Title
Defining the Function of TCF7L1 in Human Embryonic Stem Cell Pluripotency

Author
Sierra, Robert Anthony

Publication Date
2014

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA,
IRVINE

Defining the Function of *TCF7L1* in Human Embryonic Stem Cell Pluripotency

DISSERTATION

submitted in partial satisfaction of the requirements
for the degree of

DOCTOR OF PHILOSOPHY

in Biomedical Sciences

by

Robert Anthony Sierra

Dissertation Committee:
Professor Peter J. Donovan, Chair
Professor Marian L. Waterman
Professor Peter Kaiser
Professor Xing Dai

2014

DEDICATION

To

My Parents

for your enduring, benevolent, support.

TABLE OF CONTENTS

LIST OF FIGURES

ACKNOWLEDGMENTS

First, I would like to thank Dr. Peter Donovan for giving me an opportunity to study in his laboratory. The lessons I have learned, both professionally and personally, while under his tutelage have been invaluable. Thank you for your guidance and thank you for giving me a venue to grow as a scientist and as an individual.

Next, I would like to express my sincerest form of gratitude to Dr. Marian Waterman who was not merely a collaborator, but a true mentor in every sense of the word. Dr. Waterman generously welcomed me into her laboratory as a first year graduate student as if I was one of her own. She will always be an enduring example of hard work, generosity, dedication and passion. Thank you.

I would be remised if I did not say a few words about the collaborative experience between Dr. Donovan and Dr. Waterman, which was remarkable and worthy of its own sentiment. Our meetings and discussions were stimulating, enlightening and a highlight of graduate school experience. I am truly grateful for this unique opportunity.

Also, I thank my committee members Dr. Peter Kaiser and Dr. Xing Dai for their intelligent perspectives, stimulating discussions and valuable suggestions.

Lastly, I would like to thank the California Institute of Regenerative Medicine (CIRM) for supporting and funding this research (RB2-01629). Moreover, I would like to extend my gratitude to CIRM for awarding me with the pre-doctoral training fellowship (TG2-01152). The opportunity and resources made available by this fellowship have been invaluable to my research project.

CURRICULUM VITAE

Robert Anthony Sierra

EDUCATION

Doctor of Philosophy in Biomedical Sciences **2014**
University of California, Irvine *Irvine, California*

Masters of Science in Biotechnology **2008**
University of California, Irvine *Irvine, California*

Bachelor of Arts in Developmental Biology **2004**
University of California, Santa Cruz *Santa Cruz, California*

RESEARCH EXPERIENCE

Doctoral Graduate Researcher **2008–2014**
University of California, Irvine *Irvine, California*

Masters Graduate Researcher **2008**
University of California, Irvine *Irvine, California*

TEACHING EXPERIENCE

Teaching Assistant **Fall 2008**
DNA to Organisms (Bio Sci 93) *Irvine, California*

HONORS AND AWARDS

CIRM Training Fellowship **2012 - Present**
University of California, Irvine *Irvine, California*

Graduate Division Fellowship **2007**
University of California, Irvine *Irvine, California*

Honors in Major **2004**
University of California, Santa Cruz *Santa Cruz, California*

REFEREED JOURNAL PUBLICATIONS

Stringari, C., **Sierra, R.**, Donovan, P.J., and Gratton, E.
Label-free separation of human embryonic stem cells and their differentiating progenies by phasor fluorescence lifetime microscopy.
J Biomed Opt 17, 046012. **2012**

ABSTRACT OF THE DISSERTATION

Defining the Function of *TCF7L1* in Human Embryonic Stem Cell Pluripotency

By

Robert Anthony Sierra

Doctor of Philosophy in Biomedical Sciences

University of California, Irvine, 2014

Professor Peter J. Donovan, Chair

A critical stage of human development is the specification of the three primary germ layers that will comprise the embryo: a process known as gastrulation. During gastrulation pluripotent, unspecialized, cells of the epiblast differentiate into either endoderm, mesoderm or ectoderm lineages, a process coordinated in part by extrinsic signaling cues. The first event of gastrulation is the directed differentiation of epiblast cells towards mesendoderm, a bipotent progenitor that eventually becomes restricted to either endoderm and mesoderm cell types. Concomitantly with differentiation, nascent mesendoderm cells undergo an epithelial-to-mesenchymal transition (EMT) at the primitive streak, where the cells migrate from the epiblast layer to form a new layer of cells. Because of obvious ethical and legal reasons, it is not possible to study this stage of human development *in vivo*. However, studies suggest that human embryonic stem cells (hESCs) are *in vitro* equivalents of pluripotent epiblast cells. They therefore represent a feasible system for modeling human gastrulation *in vitro*. Since WNTs are known regulators of gastrulation in all metazoan species, it seemed logical to test the role of this pathway in epiblast differentiation. Specifically, I used hESCs to dissect the role of the transcriptional repressor *TCF7L1*, a downstream transcription factor of the canonical WNT pathway, in the regulation of pluripotency. I found that *TCF7L1* acts as a key regulator of mesendoderm differentiation in hESCs. I also identify a novel regulatory

