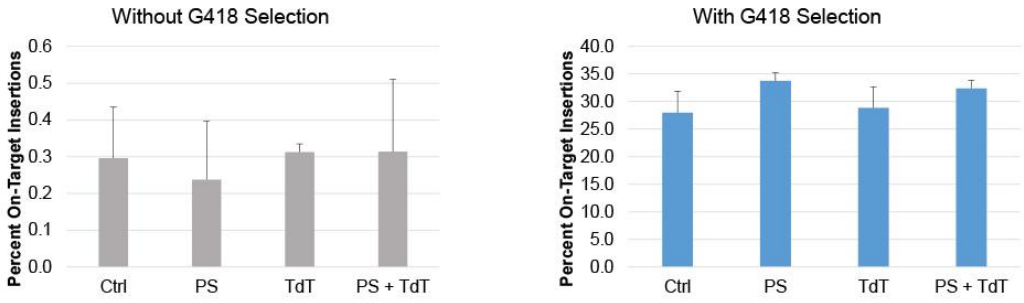
B


Without G418 Selection
Percent On-Target Insertions
0.2
0.2
0.1
0.1
0.0
Ctrl
PS
PS 2'-CH3
With G418 Selection
Percent On-Target Insertions
35.0
30.0
25.0
20.0
15.0
10.0
5.0
0.0
Ctrl
PS
PS 2'-CH3

C


Without G418 Selection
Percent On-Target Insertions
0.3
0.2
0.2
0.1
0.1
0.0
-
SAP
TdT
SAP + TdT
With G418 Selection
Percent On-Target Insertions
40.0
35.0
30.0
25.0
20.0
15.0
10.0
5.0
0.0
-
SAP
TdT
SAP + TdT

D


Without G418 Selection
Percent On-Target Insertions
0.6
0.5
0.4
0.3
0.2
0.1
0.0
Ctrl
PS
TdT
PS + TdT
With G418 Selection
Percent On-Target Insertions
40.0
35.0
30.0
25.0
20.0
15.0
10.0
5.0
0.0
Ctrl
PS
TdT
PS + TdT

Appendix 6.10 Self Destruct Sequence

COIN provides substantial increases in targeting efficiency, but leaves a selection cassette behind in the genome. To provide a convenient method to remove the selection cassette, we have generated a set of specialized donors referred to as self-destruct sequences (SDS). SDS contain selectable markers for use with COIN as well as all of the components required to stimulate their own removal (Fig 6.10A,C).

The SDS contains two major components. First, it contains a selectable marker. We have generated constructs that express either GFP (Fig 6.10A) or the Puro-ΔTK fusion protein (Fig 6.10B). The first half of Puro-ΔTK provides positive selection for Puromycin resistance. The second half is a variant of the Thymidine Kinase enzyme from Herpes Simplex virus (ΔTK)[266]. ΔTK can convert normally inert nucleoside analogues including Fialuridine (FIAU) into toxic metabolites. Thus when FIAU is added to cell media, it will preferentially kill cells still expressing the SDS.

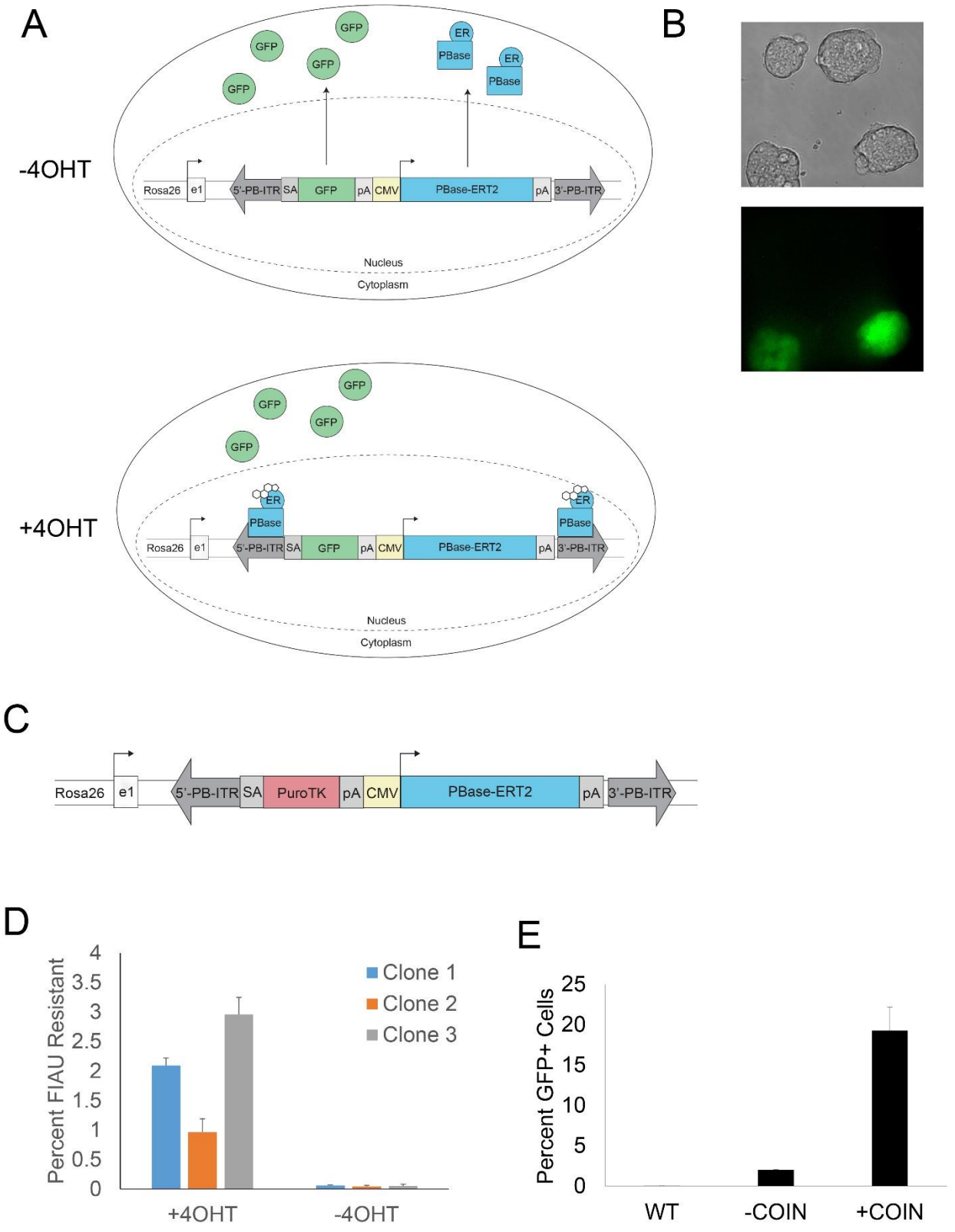
The second component of the SDS is the machinery required for removal. The entire SDS sequence is flanked by inverted terminal repeats (ITR) which are recognized by the PiggyBac transposase (PBase). Upon expression of PBase, the sequence in between the ITR undergoes a scarless excision, meaning no residual sequence is left in the genome and it is restored to the WT genomic DNA. A promoter driving expression of PBase is included between the ITR and will stimulate its own excision from the genome. To prevent immediate removal, PBase is fused to a modified estrogen receptor (ERT2) that is preferentially stimulated by 4-Hydroxy Tamoxifen (4OHT)[265]. In the absence of 4OHT, PBase-ERT2 is sequestered away from the nucleus (Fig 6.10A Top). When excision is desired, 4OHT is added to the media to stimulate nuclear translocation of PBase (Fig 6.10A Bottom).

GFP-SDS donor constructs were generated to target the Rosa26 locus in mouse ESC. Cells were transfected and GFP+ integrants were clonally isolated after 18 days. In the absence of 4-OHT, all colonies maintained GFP expression. When treated with 1uM 4OHT for three days, 1-2% of colonies lost GFP expression (Fig 6.10B), suggesting efficient excision of the SDS.

PuroΔTK-SDS constructs were similarly targeted to the Rosa26 locus. Cells with integration of PuroΔTK are resistant to Puromycin and sensitive to FIAU. Puro resistant colonies were clonally isolated and expanded. Three independent clones were then treated +/- 1uM 4OHT for three days and then subjected to FIAU selection (Fig 6.10D). 1-3% of colonies for each clone were FIAU resistant following 4OHT treatment, suggesting efficient excision, while less than 0.05% of colonies were resistant in the absence of 4OHT.

These results indicate that SDS can be efficiently removed by the addition of 4OHT to media. We also tested whether SDS could be used to stimulate an increase in targeting efficiency by COIN. PuroΔTK-SDS constructs were targeted to the Rosa26 locus by the inclusion of 500bp homology arms and used to stimulate integration of a GFP expression cassette at the *Lef1* locus. The percent of GFP+ cells was measured by flow cytometry after 18 days (Fig 6.10E). As with other Rosa26 selection donors, the SDS increased targeting efficiency by more than 10-fold, bringing the number of GFP+ cells to nearly 20%.

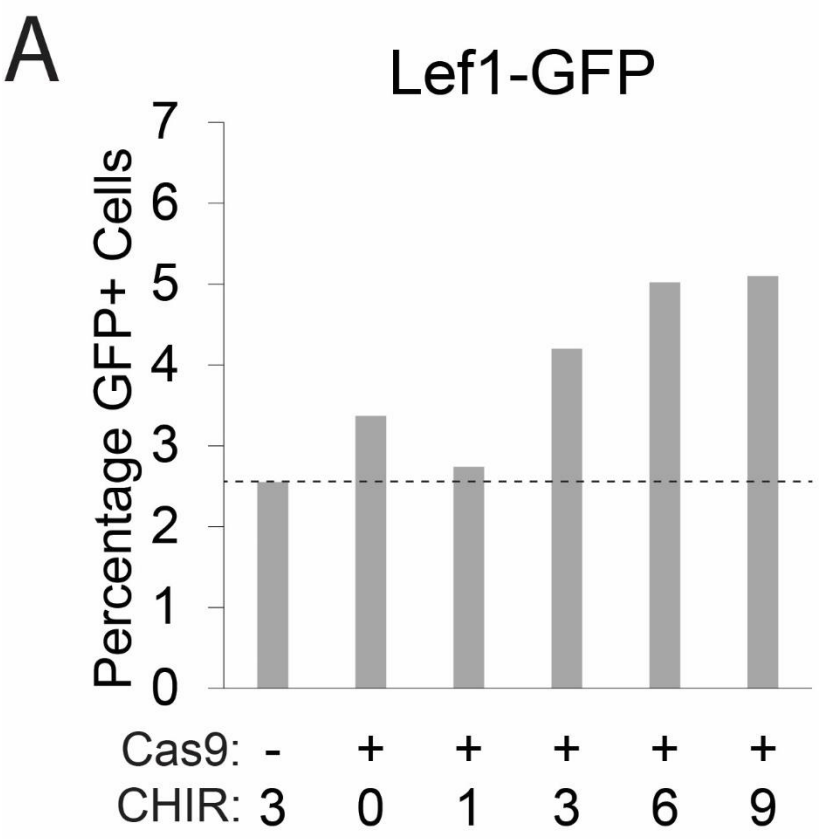
*Data for figure 6.10D was generated in collaboration with Matthew MacDougall