interaction between the BMP4 pathway (another key regulator of mesendoderm/primitive streak formation *in vivo*) and *TCF7L1*, whereby activation of the BMP4 signaling pathway induces *TCF7L1* down-regulation. Based on these findings I hypothesize that the action of BMP4 on *TCF7L1* leads to the loss of repression of mesendoderm genes and activates mesendoderm differentiation during gastrulation. Thus, this study marks *TCF7L1* as a key suppressor of mesendoderm differentiation in hESCs and provides valuable insights into the regulation of gastrulation *in vivo*. My studies also make several predictions about the control of signaling events in pluripotent stem cells *in vitro* and *in vivo* and provide new avenues in this exciting area of research.

Chapter 1

Introduction

Human embryonic stem cells (hESCs) were first derived in 1998 by Dr. James Thomson and colleagues [113]. Using immunosurgery, they harvested cells of the inner cell mass (ICM) of three-to-five day old human blastocysts and stably maintained them *in vitro* using optimized medium conditions (**Figure 1.1**). In culture, these cells exhibited two key properties: pluripotency and self-renewal. Pluripotency defines an undifferentiated cell that has the capacity to differentiate into cell types representative of all three germ layers (endoderm, mesoderm and ectoderm) present in the developing embryo. Self-renewal defines the ability of a cell to continuously divide and expand. Notably, hESC derivation came seventeen years after the derivation of mouse ESCs (mESCs), highlighting the inherent challenges and monumental significance of this achievement [68][35]. Dr. Thomson's discovery paved the way for an exciting new field of research with unprecedented implications for regenerative medicine and human developmental biology.

While mESCs have long been the paradigm for mammalian ESC research, they do have inherent limitations for human applications. Human ESC derivation was a milestone because it provided a species-matching model system to study human development and therapeutic

applications. In order to fully utilize the true potential of hESCs, however, we must first develop a comprehensive understanding of their biology. To this end, the work presented here aims to further our understanding of mechanisms regulating hESC pluripotency. More specifically, I investigate how the WNT signaling pathway - a highly conserved signaling pathway necessary for metazoan development - is involved in the regulation of hESC pluripotency and differentiation. Because hESCs are thought to be equivalent to cells of the developing epiblast, I also discuss how my findings apply to human development *in vivo*.

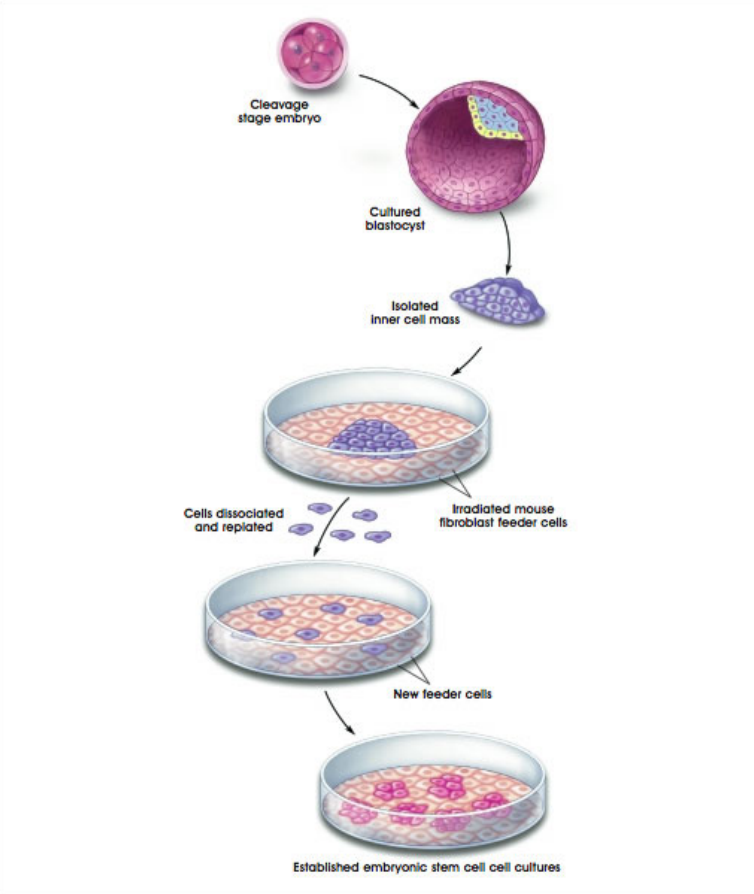


Figure 1.1: **Derivation of human embryonic stem cells**. Excess cleavage-stage human embryos, donated by *in vitro* fertilization (IVF) donors, were grown to the blastocyst stage of development. The inner cell mass cells were then isolated by immunosurgery and cultured under appropriate conditions until they gave rise to cells that would be expanded indefinitely and had the properties of pluripotent cells. *Figure taken from http://stemcells.nih.gov.*

1.1 Human ESCs: a Model of Early Human Development

As a model system, hESCs present a unique opportunity to study the development of our own species. Studies in model organisms (e.g., *M. musculus*, *D. melanogaster*, *C. elegans*, *X. laevis*, *G. gallus*) have undoubtedly made great contributions to our understanding of early embryogenesis, but deeper insights into human development are likely best derived from a matching, species-specific research system. Due to obvious ethical and moral reasons, however, human systems are not amenable to the various experimental paradigms available in these model species. Considering these limitations, hESCs are the most-fitting model system for studying early human developmental processes *in vitro* since they are capable of reconstituting early *in vivo* fate-decisions in culture [139]. Thus, hESC research might allow us to probe the mechanisms orchestrating early human developmental events.

Embryogenesis commences with the fusion of a male gamete (sperm) pronucleus and a female gamete (egg) pronucleus (day 0) to generate a diploid zygote [84]. What follows is an oocyte-to-embryo transition, which lasts for approximately 3 days (day 0-3) and involves genetic and epigenetic reprogramming events that prepare the embryonic genome for transcriptional activation (day 3). During this time, three cleavage divisions increase the number of cells in the embryo from 1 to 8 (day 1-3) (**Figure 1.2**). With each successive cleavage division the overall size of the embryo remains constant. Therefore, the dividing cells get sequentially and uniformly smaller. On the fourth day the morula is formed, which arises from a process called compaction and produces a compact cluster of tightly adherent blastomeres (day 4). Continued cell divisions over the next 24 hours lead to the generation of the blastocyst (day 5): a fluid-filled circular-structure consisting of trophectoderm and pluripotent ICM cells. After the fifth day (day 5) the embryo prepares for implantation into the uterine wall (day 7) by "hatching" from the encapsulating zona pellucida glycoprotein layer (day 6) and by

inner cell mass (ICM) divergence to pluripotent epiblast cells and cells of the hypoblast (**Figure 1.3**).

After implantation, the three primary germ layers (endoderm, mesoderm and ectoderm) are specified from the pluripotent cells of the epiblast epithelium through a process known as gastrulation [102]. At this stage of development the embryo has a bi-layer organization consisting of an overlying epiblast epithelium and an underlying hypoblast (**Figure 1.4**). Precisely controlled spaciotemporal growth factor signaling events initiate gastrulation at a site called the primitive streak: a structure arising from the migration of differentiating epiblast cells of the epiblast epithelium into the intervening space between the epiblast and hypoblast layers. The process of differentiation and migration is known as an epithelial-mesenchymal transition (EMT) [111]. It is here, at the primitive streak, that epiblast cells exit the pluripotent-state and become mesendoderm: a bi-potent progenitor cell type that will give rise to either mesoderm and endoderm. This is a stage of development that is unavailable to experimental manipulation in humans, therefore hESCs - considered to be the representation of an epiblast-stage pluripotent cell - afford us the ability to study early developmental processes such as the the epiblast-to-mesendoderm transition *in vitro* [110][85].