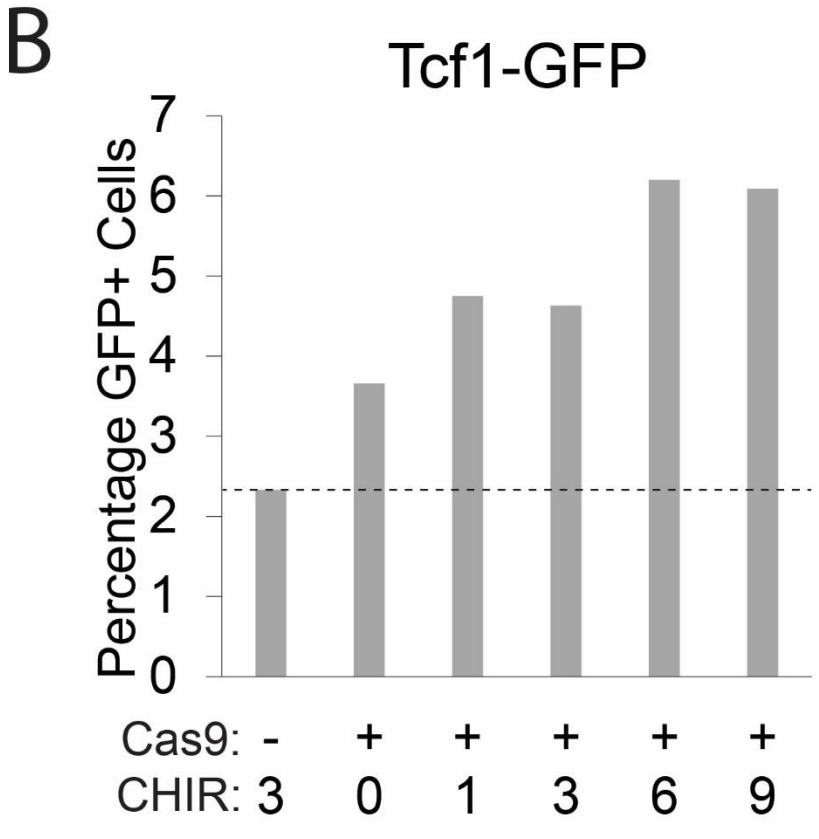

A
-4OHT
GFP
ER
PBase
Rosa26
e1
5'-PB-ITR
SA
GFP
pA
CMV
PBase-ERT2
3'-PB-ITR
Nucleus
Cytoplasm
+4OHT
B
C
PuroTK
D
Percent FIAU Resistant
Clone 1
Clone 2
Clone 3
+4OHT
-4OHT
E
Percent GFP+ Cells
WT
-COIN
+COIN

Appendix 6.11 CHIR stimulates increased integration

Donors were generated containing a GFP expression cassette flanked by 200bp homology arms targeting either the *Lef1*(A) or *Tcf1*(B) gene. These were transfected into mouse ESC with and without Cas9 plus a gRNA targeting the appropriate location. Cells were grown in 15% FBS supplemented with LIF and 0-9 uM CHIR during transfections. Differences in CHIR were maintained for 3 days after transfection, after which all cells were converted to 3uM CHIR and maintained for an additional 14 days in culture.

Stable integration of the GFP cassette was quantified by flow cytometry. For both *Lef1*(A) or *Tcf1*(B) target sites, increasing the concentration of CHIR significantly increases the rate of integration. This could be caused by transcriptional effects, as both *Lef1* and *Tcf1* are Wnt target genes. Alternatively, it could be a global effect of CHIR not directly related to transcription. For example, CHIR promotes the naïve ESC state in these culture conditions and ESC may be more prone to undergo HDR. This can be tested in two ways. First, integration at genes that are not transcriptionally regulated by Wnt stimulation can be examined. Second, transcription of genes can be stimulated directly to evaluate effects on targeting efficiency. This can be accomplished using a nuclease-dead variant of Cas9 that has been fused to one or more VP64 activation domains. These engineered proteins can efficiently stimulate transcription throughout the genome when targeted shortly upstream of the transcription start site[318].


A
Lef1-GFP
Percentage GFP+ Cells
0
1
2
3
4
5
6
7
Cas9: - + + + + +
CHIR: 3 0 1 3 6 9


B
Tcf1-GFP
Percentage GFP+ Cells
0
1
2
3
4
5
6
7
Cas9: - + + + + +
CHIR: 3 0 1 3 6 9

Appendix 6.12 Permission for reuse for Nucleic Acids Research

Obtained from https://academic.oup.com/nar/pages/Policies on February 28th, 2017:

“As part of the license agreement, authors may use their own material in other publications provided that the Journal is acknowledged as the original place of publication and Oxford University Press as the Publisher.”

Appendix 6.13 Permission for reuse for Cell Reports

Obtained from http://www.cell.com/cell-reports/faq on February 28th, 2017:

““Authors can choose to publish their work under one of two Creative Commons licenses:

One is the Creative Commons Attribution License 4.0 International License (CC BY 4.0). This license allows users to alter and build upon the article and then distribute the resulting work, even commercially, and thus encourages maximum use and redistribution.

The other is the Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License (CC BY-NC-ND 4.0). This license allows users to copy and distribute the article, in whole or in part, provided the work is attributed back to the original author and publisher, but does not allow for the distribution of derivative versions or for the article to be used commercially. Permission of the journal is required for the distribution of derivative versions or for commercial uses.

Under either license, all users must credit the authors and journal for the original creation of the work.” ”

References

1 Johnson, M. H. From mouse egg to mouse embryo: polarities, axes, and tissues. *Annual review of cell and developmental biology* **25**, 483-512, doi:10.1146/annurev.cellbio.042308.113348 (2009).

2 Hardy, K., Handyside, A. H. & Winston, R. M. The human blastocyst: cell number, death and allocation during late preimplantation development in vitro. *Development* **107**, 597-604 (1989).

3 Aiken, C. E., Swoboda, P. P., Skepper, J. N. & Johnson, M. H. The direct measurement of embryogenic volume and nucleo-cytoplasmic ratio during mouse pre-implantation development. *Reproduction* **128**, 527-535, doi:10.1530/rep.1.00281 (2004).

4 Kalkan, T. & Smith, A. Mapping the route from naive pluripotency to lineage specification. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences* **369**, doi:10.1098/rstb.2013.0540 (2014).

5 Martello, G. & Smith, A. The nature of embryonic stem cells. *Annual review of cell and developmental biology* **30**, 647-675, doi:10.1146/annurev-cellbio-100913-013116 (2014).

6 Capecchi, M. R. Gene targeting in mice: functional analysis of the mammalian genome for the twenty-first century. *Nature reviews. Genetics* **6**, 507-512, doi:10.1038/nrg1619 (2005).

7 Nichols, J. & Smith, A. Naive and primed pluripotent states. *Cell stem cell* **4**, 487-492, doi:10.1016/j.stem.2009.05.015 (2009).

8 Liu, P. *et al.* Requirement for Wnt3 in vertebrate axis formation. *Nature genetics* **22**, 361-365, doi:10.1038/11932 (1999).

9 Winnier, G., Blessing, M., Labosky, P. A. & Hogan, B. L. Bone morphogenetic protein-4 is required for mesoderm formation and patterning in the mouse. *Genes & development* **9**, 2105-2116 (1995).

10 Conlon, F. L. *et al.* A primary requirement for nodal in the formation and maintenance of the primitive streak in the mouse. *Development* **120**, 1919-1928 (1994).

11 Clements, W. K. & Traver, D. Signalling pathways that control vertebrate haematopoietic stem cell specification. *Nature reviews. Immunology* **13**, 336-348, doi:10.1038/nri3443 (2013).

12 Yi, F. *et al.* Opposing effects of Tcf3 and Tcf1 control Wnt stimulation of embryonic stem cell self-renewal. *Nature cell biology* **13**, 762-770, doi:10.1038/ncb2283 (2011).

13 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. *Nature cell biology* **13**, 838-845, doi:10.1038/ncb2267 (2011).

14 Ying, Q. L. *et al.* The ground state of embryonic stem cell self-renewal. *Nature* **453**, 519-523, doi:10.1038/nature06968 (2008).

15 Ogawa, K., Nishinakamura, R., Iwamatsu, Y., Shimosato, D. & Niwa, H. Synergistic action of Wnt and LIF in maintaining pluripotency of mouse ES cells. *Biochemical and biophysical research communications* **343**, 159-166, doi:10.1016/j.bbrc.2006.02.127 (2006).

16 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. *Nature medicine* **10**, 55-63, doi:10.1038/nm979 (2004).

17 Nostro, M. C., Cheng, X., Keller, G. M. & Gadue, P. Wnt, activin, and BMP signaling regulate distinct stages in the developmental pathway from embryonic stem cells to blood. *Cell stem cell* **2**, 60-71, doi:10.1016/j.stem.2007.10.011 (2008).

18 Kurek, D. *et al.* Endogenous WNT signals mediate BMP-induced and spontaneous differentiation of epiblast stem cells and human embryonic stem cells. *Stem cell reports* **4**, 114-128, doi:10.1016/j.stemcr.2014.11.007 (2015).

19 Lindsley, R. C., Gill, J. G., Kyba, M., Murphy, T. L. & Murphy, K. M. Canonical Wnt signaling is required for development of embryonic stem cell-derived mesoderm. *Development* **133**, 3787-3796, doi:10.1242/dev.02551 (2006).

20 Sumi, T., Oki, S., Kitajima, K. & Meno, C. Epiblast ground state is controlled by canonical Wnt/beta-catenin signaling in the postimplantation mouse embryo and epiblast stem cells. *PloS one* **8**, e63378, doi:10.1371/journal.pone.0063378 (2013).

21 Martello, G., Bertone, P. & Smith, A. Identification of the missing pluripotency mediator downstream of leukaemia inhibitory factor. *The EMBO journal* **32**, 2561-2574, doi:10.1038/emboj.2013.177 (2013).

22 Niwa, H., Burdon, T., Chambers, I. & Smith, A. Self-renewal of pluripotent embryonic stem cells is mediated via activation of STAT3. *Genes & development* **12**, 2048-2060 (1998).

23 Matsuda, T. *et al.* STAT3 activation is sufficient to maintain an undifferentiated state of mouse embryonic stem cells. *The EMBO journal* **18**, 4261-4269, doi:10.1093/emboj/18.15.4261 (1999).

24 Boeuf, H., Hauss, C., Graeve, F. D., Baran, N. & Kedinger, C. Leukemia inhibitory factor-dependent transcriptional activation in embryonic stem cells. *The Journal of cell biology* **138**, 1207-1217 (1997).

25 Smith, A. G. *et al.* Inhibition of pluripotential embryonic stem cell differentiation by purified polypeptides. *Nature* **336**, 688-690, doi:10.1038/336688a0 (1988).

26 Smith, A. G. & Hooper, M. L. Buffalo rat liver cells produce a diffusible activity which inhibits the differentiation of murine embryonal carcinoma and embryonic stem cells. *Developmental biology* **121**, 1-9 (1987).

27 Williams, R. L. *et al.* Myeloid leukaemia inhibitory factor maintains the developmental potential of embryonic stem cells. *Nature* **336**, 684-687, doi:10.1038/336684a0 (1988).

28 Kunath, T. *et al.* FGF stimulation of the Erk1/2 signalling cascade triggers transition of pluripotent embryonic stem cells from self-renewal to lineage commitment. *Development* **134**, 2895-2902, doi:10.1242/dev.02880 (2007).

29 ten Berge, D. *et al.* Embryonic stem cells require Wnt proteins to prevent differentiation to epiblast stem cells. *Nature cell biology* **13**, 1070-1075, doi:10.1038/ncb2314 (2011).

30 Martin, G. R. Teratocarcinomas and mammalian embryogenesis. *Science* **209**, 768-776 (1980).

31 Wilder, P. J. *et al.* Inactivation of the FGF-4 gene in embryonic stem cells alters the growth and/or the survival of their early differentiated progeny. *Developmental biology* **192**, 614-629, doi:10.1006/dbio.1997.8777 (1997).

32 Chai, N. *et al.* FGF is an essential regulator of the fifth cell division in preimplantation mouse embryos. *Developmental biology* **198**, 105-115 (1998).

33 Feldman, B., Poueymirou, W., Papaioannou, V. E., DeChiara, T. M. & Goldfarb, M. Requirement of FGF-4 for postimplantation mouse development. *Science* **267**, 246-249 (1995).

34 Kemp, C., Willems, E., Abdo, S., Lambiv, L. & Leyns, L. Expression of all Wnt genes and their secreted antagonists during mouse blastocyst and postimplantation development. *Developmental dynamics : an official publication of the American Association of Anatomists* **233**, 1064-1075, doi:10.1002/dvdy.20408 (2005).

35 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. *Developmental cell* **12**, 957-971, doi:10.1016/j.devcel.2007.04.001 (2007).

36 Takahashi, K. & Yamanaka, S. Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. *Cell* **126**, 663-676, doi:10.1016/j.cell.2006.07.024 (2006).

37 Martello, G. *et al.* Esrrb is a pivotal target of the Gsk3/Tcf3 axis regulating embryonic stem cell self-renewal. *Cell stem cell* **11**, 491-504, doi:10.1016/j.stem.2012.06.008 (2012).

38 Chambers, I. & Tomlinson, S. R. The transcriptional foundation of pluripotency. *Development* **136**, 2311-2322, doi:10.1242/dev.024398 (2009).

39 Silva, J. *et al.* Nanog is the gateway to the pluripotent ground state. *Cell* **138**, 722-737, doi:10.1016/j.cell.2009.07.039 (2009).

40 Silva, J., Chambers, I., Pollard, S. & Smith, A. Nanog promotes transfer of pluripotency after cell fusion. *Nature* **441**, 997-1001, doi:10.1038/nature04914 (2006).

41 Masui, S. *et al.* Pluripotency governed by Sox2 via regulation of Oct3/4 expression in mouse embryonic stem cells. *Nature cell biology* **9**, 625-635, doi:10.1038/ncb1589 (2007).

42 Niwa, H., Miyazaki, J. & Smith, A. G. Quantitative expression of Oct-3/4 defines differentiation, dedifferentiation or self-renewal of ES cells. *Nature genetics* **24**, 372-376, doi:10.1038/74199 (2000).

43 Chen, X. *et al.* Integration of external signaling pathways with the core transcriptional network in embryonic stem cells. *Cell* **133**, 1106-1117, doi:10.1016/j.cell.2008.04.043 (2008).

44 Marson, A. *et al.* Connecting microRNA genes to the core transcriptional regulatory circuitry of embryonic stem cells. *Cell* **134**, 521-533, doi:10.1016/j.cell.2008.07.020 (2008).

45 Young, R. A. Control of the embryonic stem cell state. *Cell* **144**, 940-954, doi:10.1016/j.cell.2011.01.032 (2011).

46 Cole, M. F., Johnstone, S. E., Newman, J. J., Kagey, M. H. & Young, R. A. Tcf3 is an integral component of the core regulatory circuitry of embryonic stem cells. *Genes & development* **22**, 746-755, doi:10.1101/gad.1642408 (2008).

47 Whyte, W. A. *et al.* Master transcription factors and mediator establish super-enhancers at key cell identity genes. *Cell* **153**, 307-319, doi:10.1016/j.cell.2013.03.035 (2013).

48 Hnisz, D. *et al.* Convergence of developmental and oncogenic signaling pathways at transcriptional super-enhancers. *Molecular cell* **58**, 362-370, doi:10.1016/j.molcel.2015.02.014 (2015).

49 Hnisz, D. *et al.* Super-enhancers in the control of cell identity and disease. *Cell* **155**, 934-947, doi:10.1016/j.cell.2013.09.053 (2013).

50 Clevers, H. & Nusse, R. Wnt/beta-catenin signaling and disease. *Cell* **149**, 1192-1205, doi:10.1016/j.cell.2012.05.012 (2012).

51 King, N. *et al.* The genome of the choanoflagellate Monosiga brevicollis and the origin of metazoans. *Nature* **451**, 783-788, doi:10.1038/nature06617 (2008).

52 Nusslein-Volhard, C. & Wieschaus, E. Mutations affecting segment number and polarity in Drosophila. *Nature* **287**, 795-801 (1980).

53 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).

54 Andl, T., Reddy, S. T., Gaddapara, T. & Millar, S. E. WNT signals are required for the initiation of hair follicle development. *Developmental cell* **2**, 643-653 (2002).

55 Huelsken, J., Vogel, R., Erdmann, B., Cotsarelis, G. & Birchmeier, W. beta-Catenin controls hair follicle morphogenesis and stem cell differentiation in the skin. *Cell* **105**, 533-545 (2001).

56 Choi, Y. S. *et al.* Distinct functions for Wnt/beta-catenin in hair follicle stem cell proliferation and survival and interfollicular epidermal homeostasis. *Cell stem cell* **13**, 720-733, doi:10.1016/j.stem.2013.10.003 (2013).

57 Clevers, H., Loh, K. M. & Nusse, R. Stem cell signaling. An integral program for tissue renewal and regeneration: Wnt signaling and stem cell control. *Science* **346**, 1248012, doi:10.1126/science.1248012 (2014).

58 Polakis, P. Wnt signaling in cancer. *Cold Spring Harbor perspectives in biology* **4**, doi:10.1101/cshperspect.a008052 (2012).

59 Pokutta, S. & Weis, W. I. Structure of the dimerization and beta-catenin-binding region of alpha-catenin. *Molecular cell* **5**, 533-543 (2000).

60 Huber, A. H. & Weis, W. I. The structure of the beta-catenin/E-cadherin complex and the molecular basis of diverse ligand recognition by beta-catenin. *Cell* **105**, 391-402 (2001).

61 Takada, R. *et al.* Monounsaturated fatty acid modification of Wnt protein: its role in Wnt secretion. *Developmental cell* **11**, 791-801, doi:10.1016/j.devcel.2006.10.003 (2006).

62 Biechele, S., Cox, B. J. & Rossant, J. Porcupine homolog is required for canonical Wnt signaling and gastrulation in mouse embryos. *Developmental biology* **355**, 275-285, doi:10.1016/j.ydbio.2011.04.029 (2011).

63 Mulligan, K. A. *et al.* Secreted Wingless-interacting molecule (Swim) promotes long-range signaling by maintaining Wingless solubility. *Proceedings of the National Academy of Sciences of the United States of America* **109**, 370-377, doi:10.1073/pnas.1119197109 (2012).

64 Bhanot, P. *et al.* A new member of the frizzled family from Drosophila functions as a Wingless receptor. *Nature* **382**, 225-230, doi:10.1038/382225a0 (1996).

65 Tamai, K. *et al.* LDL-receptor-related proteins in Wnt signal transduction. *Nature* **407**, 530-535, doi:10.1038/35035117 (2000).

66 Pinson, K. I., Brennan, J., Monkley, S., Avery, B. J. & Skarnes, W. C. An LDL-receptor-related protein mediates Wnt signalling in mice. *Nature* **407**, 535-538, doi:10.1038/35035124 (2000).

67 Janda, C. Y., Waghray, D., Levin, A. M., Thomas, C. & Garcia, K. C. Structural basis of Wnt recognition by Frizzled. *Science* **337**, 59-64, doi:10.1126/science.1222879 (2012).

68 Li, V. S. *et al.* Wnt signaling through inhibition of beta-catenin degradation in an intact Axin1 complex. *Cell* **149**, 1245-1256, doi:10.1016/j.cell.2012.05.002 (2012).

69 Stamos, J. L., Chu, M. L., Enos, M. D., Shah, N. & Weis, W. I. Structural basis of GSK-3 inhibition by N-terminal phosphorylation and by the Wnt receptor LRP6. *eLife* **3**, e01998, doi:10.7554/eLife.01998 (2014).

70 Zhang, N. *et al.* FoxM1 promotes beta-catenin nuclear localization and controls Wnt target-gene expression and glioma tumorigenesis. *Cancer cell* **20**, 427-442, doi:10.1016/j.ccr.2011.08.016 (2011).

71 Xing, Y. *et al.* Crystal structure of a full-length beta-catenin. *Structure* **16**, 478-487, doi:10.1016/j.str.2007.12.021 (2008).

72 Cadigan, K. M. & Waterman, M. L. TCF/LEFs and Wnt signaling in the nucleus. *Cold Spring Harbor perspectives in biology* **4**, doi:10.1101/cshperspect.a007906 (2012).

73 Arce, L., Yokoyama, N. N. & Waterman, M. L. Diversity of LEF/TCF action in development and disease. *Oncogene* **25**, 7492-7504, doi:10.1038/sj.onc.1210056 (2006).

74 Waterman, M. L., Fischer, W. H. & Jones, K. A. A thymus-specific member of the HMG protein family regulates the human T cell receptor C alpha enhancer. *Genes & development* **5**, 656-669 (1991).

75 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 & development* **5**, 880-894 (1991).

76 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. *The EMBO journal* **10**, 123-132 (1991).

77 Korinek, V. *et al.* Two members of the Tcf family implicated in Wnt/beta-catenin signaling during embryogenesis in the mouse. *Molecular and cellular biology* **18**, 1248-1256 (1998).