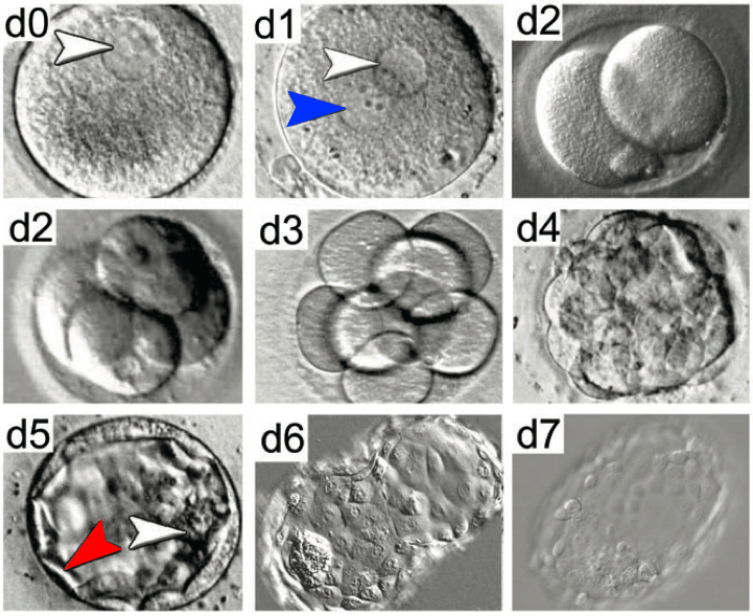


Figure 1.2: **Stages of human pre-implantation development**. Phase-contrast images showing the first 7 days of human embryonic development. Fertilization and fusion of the two gamete pronuclei (white and blue arrows) occurs during the first 24 hours (d0 - d1). From d1 to d3 the egg undergoes 3 rounds of cleavage divisions followed by compaction (d4) to from the morula. At d5 the blastocyst is formed with the inner cell mass (white arrow) located inside of a fluid-filled sphere of trophectoderm (red arrow). "Hatching" of the blastocyst from the zona pellucida occurs on d6, at which time the embryo is ready for implantation on d7. *Figure taken from [84] with permission.*