78 Graham, T. A., Weaver, C., Mao, F., Kimelman, D. & Xu, W. Crystal structure of a beta-catenin/Tcf complex. *Cell* **103**, 885-896 (2000).
79 Molenaar, M. *et al.* XTcf-3 transcription factor mediates beta-catenin-induced axis formation in Xenopus embryos. *Cell* **86**, 391-399 (1996).
80 Behrens, J. *et al.* Functional interaction of beta-catenin with the transcription factor LEF-1. *Nature* **382**, 638-642, doi:10.1038/382638a0 (1996).
81 van Beest, M. *et al.* Sequence-specific high mobility group box factors recognize 10-12-base pair minor groove motifs. *The Journal of biological chemistry* **275**, 27266-27273, doi:10.1074/jbc.M004102200 (2000).
82 Love, J. J. *et al.* Structural basis for DNA bending by the architectural transcription factor LEF-1. *Nature* **376**, 791-795, doi:10.1038/376791a0 (1995).
83 van de Wetering, M. & Clevers, H. Sequence-specific interaction of the HMG box proteins TCF-1 and SRY occurs within the minor groove of a Watson-Crick double helix. *The EMBO journal* **11**, 3039-3044 (1992).
84 Giese, K., Amsterdam, A. & Grosschedl, R. DNA-binding properties of the HMG domain of the lymphoid-specific transcriptional regulator LEF-1. *Genes & development* **5**, 2567-2578 (1991).
85 Galceran, J., Farinas, I., Depew, M. J., Clevers, H. & Grosschedl, R. Wnt3a-/--like phenotype and limb deficiency in Lef1(-/-)Tcf1(-/-) mice. *Genes & development* **13**, 709-717 (1999).
86 Roose, J. *et al.* The Xenopus Wnt effector XTcf-3 interacts with Groucho-related transcriptional repressors. *Nature* **395**, 608-612, doi:10.1038/26989 (1998).
87 Cavallo, R. A. *et al.* Drosophila Tcf and Groucho interact to repress Wingless signalling activity. *Nature* **395**, 604-608, doi:10.1038/26982 (1998).
88 Hikasa, H. *et al.* Regulation of TCF3 by Wnt-dependent phosphorylation during vertebrate axis specification. *Dev Cell* **19**, 521-532, doi:S1534-5807(10)00423-5 [pii]10.1016/j.devcel.2010.09.005 [doi] (2010).
89 Merrill, B. J. *et al.* Tcf3: a transcriptional regulator of axis induction in the early embryo. *Development* **131**, 263-274, doi:10.1242/dev.00935 (2004).
90 Verbeek, S. *et al.* An HMG-box-containing T-cell factor required for thymocyte differentiation. *Nature* **374**, 70-74, doi:10.1038/374070a0 (1995).
91 van Genderen, C. *et al.* Development of several organs that require inductive epithelial-mesenchymal interactions is impaired in LEF-1-deficient mice. *Genes & development* **8**, 2691-2703 (1994).
92 Oosterwegel, M. *et al.* Differential expression of the HMG box factors TCF-1 and LEF-1 during murine embryogenesis. *Development* **118**, 439-448 (1993).
93 Pereira, L., Yi, F. & Merrill, B. J. Repression of Nanog gene transcription by Tcf3 limits embryonic stem cell self-renewal. *Molecular and cellular biology* **26**, 7479-7491, doi:10.1128/MCB.00368-06 (2006).
94 Yang, S. H., Kalkan, T., Morrisroe, C., Smith, A. & Sharrocks, A. D. A genome-wide RNAi screen reveals MAP kinase phosphatases as key ERK pathway regulators during embryonic stem cell differentiation. *PLoS genetics* **8**, e1003112, doi:10.1371/journal.pgen.1003112 (2012).
95 Leeb, M., Dietmann, S., Paramor, M., Niwa, H. & Smith, A. Genetic exploration of the exit from self-renewal using haploid embryonic stem cells. *Cell stem cell* **14**, 385-393, doi:10.1016/j.stem.2013.12.008 (2014).
96 Betschinger, J. *et al.* Exit from pluripotency is gated by intracellular redistribution of the bHLH transcription factor Tfe3. *Cell* **153**, 335-347, doi:10.1016/j.cell.2013.03.012 (2013).
97 Kelly, K. F. *et al.* beta-catenin enhances Oct-4 activity and reinforces pluripotency through a TCF-independent mechanism. *Cell stem cell* **8**, 214-227, doi:10.1016/j.stem.2010.12.010 (2011).

98 Abu-Remaileh, M. *et al.* Oct-3/4 regulates stem cell identity and cell fate decisions by modulating Wnt/beta-catenin signalling. *The EMBO journal* **29**, 3236-3248, doi:10.1038/emboj.2010.200 (2010).

99 Takao, Y., Yokota, T. & Koide, H. Beta-catenin up-regulates Nanog expression through interaction with Oct-3/4 in embryonic stem cells. *Biochemical and biophysical research communications* **353**, 699-705, doi:10.1016/j.bbrc.2006.12.072 (2007).

100 Hoffmeyer, K. *et al.* Wnt/beta-catenin signaling regulates telomerase in stem cells and cancer cells. *Science* **336**, 1549-1554, doi:10.1126/science.1218370 (2012).

101 Cooke, M. S., Evans, M. D., Dizdaroglu, M. & Lunec, J. Oxidative DNA damage: mechanisms, mutation, and disease. *FASEB journal : official publication of the Federation of American Societies for Experimental Biology* **17**, 1195-1214, doi:10.1096/fj.02-0752rev (2003).

102 Marteijn, J. A., Lans, H., Vermeulen, W. & Hoeijmakers, J. H. Understanding nucleotide excision repair and its roles in cancer and ageing. *Nature reviews. Molecular cell biology* **15**, 465-481, doi:10.1038/nrm3822 (2014).

103 Houtgraaf, J. H., Versmissen, J. & van der Giessen, W. J. A concise review of DNA damage checkpoints and repair in mammalian cells. *Cardiovascular revascularization medicine : including molecular interventions* **7**, 165-172, doi:10.1016/j.carrev.2006.02.002 (2006).

104 Pfeiffer, P., Goedecke, W. & Obe, G. Mechanisms of DNA double-strand break repair and their potential to induce chromosomal aberrations. *Mutagenesis* **15**, 289-302 (2000).

105 Merrill, B. J. & Holm, C. A requirement for recombinational repair in Saccharomyces cerevisiae is caused by DNA replication defects of mec1 mutants. *Genetics* **153**, 595-605 (1999).

106 Jackson, S. P. & Bartek, J. The DNA-damage response in human biology and disease. *Nature* **461**, 1071-1078, doi:10.1038/nature08467 (2009).

107 Johnson, R. D. & Jasin, M. Sister chromatid gene conversion is a prominent double-strand break repair pathway in mammalian cells. *The EMBO journal* **19**, 3398-3407, doi:10.1093/emboj/19.13.3398 (2000).

108 Johnson, R. D., Liu, N. & Jasin, M. Mammalian XRCC2 promotes the repair of DNA double-strand breaks by homologous recombination. *Nature* **401**, 397-399, doi:10.1038/43932 (1999).

109 Chapman, J. R., Taylor, M. R. & Boulton, S. J. Playing the end game: DNA double-strand break repair pathway choice. *Molecular cell* **47**, 497-510, doi:10.1016/j.molcel.2012.07.029 (2012).

110 Jinek, M. *et al.* RNA-programmed genome editing in human cells. *eLife* **2**, e00471, doi:10.7554/eLife.00471 (2013).

111 Mali, P. *et al.* RNA-guided human genome engineering via Cas9. *Science* **339**, 823-826, doi:10.1126/science.1232033 (2013).

112 Ran, F. A. *et al.* Double nicking by RNA-guided CRISPR Cas9 for enhanced genome editing specificity. *Cell* **154**, 1380-1389, doi:10.1016/j.cell.2013.08.021 (2013).

113 Pardo, B., Gomez-Gonzalez, B. & Aguilera, A. DNA repair in mammalian cells: DNA double-strand break repair: how to fix a broken relationship. *Cellular and molecular life sciences : CMLS* **66**, 1039-1056, doi:10.1007/s00018-009-8740-3 (2009).

114 Daley, J. M. & Wilson, T. E. Rejoining of DNA double-strand breaks as a function of overhang length. *Molecular and cellular biology* **25**, 896-906, doi:10.1128/MCB.25.3.896-906.2005 (2005).

115 Pang, D., Yoo, S., Dynan, W. S., Jung, M. & Dritschilo, A. Ku proteins join DNA fragments as shown by atomic force microscopy. *Cancer Res* **57**, 1412-1415 (1997).

116 Walker, J. R., Corpina, R. A. & Goldberg, J. Structure of the Ku heterodimer bound to DNA and its implications for double-strand break repair. *Nature* **412**, 607-614, doi:10.1038/35088000 (2001).

117 Mari, P. O. *et al.* Dynamic assembly of end-joining complexes requires interaction between Ku70/80 and XRCC4. *Proceedings of the National Academy of Sciences of the United States of America* **103**, 18597-18602, doi:10.1073/pnas.0609061103 (2006).

118 Gottlieb, T. M. & Jackson, S. P. The DNA-dependent protein kinase: requirement for DNA ends and association with Ku antigen. *Cell* **72**, 131-142 (1993).

119 Kurimasa, A. *et al.* Requirement for the kinase activity of human DNA-dependent protein kinase catalytic subunit in DNA strand break rejoining. *Molecular and cellular biology* **19**, 3877-3884 (1999).

120 Spagnolo, L., Rivera-Calzada, A., Pearl, L. H. & Llorca, O. Three-dimensional structure of the human DNA-PKcs/Ku70/Ku80 complex assembled on DNA and its implications for DNA DSB repair. *Molecular cell* **22**, 511-519, doi:10.1016/j.molcel.2006.04.013 (2006).

121 Clements, P. M. *et al.* The ataxia-oculomotor apraxia 1 gene product has a role distinct from ATM and interacts with the DNA strand break repair proteins XRCC1 and XRCC4. *DNA repair* **3**, 1493-1502, doi:10.1016/j.dnarep.2004.06.017 (2004).

122 Ahel, I. *et al.* The neurodegenerative disease protein aprataxin resolves abortive DNA ligation intermediates. *Nature* **443**, 713-716, doi:10.1038/nature05164 (2006).

123 Pardo, B., Ma, E. & Marcand, S. Mismatch tolerance by DNA polymerase Pol4 in the course of nonhomologous end joining in Saccharomyces cerevisiae. *Genetics* **172**, 2689-2694, doi:10.1534/genetics.105.053512 (2006).

124 Nick McElhinny, S. A. *et al.* A gradient of template dependence defines distinct biological roles for family X polymerases in nonhomologous end joining. *Molecular cell* **19**, 357-366, doi:10.1016/j.molcel.2005.06.012 (2005).

125 Daley, J. M., Laan, R. L., Suresh, A. & Wilson, T. E. DNA joint dependence of pol X family polymerase action in nonhomologous end joining. *The Journal of biological chemistry* **280**, 29030-29037, doi:10.1074/jbc.M505277200 (2005).

126 Gu, J., Lu, H., Tsai, A. G., Schwarz, K. & Lieber, M. R. Single-stranded DNA ligation and XLF-stimulated incompatible DNA end ligation by the XRCC4-DNA ligase IV complex: influence of terminal DNA sequence. *Nucleic acids research* **35**, 5755-5762, doi:10.1093/nar/gkm579 (2007).

127 Boulton, S. J. & Jackson, S. P. Saccharomyces cerevisiae Ku70 potentiates illegitimate DNA double-strand break repair and serves as a barrier to error-prone DNA repair pathways. *The EMBO journal* **15**, 5093-5103 (1996).

128 Wilson, T. E., Grawunder, U. & Lieber, M. R. Yeast DNA ligase IV mediates non-homologous DNA end joining. *Nature* **388**, 495-498, doi:10.1038/41365 (1997).

129 Lieber, M. R. The mechanism of double-strand DNA break repair by the nonhomologous DNA end-joining pathway. *Annual review of biochemistry* **79**, 181-211, doi:10.1146/annurev.biochem.052308.093131 (2010).

130 Sun, H., Treco, D. & Szostak, J. W. Extensive 3'-overhanging, single-stranded DNA associated with the meiosis-specific double-strand breaks at the ARG4 recombination initiation site. *Cell* **64**, 1155-1161 (1991).

131 Alani, E., Thresher, R., Griffith, J. D. & Kolodner, R. D. Characterization of DNA-binding and strand-exchange stimulation properties of y-RPA, a yeast single-strand-DNA-binding protein. *Journal of molecular biology* **227**, 54-71 (1992).

132 Moreau, S., Morgan, E. A. & Symington, L. S. Overlapping functions of the Saccharomyces cerevisiae Mre11, Exo1 and Rad27 nucleases in DNA metabolism. *Genetics* **159**, 1423-1433 (2001).

133 Trujillo, K. M., Yuan, S. S., Lee, E. Y. & Sung, P. Nuclease activities in a complex of human recombination and DNA repair factors Rad50, Mre11, and p95. *The Journal of biological chemistry* **273**, 21447-21450 (1998).

134 Llorente, B. & Symington, L. S. The Mre11 nuclease is not required for 5' to 3' resection at multiple HO-induced double-strand breaks. *Molecular and cellular biology* **24**, 9682-9694, doi:10.1128/MCB.24.21.9682-9694.2004 (2004).

135 Krejci, L., Altmannova, V., Spirek, M. & Zhao, X. Homologous recombination and its regulation. *Nucleic acids research* **40**, 5795-5818, doi:10.1093/nar/gks270 (2012).
136 Ira, G. & Haber, J. E. Characterization of RAD51-independent break-induced replication that acts preferentially with short homologous sequences. *Molecular and cellular biology* **22**, 6384-6392 (2002).
137 Goetz, J. D., Motycka, T. A., Han, M., Jasin, M. & Tomkinson, A. E. Reduced repair of DNA double-strand breaks by homologous recombination in a DNA ligase I-deficient human cell line. *DNA repair* **4**, 649-654, doi:10.1016/j.dnarep.2005.02.004 (2005).
138 Sartori, A. A. *et al.* Human CtIP promotes DNA end resection. *Nature* **450**, 509-514, doi:10.1038/nature06337 (2007).
139 Buis, J., Stoneham, T., Spehalski, E. & Ferguson, D. O. Mre11 regulates CtIP-dependent double-strand break repair by interaction with CDK2. *Nature structural & molecular biology* **19**, 246-252, doi:10.1038/nsmb.2212 (2012).
140 Chen, L., Nievera, C. J., Lee, A. Y. & Wu, X. Cell cycle-dependent complex formation of BRCA1.CtIP.MRN is important for DNA double-strand break repair. *The Journal of biological chemistry* **283**, 7713-7720, doi:10.1074/jbc.M710245200 (2008).
141 Huertas, P. & Jackson, S. P. Human CtIP mediates cell cycle control of DNA end resection and double strand break repair. *The Journal of biological chemistry* **284**, 9558-9565, doi:10.1074/jbc.M808906200 (2009).
142 Wong, E. A. & Capecchi, M. R. Analysis of homologous recombination in cultured mammalian cells in transient expression and stable transformation assays. *Somatic cell and molecular genetics* **12**, 63-72 (1986).
143 Escribano-Diaz, C. *et al.* A cell cycle-dependent regulatory circuit composed of 53BP1-RIF1 and BRCA1-CtIP controls DNA repair pathway choice. *Molecular cell* **49**, 872-883, doi:10.1016/j.molcel.2013.01.001 (2013).
144 Capecchi, M. R. High efficiency transformation by direct microinjection of DNA into cultured mammalian cells. *Cell* **22**, 479-488 (1980).
145 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**, 230-234 (1985).
146 Thomas, K. R. & Capecchi, M. R. Site-directed mutagenesis by gene targeting in mouse embryo-derived stem cells. *Cell* **51**, 503-512 (1987).
147 Mansour, S. L., Thomas, K. R. & Capecchi, M. R. Disruption of the proto-oncogene int-2 in mouse embryo-derived stem cells: a general strategy for targeting mutations to non-selectable genes. *Nature* **336**, 348-352, doi:10.1038/336348a0 (1988).
148 Sargent, R. G., Brenneman, M. A. & Wilson, J. H. Repair of site-specific double-strand breaks in a mammalian chromosome by homologous and illegitimate recombination. *Molecular and cellular biology* **17**, 267-277 (1997).
149 Vasquez, K. M., Marburger, K., Intody, Z. & Wilson, J. H. Manipulating the mammalian genome by homologous recombination. *Proceedings of the National Academy of Sciences of the United States of America* **98**, 8403-8410, doi:10.1073/pnas.111009698 (2001).
150 Sedivy, J. M. & Sharp, P. A. Positive genetic selection for gene disruption in mammalian cells by homologous recombination. *Proceedings of the National Academy of Sciences of the United States of America* **86**, 227-231 (1989).
151 Austin, C. P. *et al.* The knockout mouse project. *Nature genetics* **36**, 921-924, doi:10.1038/ng0904-921 (2004).

152 Hall, B., Limaye, A. & Kulkarni, A. B. Overview: generation of gene knockout mice. *Current protocols in cell biology / editorial board, Juan S. Bonifacino ... [et al.]* **Chapter 19**, Unit 19 12 19 12 11-17, doi:10.1002/0471143030.cb1912s44 (2009).

153 Hanson, K. D. & Sedivy, J. M. Analysis of biological selections for high-efficiency gene targeting. *Molecular and cellular biology* **15**, 45-51 (1995).

154 Savatier, P., Lapillonne, H., Jirmanova, L., Vitelli, L. & Samarut, J. Analysis of the cell cycle in mouse embryonic stem cells. *Methods in molecular biology* **185**, 27-33 (2002).

155 Choulika, A., Perrin, A., Dujon, B. & Nicolas, J. F. Induction of homologous recombination in mammalian chromosomes by using the I-SceI system of Saccharomyces cerevisiae. *Molecular and cellular biology* **15**, 1968-1973 (1995).

156 Plessis, A., Perrin, A., Haber, J. E. & Dujon, B. Site-specific recombination determined by I-SceI, a mitochondrial group I intron-encoded endonuclease expressed in the yeast nucleus. *Genetics* **130**, 451-460 (1992).

157 Rouet, P., Smih, F. & Jasin, M. Introduction of double-strand breaks into the genome of mouse cells by expression of a rare-cutting endonuclease. *Molecular and cellular biology* **14**, 8096-8106 (1994).

158 Rudin, N., Sugarman, E. & Haber, J. E. Genetic and physical analysis of double-strand break repair and recombination in Saccharomyces cerevisiae. *Genetics* **122**, 519-534 (1989).

159 Bibikova, M., Beumer, K., Trautman, J. K. & Carroll, D. Enhancing gene targeting with designed zinc finger nucleases. *Science* **300**, 764, doi:10.1126/science.1079512 (2003).

160 Porteus, M. H. Mammalian gene targeting with designed zinc finger nucleases. *Molecular therapy : the journal of the American Society of Gene Therapy* **13**, 438-446, doi:10.1016/j.ymthe.2005.08.003 (2006).

161 Porteus, M. H. & Baltimore, D. Chimeric nucleases stimulate gene targeting in human cells. *Science* **300**, 763, doi:10.1126/science.1078395 (2003).

162 Christian, M. *et al.* Targeting DNA double-strand breaks with TAL effector nucleases. *Genetics* **186**, 757-761, doi:10.1534/genetics.110.120717 (2010).

163 Yusa, K. *et al.* Targeted gene correction of alpha1-antitrypsin deficiency in induced pluripotent stem cells. *Nature* **478**, 391-394, doi:10.1038/nature10424 (2011).

164 Doudna, J. A. & Charpentier, E. Genome editing. The new frontier of genome engineering with CRISPR-Cas9. *Science* **346**, 1258096, doi:10.1126/science.1258096 (2014).

165 Ishino, Y., Shinagawa, H., Makino, K., Amemura, M. & Nakata, A. Nucleotide sequence of the iap gene, responsible for alkaline phosphatase isozyme conversion in Escherichia coli, and identification of the gene product. *Journal of bacteriology* **169**, 5429-5433 (1987).

166 Barrangou, R. *et al.* CRISPR provides acquired resistance against viruses in prokaryotes. *Science* **315**, 1709-1712, doi:10.1126/science.1138140 (2007).

167 Jinek, M. *et al.* A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science* **337**, 816-821, doi:10.1126/science.1225829 (2012).

168 Cong, L. *et al.* Multiplex genome engineering using CRISPR/Cas systems. *Science* **339**, 819-823, doi:10.1126/science.1231143 (2013).

169 Lin, S., Staahl, B. T., Alla, R. K. & Doudna, J. A. Enhanced homology-directed human genome engineering by controlled timing of CRISPR/Cas9 delivery. *eLife* **3**, e04766, doi:10.7554/eLife.04766 (2014).

170 Fu, Y., Sander, J. D., Reyon, D., Cascio, V. M. & Joung, J. K. Improving CRISPR-Cas nuclease specificity using truncated guide RNAs. *Nature biotechnology* **32**, 279-284, doi:10.1038/nbt.2808 (2014).

171 Tsai, S. Q. *et al.* GUIDE-seq enables genome-wide profiling of off-target cleavage by CRISPR-Cas nucleases. *Nature biotechnology* **33**, 187-197, doi:10.1038/nbt.3117 (2015).

172 Yagi, T. *et al.* Homologous recombination at c-fyn locus of mouse embryonic stem cells with use of diphtheria toxin A-fragment gene in negative selection. *Proceedings of the National Academy of Sciences of the United States of America* **87**, 9918-9922 (1990).
173 Huelsken, J. *et al.* Requirement for beta-catenin in anterior-posterior axis formation in mice. *The Journal of cell biology* **148**, 567-578 (2000).
174 Stewart, C. L. *et al.* Blastocyst implantation depends on maternal expression of leukaemia inhibitory factor. *Nature* **359**, 76-79, doi:10.1038/359076a0 (1992).
175 Hondo, E. & Stewart, C. L. Profiling gene expression in growth-arrested mouse embryos in diapause. *Genome biology* **6**, 202, doi:10.1186/gb-2004-6-1-202 (2005).
176 Boroviak, T. & Nichols, J. The birth of embryonic pluripotency. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences* **369**, doi:10.1098/rstb.2013.0541 (2014).
177 Nichols, J., Chambers, I., Taga, T. & Smith, A. Physiological rationale for responsiveness of mouse embryonic stem cells to gp130 cytokines. *Development* **128**, 2333-2339 (2001).
178 Buecker, C. *et al.* Reorganization of enhancer patterns in transition from naive to primed pluripotency. *Cell stem cell* **14**, 838-853, doi:10.1016/j.stem.2014.04.003 (2014).
179 Hayashi, K., Ohta, H., Kurimoto, K., Aramaki, S. & Saitou, M. Reconstitution of the mouse germ cell specification pathway in culture by pluripotent stem cells. *Cell* **146**, 519-532, doi:10.1016/j.cell.2011.06.052 (2011).
180 Tesar, P. J. *et al.* New cell lines from mouse epiblast share defining features with human embryonic stem cells. *Nature* **448**, 196-199, doi:10.1038/nature05972 (2007).
181 Hoffman, J. A., Wu, C. I. & Merrill, B. J. Tcf7l1 prepares epiblast cells in the gastrulating mouse embryo for lineage specification. *Development* **140**, 1665-1675, doi:10.1242/dev.087387 (2013).
182 Hsu, P. D. *et al.* DNA targeting specificity of RNA-guided Cas9 nucleases. *Nature biotechnology* **31**, 827-832, doi:10.1038/nbt.2647 (2013).
183 Chang, Y. F., Imam, J. S. & Wilkinson, M. F. The nonsense-mediated decay RNA surveillance pathway. *Annual review of biochemistry* **76**, 51-74, doi:10.1146/annurev.biochem.76.050106.093909 (2007).
184 Lyashenko, N. *et al.* Differential requirement for the dual functions of beta-catenin in embryonic stem cell self-renewal and germ layer formation. *Nature cell biology* **13**, 753-761, doi:10.1038/ncb2260 (2011).
185 Yang, S. H. *et al.* Otx2 and Oct4 drive early enhancer activation during embryonic stem cell transition from naive pluripotency. *Cell reports* **7**, 1968-1981, doi:10.1016/j.celrep.2014.05.037 (2014).
186 Yamaguchi, T. P., Takada, S., Yoshikawa, Y., Wu, N. & McMahon, A. P. T (Brachyury) is a direct target of Wnt3a during paraxial mesoderm specification. *Genes & development* **13**, 3185-3190 (1999).
187 Nusse, R. Wnt signaling. *Cold Spring Harbor perspectives in biology* **4**, doi:4/5/a011163 [pii]10.1101/cshperspect.a011163 [doi] (2012).
188 Stamos, J. L. & Weis, W. I. The beta-catenin destruction complex. *Cold Spring Harbor perspectives in biology* **5**, a007898, doi:cshperspect.a007898 [pii]10.1101/cshperspect.a007898 [doi] (2013).
189 Yost, C. *et al.* The axis-inducing activity, stability, and subcellular distribution of beta-catenin is regulated in Xenopus embryos by glycogen synthase kinase 3. *Genes & development* **10**, 1443-1454 (1996).
190 Aberle, H., Bauer, A., Stappert, J., Kispert, A. & Kemler, R. beta-catenin is a target for the ubiquitin-proteasome pathway. *The EMBO journal* **16**, 3797-3804, doi:10.1093/emboj/16.13.3797 [doi] (1997).