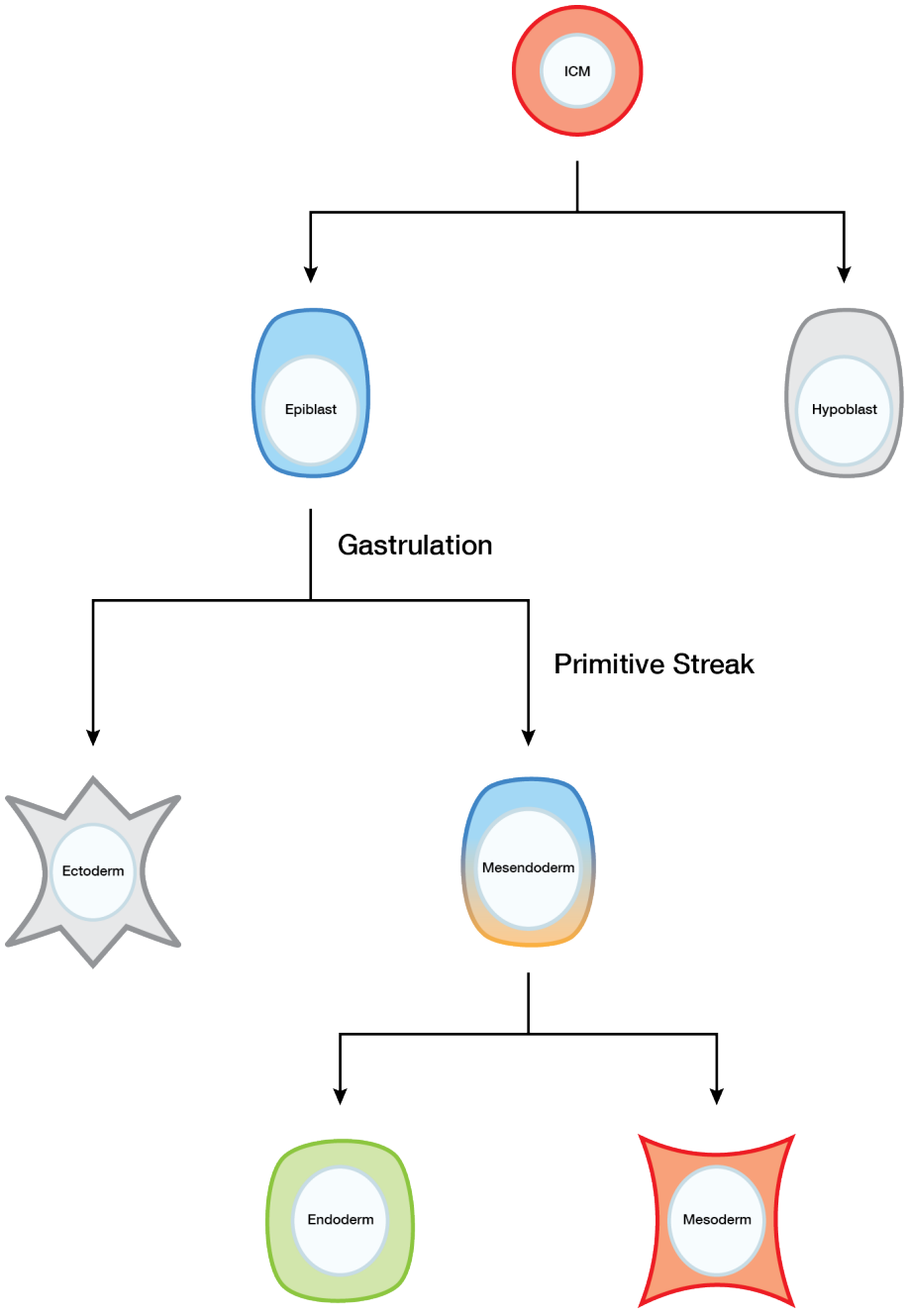


Figure 1.3: **Early developmental cell fate decisions**. Pluripotent inner cell mass (ICM) cells give rise to pluripotent epiblast cells and cells of the hypoblast. During gastrulation epiblast cells differentiate into all three primary germ layers: ectoderm, endoderm and mesoderm. Epiblast cells within the region of the developing primitive streak become a bi-potent progenitor cell type called mesendoderm, which then becomes either endoderm or mesoderm.

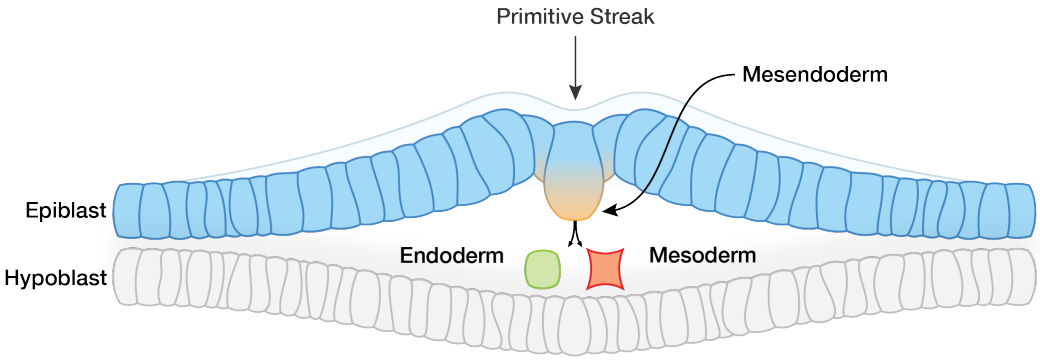


Figure 1.4: **Human primitive streak formation**. A cross section of a bi-laminar human embryo illustrating primitive streak formation. Spaciotemporal growth factor signaling events at the posterior end of the embryo trigger epiblast differentiation into mesendoderm at the nascent primitive streak. As this occurs, differentiating epiblast cells migrate into the space between the upper epiblast epithelium layer (blue) and lower hypoblast layer (grey). The mesendoderm (blue/orange) cells will go on to become either endoderm (green) or mesoderm (red).

1.2 Key Developmental Signaling Pathways

Because hESCs are developmentally potent cells, they can be used as a proxy for studying the mechanisms regulating fate decisions *in vitro*. Arguably, one of the most interesting developmental events during embryogenesis is gastrulation: the process of germ layer specification. By using hESCs, we can test hypotheses and fill vital gaps in our knowledge about how this process is regulated *in vivo*. Of particular interest, is the very first differentiation event where pluripotent epiblast cells cells transition into mesendoderm cells as gastrulation commences and in doing so form a structure know as the primitive streak. Coordinated and specific signaling pathway activation by *NODAL*, *WNT3* and *BMP4*, drive this process by initiating epiblast cells to differentiate (**Figure 1.5**) [106]. These three pathways are outlined below and are continuously referenced throughout this thesis and frequently referred to in experiments. Studies in a variety of systems have confirmed the conserved role of these factors in vertebrate gastrulation.

A great deal of early mammalian development research comes from mouse genetics studies, so it is worth describing the structure and terminology used to describe the mouse embryo. The image below illustrates the post-implantation embryo as gastrulation is set to begin (**Figure 1.5**). This stage is commonly referred to as the "egg cylinder" stage and has two axes: proximal-distal (top-bottom) and anterior-posterior (left-right). The visceral endoderm (VE - green) encloses the extraembryonic ectoderm (ExE - red) and pluripotent epiblast cell layer (pink) and is involved in the patterning of the embryo. The ExE consists of epithelial cells that lie directly above the epiblast cells and posterior ExE cells secret morphogens (i.e., Bmp4) which initiate epiblast cell differentiation and formation of the primitive streak. Initiation and maintenance of this epiblast differentiation occurs on the posterior side of the embryo and requires the collective actions of Bmp4, Nodal and Wnt3 signaling. In addition, a specialized region of the VE forms the distal VE (DVE), which serves as a unique source of regionalized signals that establish anterior-posterior and proximal-

distal patterning. In all, highly complex spaciotemporal signaling events properly pattern and develop the embryo during embryogenesis.

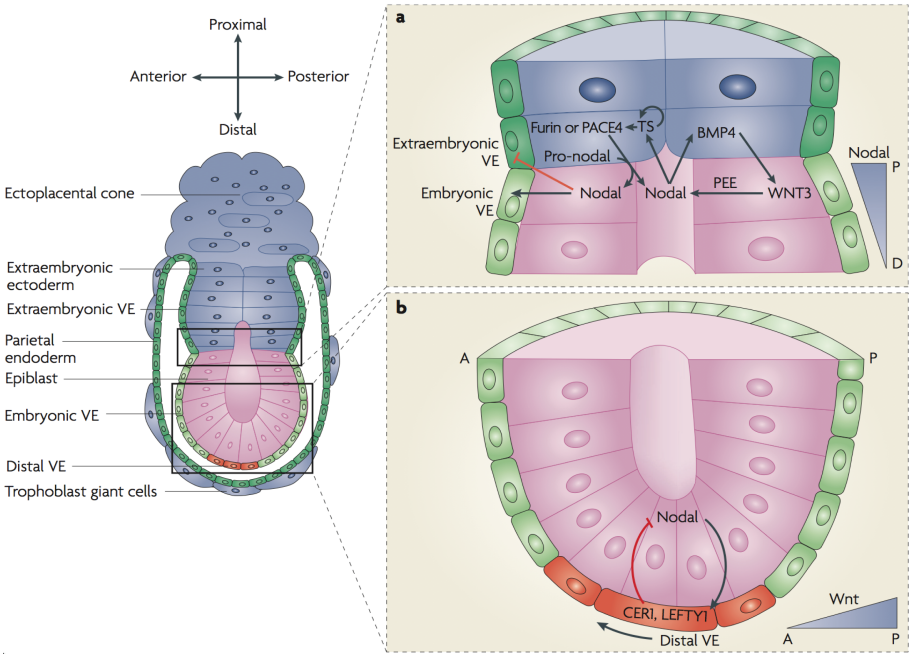


Figure 1.5: **Pre-gastrulation mouse embryo**. This image illustrates the structure and organization of an E5.5 mouse embryo right before gastrulation begins. This image also highlights how signaling pathways are involved in this process. The epiblast cells (pink) are pluripotent and receive differentiation signals from the overlying extra-embryonic ectoderm (blue). On the posterior side of the embryo, epiblast cells that are stimulated by BMP4, WNT3 and NODAL begin mesendoderm differentiation and create the primitive streak. *Figure taken from [2] with permission.*

1.2.1 The *BMP4* Signaling Pathway

Studies in mice have shown that *Bmp4* (bone morphogenetic protein 4) plays a critical role in gastrulation [106]. During normal development at the onset of gastrulation, *Bmp4* is expressed in a posterior region of the extra-embryonic ectoderm adjacent to the nascent primitive streak, and triggers the underlying epiblast cells to differentiate towards

a mesendoderm fate (**Figure 1.5**). Homozygous-null *Bmp4* mutant mice die between E6.5 - 9.5 days postcoitum (dpc) due to severely compromised mesoderm formation and in some cases a complete failure to form a primitive streak [127]. Therefore, these data demonstrate the key role that *Bmp4* plays in gastrulation.

BMP4 is a member of the transforming growth factor β (TGFβ) superfamily of signaling morphogens [128][21]. When BMP4 binds with its cognate serine-threonine kinase receptors, it initiates a signaling cascade that is propagated by receptor-regulated SMADs (R-SMADs) proteins, which in turn regulate target gene transcription. In this case, BMP4 transduces its signal by binding to BMPR1A (type I) and BMPR2 (type II) receptors, both of which are also required for progression through the primitive streak stage of development [7][75]. Dimerization of these receptors leads to intracellular phosphorylation of SMAD1, SMAD5 and SMAD8 (collectively referred to as SMAD1/5/8) [77]. Once phosphorylated, phospho-SMAD1/5/8 forms a complex with the mediator SMAD4 and then translocates into the nucleus to regulate target gene transcription.

1.2.2 The *NODAL* Signaling Pathway

Like *BMP4*, *NODAL* is a TGFβ superfamily signaling morphogen that is absolutely required for primitive streak formation. Signaling transduction occurs via the interaction between NODAL its cognate serine-threonine kinase receptors ACVR1B (type I) and ACVR2A/B (type lI), which induces SMAD2 and SMAD3 (SMAD2/3) phosphorylation [99]. Like phospho-SMAD1/5/8, phospho-SMAD2/3 heterodimerizes with the mediator SMAD4, translocates into the nucleus and regulates transcription of its target genes. Unlike SMAD1/5/8, SMAD2/3 recognizes a different DNA sequence and, therefore, regulates its own distinct set of target genes [132].