191 Hart, M. *et al.* The F-box protein beta-TrCP associates with phosphorylated beta-catenin and regulates its activity in the cell. *Curr Biol* **9**, 207-210 (1999).
192 Liu, C. *et al.* Control of beta-catenin phosphorylation/degradation by a dual-kinase mechanism. *Cell* **108**, 837-847 (2002).
193 Daniels, D. L. & Weis, W. I. Beta-catenin directly displaces Groucho/TLE repressors from Tcf/Lef in Wnt-mediated transcription activation. *Nature structural & molecular biology* **12**, 364-371 (2005).
194 Barker, N. *et al.* The chromatin remodelling factor Brg-1 interacts with beta-catenin to promote target gene activation. *The EMBO journal* **20**, 4935-4943, doi:10.1093/emboj/20.17.4935 [doi] (2001).
195 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 & development* **11**, 2359-2370 (1997).
196 Molenaar, M. *et al.* XTcf-3 transcription factor mediates beta-catenin-induced axis formation in Xenopus embryos. *Cell* **86**, 391-399. (1996).
197 van de Wetering, M. *et al.* Armadillo coactivates transcription driven by the product of the Drosophila segment polarity gene dTCF. *Cell* **88**, 789-799 (1997).
198 Merrill, B. J., Gat, U., DasGupta, R. & Fuchs, E. Tcf3 and Lef1 regulate lineage differentiation of multipotent stem cells in skin. *Genes & development* **15**, 1688-1705 (2001).
199 Wu, C. I. *et al.* Function of Wnt/β-catenin in counteracting Tcf3-repression through the Tcf3-β-catenin interaction. *Development* **139**, 2118-2129 (2012).
200 Guo, G., Huang, Y., Humphreys, P., Wang, X. & Smith, A. A PiggyBac-Based Recessive Screening Method to Identify Pluripotency Regulators. *PloS one* **6**, e18189, doi:10.1371/journal.pone.0018189 [doi] (2011).
201 Ben-Porath, I. *et al.* An embryonic stem cell-like gene expression signature in poorly differentiated aggressive human tumors. *Nature genetics* **40**, 499-507, doi:10.1038/ng.127 (2008).
202 DasGupta, R. & Fuchs, E. Multiple roles for activated LEF/TCF transcription complexes during hair follicle development and differentiation. *Development* **126**, 4557-4568 (1999).
203 Ivanova, N. B. *et al.* A stem cell molecular signature. *Science* **298**, 601-604 (2002).
204 Tumbar, T. *et al.* Defining the epithelial stem cell niche in skin. *Science* **303**, 359-363, doi:10.1126/science.1092436 [doi]1092436 [pii] (2004).
205 Nusse, R., van Ooyen, A., Cox, D., Fung, Y. K. & Varmus, H. Mode of proviral activation of a putative mammary oncogene (int-1) on mouse chromosome 15. *Nature* **307**, 131-136 (1984).
206 Alexander, C. M., Goel, S., Fakhraldeen, S. A. & Kim, S. Wnt Signaling in Mammary Glands: Plastic Cell Fates and Combinatorial Signaling. *Cold Spring Harbor perspectives in biology*, doi:cshperspect.a008037 [pii]10.1101/cshperspect.a008037 [doi] (2012).
207 Slyper, M. *et al.* Control of breast cancer growth and initiation by the stem cell-associated transcription factor TCF3. *Cancer Res* **72**, 5613-5624, doi:0008-5472.CAN-12-0119 [pii]10.1158/0008-5472.CAN-12-0119 [doi] (2012).
208 Khramtsov, A. I. *et al.* Wnt/beta-catenin pathway activation is enriched in basal-like breast cancers and predicts poor outcome. *Am J Pathol* **176**, 2911-2920, doi:10.2353/ajpath.2010.091125 (2010).
209 Geyer, F. C. *et al.* beta-Catenin pathway activation in breast cancer is associated with triple-negative phenotype but not with CTNNB1 mutation. *Mod Pathol* **24**, 209-231, doi:10.1038/modpathol.2010.205 (2011).

210 Lopez-Knowles, E. *et al.* Cytoplasmic localization of beta-catenin is a marker of poor outcome in breast cancer patients. *Cancer Epidemiol Biomarkers Prev* **19**, 301-309, doi:19/1/301 [pii]10.1158/1055-9965.EPI-09-0741 [doi] (2010).

211 Salomonis, N. *et al.* Alternative splicing regulates mouse embryonic stem cell pluripotency and differentiation. *Proceedings of the National Academy of Sciences of the United States of America* **107**, 10514-10519, doi:0912260107 [pii]10.1073/pnas.0912260107 [doi] (2010).

212 Hikasa, H. & Sokol, S. Y. Phosphorylation of TCF proteins by homeodomain-interacting protein kinase 2. *The Journal of biological chemistry* **286**, 12093-12100, doi:10.1074/jbc.M110.185280 (2011).

213 Ishitani, T., Ninomiya-Tsuji, J. & Matsumoto, K. Regulation of lymphoid enhancer factor 1/T-cell factor by mitogen-activated protein kinase-related Nemo-like kinase-dependent phosphorylation in Wnt/beta-catenin signaling. *Molecular and cellular biology* **23**, 1379-1389 (2003).

214 Ishitani, T. *et al.* The TAK1-NLK-MAPK-related pathway antagonizes signalling between beta-catenin and transcription factor TCF. *Nature* **399**, 798-802, doi:10.1038/21674 (1999).

215 Thorpe, C. J. & Moon, R. T. nemo-like kinase is an essential co-activator of Wnt signaling during early zebrafish development. *Development* **131**, 2899-2909, doi:10.1242/dev.01171 (2004).

216 Spike, B. T. *et al.* A mammary stem cell population identified and characterized in late embryogenesis reveals similarities to human breast cancer. *Cell stem cell* **10**, 183-197, doi:S1934-5909(12)00002-1 [pii]10.1016/j.stem.2011.12.018 [doi] (2012).

217 Mizuno, H., Spike, B. T., Wahl, G. M. & Levine, A. J. Inactivation of p53 in breast cancers correlates with stem cell transcriptional signatures. *Proceedings of the National Academy of Sciences of the United States of America* **107**, 22745-22750, doi:1017001108 [pii]10.1073/pnas.1017001108 [doi] (2010).

218 Wu, Z. Q. *et al.* Canonical Wnt signaling regulates Slug activity and links epithelial-mesenchymal transition with epigenetic Breast Cancer 1, Early Onset (BRCA1) repression. *Proceedings of the National Academy of Sciences of the United States of America* **109**, 16654-16659, doi:1205822109 [pii]10.1073/pnas.1205822109 [doi] (2012).

219 Tutter, A. V., Fryer, C. J. & Jones, K. A. Chromatin-specific regulation of LEF-1-beta-catenin transcription activation and inhibition in vitro. *Genes & development* **15**, 3342-3354, doi:10.1101/gad.946501 [doi] (2001).

220 Reiner, A. *et al.* Immunocytochemical localization of estrogen and progesterone receptor and prognosis in human primary breast cancer. *Cancer Res* **50**, 7057-7061 (1990).

221 Joung, J. K. *et al.* The CRISPR/Cas bacterial immune system cleaves bacteriophage and plasmid DNA. *Nature biotechnology* **468**, 67-71, doi:10.1038/nbt.311710.1038/nature09523 (2010).

222 Ye, L. *et al.* Seamless modification of wild-type induced pluripotent stem cells to the natural CCR5Delta32 mutation confers resistance to HIV infection. *Proceedings of the National Academy of Sciences of the United States of America* **111**, 9591-9596, doi:10.1073/pnas.1407473111 (2014).

223 Ousterout, D. G. *et al.* Multiplex CRISPR/Cas9-based genome editing for correction of dystrophin mutations that cause Duchenne muscular dystrophy. *Nature communications* **6**, 6244, doi:10.1038/ncomms7244 (2015).

224 Long, C. *et al.* Prevention of muscular dystrophy in mice by CRISPR/Cas9-mediated editing of germline DNA. *Science* **345**, 1184-1188, doi:10.1126/science.1254445 (2014).