During gastrulation in mice, *Nodal*'s function has been described as an inducer/maintainer

of the primitive streak [28]. As gastrulation commences, *Nodal* expression becomes localized to the proximal posterior ectoderm and drives epiblast differentiation within the forming primitive streak (**Figure 1.5**) [28]. Studies in mice have shown that genetic ablation of *Nodal* inhibits primitive streak formation causing developmental arrest shortly after gastrulation [28][13]. Interestingly, Nodal is initially secreted as a protein precursor that contains a pro domain that increases its stability and signaling range [56]. Proteolytic cleavage by the extracellular convertases Furin (*Spc1*) and Pace4 (*Spc4*) removes this pro domain, which increases Nodal autocrine signaling and decreases its stability [6].

1.2.3 The WNT Signaling Pathway

The canonical WNT signaling pathway and its role in hESC pluripotency and differentiation is a focal point of this study and therefore it will be discussed in greater detail.

The WNT signaling pathway is highly conserved across all metazoan species and involved in key processes during embryogenesis like: cell proliferation, migration, differentiation, cell polarity and survival [119][118][63]. WNT ligands are evolutionarily conserved cystine-rich secreted morphogens that bind to FRIZZLED and low-density lipoprotein (LRP) family receptors, as well as other co-receptors like receptor tyrosine kinase-like orphan receptor (ROR), protein tyrosine kinase 7 (PTK7) and receptor tyrosine kinase (RYK) [86]. In mammals, there are 19 WNT genes which are commonly segregated into two categories: the non-canonical WNT ligands and the canonical WNT ligands. Non-canonical WNT ligands (WNT-4, -5A, -5B, -6, -7A, -7B, -11) activate GTPases (RAC1, RHOA and JNK) involved in regulating planar cell polarity [86]. Additionally, non-canonical WNT ligands can also regulate intracellular calcium levels to affect cell adhesion and migration [86]. The canonical WNT pathway is activated by the canonical WNT ligands (WNT-1, -2, -3, -3A, -8A, -8B, -10A and -10B) and signal transduction depends on the activity of the transcriptional

activator β-catenin (*CTNNB1*) (**Figure 1.6**). Since the canonical pathway is the major focus of this work, it will be discussed in greater detail below.

The canonical WNT signaling pathway exists in either an ON or an OFF state depending on the presence of extracellular WNT ligands. In the OFF state β-catenin is targeted for proteasomal degradation by a "destruction complex" containing various proteins such as GSK3β, AXIN2, APC, PP2A and CK1 (**Figure 1.6**) [55]. Under these circumstances, β-catenin is absent from the nucleus and WNT target genes are transcriptionally inactive. In the ON state, extracellular WNT ligands promote disassembly of the "destruction complex" leading to stabilization and subsequent nuclear translocation of β-catenin (**Figure 1.6**). Once in the nucleus, β-catenin binds to lymphoid enhancer factor/T cell factor (LEF/TCF) family transcription factors and transcriptionally activates WNT target genes [1].

As downstream transcription factor effectors of the canonical WNT signaling pathway, the LEF/TCFs are particularly interesting because they regulate transcription of WNT target genes in the ON and OFF states. In all, there are four members of this family: *LEF1*, *TCF7*, *TCF7L1* and *TCF7L2* which share a highly conserved domain structure (**Figure 1.7**). All four members have a highly conserved high mobility group (HMG) DNA binding domain that recognizes the DNA sequence "CTTTGWW" (W represents adenine or thymine) known as a WNT response element (WRE) [3]. At their N-terminus is a highly conserved β-catenin binding domain that binds β-catenin to direct transcriptional activation of WNT target genes. When the pathway is inactive, however, LEF/TCFs can recruit transcriptional repressors like Groucho/TLE and histone deacetylases (HDACs) to silence WNT target gene transcription. Curiously, the LEF/TCFs have different intrinsic abilities for activation and repression; *LEF1* and *TCF7* are strong transcriptional activators, whereas *TCF7L1* and *TCF7L2* are strong transcriptional repressors [73]. While the molecular basis of these differences have not been fully elucidated, it has been shown that TCF7L1 and TCF7L2 have a higher binding affinity for the repressor TLE1, whereas LEF1 and TCF7 do not bind

to TLE1 [24]. Thus, the transcriptional activity of WNT target genes is influenced by the repertoire of LEF/TCFs expressed in cells.