225 Xie, F. *et al.* Seamless gene correction of beta-thalassemia mutations in patient-specific iPSCs using CRISPR/Cas9 and piggyBac. *Genome research* **24**, 1526-1533, doi:10.1101/gr.173427.114 (2014).
226 Xue, W. *et al.* CRISPR-mediated direct mutation of cancer genes in the mouse liver. *Nature* **514**, 380-384, doi:10.1038/nature13589 (2014).
227 Platt, R. J. *et al.* CRISPR-Cas9 knockin mice for genome editing and cancer modeling. *Cell* **159**, 440-455, doi:10.1016/j.cell.2014.09.014 (2014).
228 Wang, H. *et al.* One-step generation of mice carrying mutations in multiple genes by CRISPR/Cas-mediated genome engineering. *Cell* **153**, 910-918, doi:10.1016/j.cell.2013.04.025 (2013).
229 Yu, C. *et al.* Small Molecules Enhance CRISPR Genome Editing in Pluripotent Stem Cells. *Cell stem cell* **16**, 142-147, doi:10.1016/j.stem.2015.01.003 (2015).
230 Byrne, S. M., Ortiz, L., Mali, P., Aach, J. & Church, G. M. Multi-kilobase homozygous targeted gene replacement in human induced pluripotent stem cells. *Nucleic acids research* **43**, e21, doi:10.1093/nar/gku1246 (2015).
231 Canver, M. C. *et al.* Characterization of genomic deletion efficiency mediated by clustered regularly interspaced palindromic repeats (CRISPR)/Cas9 nuclease system in mammalian cells. *The Journal of biological chemistry* **289**, 21312-21324, doi:10.1074/jbc.M114.564625 (2014).
232 Holkers, M. *et al.* Adenoviral vector DNA for accurate genome editing with engineered nucleases. *Nature methods* **11**, 1051-1057, doi:10.1038/nmeth.3075 (2014).
233 Singh, P., Schimenti, J. C. & Bolcun-Filas, E. A mouse geneticist's practical guide to CRISPR applications. *Genetics* **199**, 1-15, doi:10.1534/genetics.114.169771 (2015).
234 Tsai, S. Q. & Zheng, Z. GUIDE-seq enables genome-wide profiling of off-target cleavage by CRISPR-Cas nucleases. *Nature biotechnology* **33**, 187-197, doi:10.1038/nbt.3117 (2015).
235 Wan, H. *et al.* One-step generation of p53 gene biallelic mutant Cynomolgus monkey via the CRISPR/Cas system. *Cell research* **25**, 258-261, doi:10.1038/cr.2014.158 (2015).
236 Jakociunas, T. *et al.* CasEMBLR: Cas9-facilitated multi-loci genomic integration of in vivo assembled DNA parts in Saccharomyces cerevisiae. *ACS synthetic biology*, doi:10.1021/acssynbio.5b00007 (2015).
237 Wu, C. I. *et al.* Function of Wnt/beta-catenin in counteracting Tcf3 repression through the Tcf3-beta-catenin interaction. *Development* **139**, 2118-2129, doi:10.1242/dev.076067 (2012).
238 Iwafuchi-Doi, M. & Zaret, K. S. Pioneer transcription factors in cell reprogramming. *Genes & development* **28**, 2679-2692, doi:10.1101/gad.253443.114 (2014).
239 Stergachis, A. B. *et al.* Developmental fate and cellular maturity encoded in human regulatory DNA landscapes. *Cell* **154**, 888-903, doi:10.1016/j.cell.2013.07.020 (2013).
240 Parker, S. C. *et al.* Chromatin stretch enhancer states drive cell-specific gene regulation and harbor human disease risk variants. *Proceedings of the National Academy of Sciences of the United States of America* **110**, 17921-17926, doi:10.1073/pnas.1317023110 (2013).
241 Langst, G. & Manelyte, L. Chromatin Remodelers: From Function to Dysfunction. *Genes* **6**, 299-324, doi:10.3390/genes6020299 (2015).
242 Eberhart, C. G., Tihan, T. & Burger, P. C. Nuclear localization and mutation of beta-catenin in medulloblastomas. *J Neuropathol Exp Neurol* **59**, 333-337 (2000).
243 Zurawel, R. H., Chiappa, S. A., Allen, C. & Raffel, C. Sporadic medulloblastomas contain oncogenic beta-catenin mutations. *Cancer Res* **58**, 896-899 (1998).
244 Devereux, T. R. *et al.* Mutation of beta-catenin is an early event in chemically induced mouse hepatocellular carcinogenesis. *Oncogene* **18**, 4726-4733, doi:10.1038/sj.onc.1202858 (1999).
245 Wend, P., Holland, J. D., Ziebold, U. & Birchmeier, W. Wnt signaling in stem and cancer stem cells. *Semin Cell Dev Biol* **21**, 855-863, doi:10.1016/j.semcdb.2010.09.004 (2010).

246 Huang, H. *et al.* APC mutations in sporadic medulloblastomas. *Am J Pathol* **156**, 433-437, doi:10.1016/s0002-9440(10)64747-5 (2000).
247 Yang, L. *et al.* FZD7 has a critical role in cell proliferation in triple negative breast cancer. *Oncogene* **30**, 4437-4446, doi:10.1038/onc.2011.145 (2011).
248 Rocheleau, C. E. *et al.* WRM-1 activates the LIT-1 protein kinase to transduce anterior/posterior polarity signals in C. elegans. *Cell* **97**, 717-726 (1999).
249 Yamada, M. *et al.* NARF, an nemo-like kinase (NLK)-associated ring finger protein regulates the ubiquitylation and degradation of T cell factor/lymphoid enhancer factor (TCF/LEF). *The Journal of biological chemistry* **281**, 20749-20760, doi:10.1074/jbc.M602089200 (2006).
250 Li, M. *et al.* TAB2 scaffolds TAK1 and NLK in repressing canonical Wnt signaling. *The Journal of biological chemistry* **285**, 13397-13404, doi:10.1074/jbc.M109.083246 (2010).
251 Lee, E., Salic, A. & Kirschner, M. W. Physiological regulation of [beta]-catenin stability by Tcf3 and CK1epsilon. *The Journal of cell biology* **154**, 983-993, doi:10.1083/jcb.200102074 (2001).
252 Gao, Y. & Wang, H. Y. Casein kinase 2 Is activated and essential for Wnt/beta-catenin signaling. *The Journal of biological chemistry* **281**, 18394-18400, doi:10.1074/jbc.M601112200 (2006).
253 Miravet, S. *et al.* The transcriptional factor Tcf-4 contains different binding sites for beta-catenin and plakoglobin. *The Journal of biological chemistry* **277**, 1884-1891, doi:10.1074/jbc.M110248200 (2002).
254 Yamamoto, H., Ihara, M., Matsuura, Y. & Kikuchi, A. Sumoylation is involved in beta-catenin-dependent activation of Tcf-4. *The EMBO journal* **22**, 2047-2059, doi:10.1093/emboj/cdg204 (2003).
255 Sachdev, S. *et al.* PIASy, a nuclear matrix-associated SUMO E3 ligase, represses LEF1 activity by sequestration into nuclear bodies. *Genes & development* **15**, 3088-3103 (2001).
256 Gay, F. *et al.* Acetylation regulates subcellular localization of the Wnt signaling nuclear effector POP-1. *Genes & development* **17**, 717-722, doi:10.1101/gad.1042403 (2003).
257 Elfert, S. *et al.* Acetylation of human TCF4 (TCF7L2) proteins attenuates inhibition by the HBP1 repressor and induces a conformational change in the TCF4::DNA complex. *PloS one* **8**, e61867, doi:10.1371/journal.pone.0061867 (2013).
258 Parker, C. E., Mocanu, V., Mocanu, M., Dicheva, N. & Warren, M. R. in *Neuroproteomics Frontiers in Neuroscience* (ed O. Alzate) (2010).
259 Hornbeck, P. V., Chabra, I., Kornhauser, J. M., Skrzypek, E. & Zhang, B. PhosphoSite: A bioinformatics resource dedicated to physiological protein phosphorylation. *Proteomics* **4**, 1551-1561, doi:10.1002/pmic.200300772 (2004).
260 Mayne, L. V. *et al.* SV 40-transformed normal and DNA-repair-deficient human fibroblasts can be transfected with high frequency but retain only limited amounts of integrated DNA. *Gene* **66**, 65-76 (1988).
261 Hoeijmakers, J. H., Odijk, H. & Westerveld, A. Differences between rodent and human cell lines in the amount of integrated DNA after transfection. *Experimental cell research* **169**, 111-119 (1987).
262 Lohrer, H., Blum, M. & Herrlich, P. Ataxia telangiectasia resists gene cloning: an account of parameters determining gene transfer into human recipient cells. *Molecular & general genetics : MGG* **212**, 474-480 (1988).
263 Colbere-Garapin, F., Ryhiner, M. L., Stephany, I., Kourilsky, P. & Garapin, A. C. Patterns of integration of exogenous DNA sequences transfected into mammalian cells of primate and rodent origin. *Gene* **50**, 279-288 (1986).
264 Kosuri, S. & Church, G. M. Large-scale de novo DNA synthesis: technologies and applications. *Nature methods* **11**, 499-507, doi:10.1038/nmeth.2918 (2014).

265 Cadinanos, J. & Bradley, A. Generation of an inducible and optimized piggyBac transposon system. *Nucleic acids research* **35**, e87, doi:10.1093/nar/gkm446 (2007).
266 Chen, Y. T. & Bradley, A. A new positive/negative selectable marker, puDeltatk, for use in embryonic stem cells. *Genesis* **28**, 31-35 (2000).
267 Yusa, K., Zhou, L., Li, M. A., Bradley, A. & Craig, N. L. A hyperactive piggyBac transposase for mammalian applications. *Proceedings of the National Academy of Sciences of the United States of America* **108**, 1531-1536, doi:10.1073/pnas.1008322108 (2011).
268 Hayashi, S. & McMahon, A. P. Efficient recombination in diverse tissues by a tamoxifen-inducible form of Cre: a tool for temporally regulated gene activation/inactivation in the mouse. *Developmental biology* **244**, 305-318, doi:10.1006/dbio.2002.0597 (2002).
269 Li, X. *et al.* piggyBac transposase tools for genome engineering. *Proceedings of the National Academy of Sciences of the United States of America* **110**, E2279-2287, doi:10.1073/pnas.1305987110 (2013).
270 Kim, J. H. *et al.* High cleavage efficiency of a 2A peptide derived from porcine teschovirus-1 in human cell lines, zebrafish and mice. *PloS one* **6**, e18556, doi:10.1371/journal.pone.0018556 (2011).
271 Wong, E. A. & Capecchi, M. R. Homologous recombination between coinjected DNA sequences peaks in early to mid-S phase. *Molecular and cellular biology* **7**, 2294-2295 (1987).
272 Lin, F. L., Sperle, K. M. & Sternberg, N. L. Extrachromosomal recombination in mammalian cells as studied with single- and double-stranded DNA substrates. *Molecular and cellular biology* **7**, 129-140 (1987).
273 Dellaire, G., Yan, J., Little, K. C., Drouin, R. & Chartrand, P. Evidence that extrachromosomal double-strand break repair can be coupled to the repair of chromosomal double-strand breaks in mammalian cells. *Chromosoma* **111**, 304-312, doi:10.1007/s00412-002-0212-6 (2002).
274 Davis, A. J. & Chen, D. J. DNA double strand break repair via non-homologous end-joining. *Translational cancer research* **2**, 130-143, doi:10.3978/j.issn.2218-676X.2013.04.02 (2013).
275 Maruyama, T. *et al.* Increasing the efficiency of precise genome editing with CRISPR-Cas9 by inhibition of nonhomologous end joining. *Nature biotechnology* **33**, 538-542, doi:10.1038/nbt.3190 (2015).
276 Chu, V. T. *et al.* Increasing the efficiency of homology-directed repair for CRISPR-Cas9-induced precise gene editing in mammalian cells. *Nature biotechnology* **33**, 543-548, doi:10.1038/nbt.3198 (2015).
277 Wang, H. *et al.* Biochemical evidence for Ku-independent backup pathways of NHEJ. *Nucleic acids research* **31**, 5377-5388 (2003).
278 Boulton, S. J. & Jackson, S. P. Identification of a Saccharomyces cerevisiae Ku80 homologue: roles in DNA double strand break rejoining and in telomeric maintenance. *Nucleic acids research* **24**, 4639-4648 (1996).
279 Frit, P., Barboule, N., Yuan, Y., Gomez, D. & Calsou, P. Alternative end-joining pathway(s): bricolage at DNA breaks. *DNA repair* **17**, 81-97, doi:10.1016/j.dnarep.2014.02.007 (2014).
280 Simsek, D. *et al.* DNA ligase III promotes alternative nonhomologous end-joining during chromosomal translocation formation. *PLoS genetics* **7**, e1002080, doi:10.1371/journal.pgen.1002080 (2011).
281 Liang, L. *et al.* Human DNA ligases I and III, but not ligase IV, are required for microhomology-mediated end joining of DNA double-strand breaks. *Nucleic acids research* **36**, 3297-3310, doi:10.1093/nar/gkn184 (2008).
282 Wang, H. *et al.* DNA ligase III as a candidate component of backup pathways of nonhomologous end joining. *Cancer Res* **65**, 4020-4030, doi:10.1158/0008-5472.CAN-04-3055 (2005).

283 Oh, S. *et al.* DNA ligase III and DNA ligase IV carry out genetically distinct forms of end joining in human somatic cells. *DNA repair* **21**, 97-110, doi:10.1016/j.dnarep.2014.04.015 (2014).
284 Decottignies, A. Alternative end-joining mechanisms: a historical perspective. *Frontiers in genetics* **4**, 48, doi:10.3389/fgene.2013.00048 (2013).
285 Roth, D. B. & Wilson, J. H. Nonhomologous recombination in mammalian cells: role for short sequence homologies in the joining reaction. *Molecular and cellular biology* **6**, 4295-4304 (1986).
286 Roth, D. B., Porter, T. N. & Wilson, J. H. Mechanisms of nonhomologous recombination in mammalian cells. *Molecular and cellular biology* **5**, 2599-2607 (1985).
287 Mason, R. M., Thacker, J. & Fairman, M. P. The joining of non-complementary DNA double-strand breaks by mammalian extracts. *Nucleic acids research* **24**, 4946-4953 (1996).
288 Dellaire, G. & Chartrand, P. Direct evidence that transgene integration is random in murine cells, implying that naturally occurring double-strand breaks may be distributed similarly within the genome. *Radiation research* **149**, 325-329 (1998).
289 Konopka, A. K. Compilation of DNA strand exchange sites for non-homologous recombination in somatic cells. *Nucleic acids research* **16**, 1739-1758 (1988).
290 Wurtele, H., Little, K. C. & Chartrand, P. Illegitimate DNA integration in mammalian cells. *Gene therapy* **10**, 1791-1799, doi:10.1038/sj.gt.3302074 (2003).
291 Merrihew, R. V., Marburger, K., Pennington, S. L., Roth, D. B. & Wilson, J. H. High-frequency illegitimate integration of transfected DNA at preintegrated target sites in a mammalian genome. *Molecular and cellular biology* **16**, 10-18 (1996).
292 Yan, B. W., Zhao, Y. F., Cao, W. G., Li, N. & Gou, K. M. Mechanism of random integration of foreign DNA in transgenic mice. *Transgenic research* **22**, 983-992, doi:10.1007/s11248-013-9701-z (2013).
293 Kool M, K. J., Bunt J, Hasselt NE, Lakeman A, van Sluis P, Troost D, Meeteren NS, Caron HN, Cloos J, Mrsić A, Ylstra B, Grajkowska W, Hartmann W, Pietsch T, Ellison D, Clifford SC, Versteeg R. Integrated genomics identifies five medulloblastoma subtypes with distinct genetic profiles, pathway signatures and clinicopathological features. *PloS one* **3** (2008).
294 Rogers HA, M. S., Lowe J, Brundler MA, Coyle B, Grundy RG. An investigation of WNT pathway activation and association with survival in central nervous system primitive neuroectodermal tumours (CNS PNET). *Br J Cancer* **100**, 1292-1302 (2009).
295 Blackwood DH, F. A., Walker MT, St Clair DM, Porteous DJ, Muir WJ. Schizophrenia and affective disorders--cosegregation with a translocation at chromosome 1q42 that directly disrupts brain-expressed genes: clinical and P300 findings in a family. *Am J Hum Genet.* **69**, 428-433 (2001).
296 Chubb JE, B. N., Soares DC, Porteous DJ, Millar JK. The DISC locus in psychiatric illness. *Mol Psychiatry* **13**, 36-64 (2008).
297 Hikida T, J.-P. H., Seshadri S, Oishi K, Hookway C, Kong S, Wu D, Xue R, Andradé M, Tankou S, Mori S, Gallagher M, Ishizuka K, Pletnikov M, Kida S, Sawa A. Dominant-negative DISC1 transgenic mice display schizophrenia-associated phenotypes detected by measures translatable to humans. *Proc Natl Acad Sci* **104**, 14501-14506 (2007).
298 Mao Y, G. X., Frank CL, Madison JM, Koehler AN, Doud MK, Tassa C, Berry EM, Soda T, Singh KK, Biechele T, Petryshen TL, Moon RT, Haggarty SJ, Tsai LH. Disrupted in schizophrenia 1 regulates neuronal progenitor proliferation via modulation of GSK3beta/beta-catenin signaling. *Cell* **136**, 1017-1031 (2009).
299 Doble, B. W. & Woodgett, J. R. Role of glycogen synthase kinase-3 in cell fate and epithelial-mesenchymal transitions. *Cells Tissues Organs* **185**, 73-84 (2007).
300 Kim WY, W. X., Wu Y, Doble BW, Patel S, Woodgett JR, Snider WD. GSK-3 is a master regulator of neural progenitor homeostasis. *Nat Neurosci.* **12**, 1390-1397 (2009).

301 Chenn A, W. C. Regulation of cerebral cortical size by control of cell cycle exit in neural precursors. *Science* **297**, 365-369 (2002).

302 Woodhead GJ, M. C., Olson EC, Chenn A. Cell-autonomous beta-catenin signaling regulates cortical precursor proliferation. *J Neurosci* **26**, 12620-12630 (2006).

303 Schumann CM, H. J., Goodlin-Jones BL, Lotspeich LJ, Kwon H, Buonocore MH, Lammers CR, Reiss AL, Amaral DG. The amygdala is enlarged in children but not adolescents with autism; the hippocampus is enlarged at all ages. *J Neurosci* **24**, 6392-6401 (2004).

304 Sparks BF, F. S., Shaw DW, Aylward EH, Echelard D, & AA, A. Brain structural abnormalities in young children with autism spectrum disorder. *Neurology* **59**, 184-192 (2002).

305 Flagstad, P. *et al.* Disruption of neurogenesis on gestational day 17 in the rat causes behavioral changes relevant to positive and negative schizophrenia symptoms and alters amphetamine-induced dopamine release in nucleus accumbens. *Neuropsychopharmacology* **29**, 2052-2064 (2004).

306 Götz, M. & Huttner, W. B. THE CELL BIOLOGY OF NEUROGENESIS. *Nature reviews. Molecular cell biology* **6**, 777-788 (2005).

307 Borello, U. & Pierani, A. Patterning the cerebral cortex: traveling with morphogens. *Curr Opin Genet Dev* **20**, 408-415 (2010).

308 Lee, S. M., Tole, S., Grove, E. & McMahon, A. P. A local Wnt-3a signal is required for development of the mammalian hippocampus. *Development* **127**, 457-467 (2000).

309 Galceran, J., Miyashita-Lin, E. M., Devaney, E., Rubenstein, J. L. R. & Grosschedl, R. Hippocampus development and generation of dentate gyrus granule cells is regulated by LEF1. *Development*, 469-482 (2000).

310 Maretto, S. *et al.* Mapping Wnt/beta-catenin signaling during mouse development and in colorectal tumors. *Proceedings of the National Academy of Sciences of the United States of America* **100**, 3299-3304 (2003).

311 Yi, F. *et al.* Opposing effects of Tcf3 and Tcf1 control Wnt stimulation of embryonic stem cell self-renewal. *Nature cell biology* **13**, 762-770 (2011).

312 Wilson, R. Preparation of single-stranded DNA from PCR products with streptavidin magnetic beads. *Nucleic acid therapeutics* **21**, 437-440, doi:10.1089/nat.2011.0322 (2011).

313 Goldmark, P. J. & Linn, S. Purification and properties of the recBC DNase of Escherichia coli K-12. *The Journal of biological chemistry* **247**, 1849-1860 (1972).

314 Wakimoto, Y., Jiang, J. & Wakimoto, H. Isolation of single-stranded DNA. *Current protocols in molecular biology* **107**, 2 15 11-19, doi:10.1002/0471142727.mb0215s107 (2014).

315 Srivastava, M. *et al.* An inhibitor of nonhomologous end-joining abrogates double-strand break repair and impedes cancer progression. *Cell* **151**, 1474-1487, doi:10.1016/j.cell.2012.11.054 (2012).

316 Veuger, S. J., Curtin, N. J., Richardson, C. J., Smith, G. C. & Durkacz, B. W. Radiosensitization and DNA repair inhibition by the combined use of novel inhibitors of DNA-dependent protein kinase and poly(ADP-ribose) polymerase-1. *Cancer Res* **63**, 6008-6015 (2003).

317 IDT. Antisense Technologies. *Integrated DNA Technologies*, doi:https://www.idtdna.com/pages/docs/educational-resources/antisense-technologies.pdf (2011).

318 Konermann, S. *et al.* Genome-scale transcriptional activation by an engineered CRISPR-Cas9 complex. *Nature* **517**, 583-588, doi:10.1038/nature14136 (2015).

VITA

NAME: Brian Rudolph Shy

EDUCATION: B.A., Economics, Michigan State University, East Lansing, MI, 2006

M.D., College of Medicine, University of Illinois at Chicago, Chicago, IL, 2017

Ph.D., Biochemistry and Molecular Genetics, University of Illinois at Chicago, Chicago, IL, 2017

HONORS: Dean's List, Michigan State University, East Lansing, MI, 2001-2006

Magna Cum Laude, Michigan State University, East Lansing, MI, 2006

Travel Fellowship, Peripheral Nerve Society, 2007

Predoctoral Fellowship, Lung & Vascular Biology Training grant (NIH T32 HL7829-19), Department of Pharmacology, University of Illinois at Chicago, Chicago, IL 2013-2015

Chicago Biomedical Consortium (CBC) Scholar, CBC, Chicago, IL, 2013-2015

Otto Saphir Award in Pathology, Department of Pathology, University of Illinois at Chicago, Chicago, IL, 2016

PUBLICATIONS: Shy BR, MacDougall MS, Clarke R, Merrill BJ. Co-incident insertion enables high efficiency genome engineering in mouse embryonic stem cells. Nucleic Acids Research, August 2016.

Shy BR, Wu C, Zhang J, Khramstova GF, Olopade OI, Goss KH, Merrill BJ. Regulation of Tcf7l1 protein stability and DNA binding as principle mechanisms of Wnt/β-catenin signaling. Cell Reports, July 2013. (*Cover)

Saporta MA, Shy BR, Patzko A, Bai Y, Pennuto M, Ferri C, Tinelli E, Saveri P, Kirschner D, Crowther M, Southwood C, Wu X, Gow A, Feltri ML, Wrabetz L, Shy ME. MpzR98C arrests Schwann cell development in a mouse model of early-onset Charcot-Marie-Tooth disease type 1B. Brain, July 2012.

Wu C, Hoffman JA, Shy BR, Ford EM, Fuchs E, Nguyen H, Merrill BJ. Function of Wnt/β-catenin in counteracting Tcf3 repression through the Tcf3-β-catenin interaction. Development, June 2012.

Yi F, Pereira L, Hoffman JA, Shy BR, Yuen CM, Liu DR, Merrill BJ. Opposing effects of Tcf3 and Tcf1 control Wnt stimulation of embryonic stem cell self-renewal. Nature Cell Biology, June 2011.