Like BMP4 and NODAL pathways, canonical WNT signaling is also required for proper primitive streak formation [52][67]. Genetic experiments have shown that *Wnt3*-/- mice fail to form a primitive streak and die during embryogenesis [61]. In fact, the epiblast cells themselves must express Wnt3 in order for epiblast cells to maintain mesendoderm differentiation within the forming primitive streak [114]. Thus, WNT3 has a central role in early development at a stage equivalent to that represented by hESCs. My studies, therefore, aimed to use hESCs to model this stage of development and to decipher the role of WNT signaling pathway components in regulating differentiation decisions.

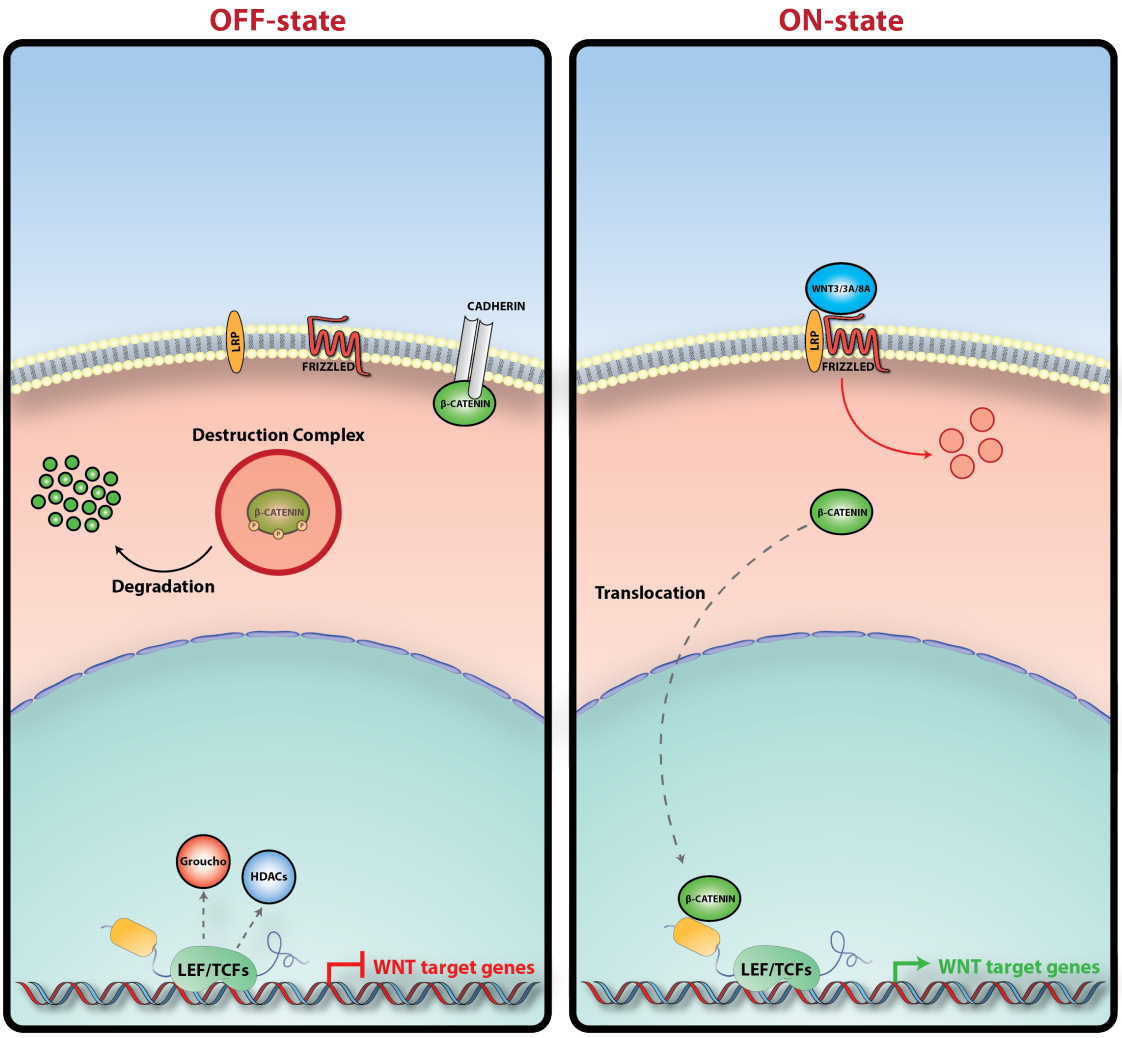


Figure 1.6: **The canonical WNT signaling pathway**. The canonical WNT signaling pathway depends on β-Catenin to mediate its downstream transcriptional responses. If there are no extracellular canonical WNT ligands (e.g., WNT3, WNT3A, WNT8A) the pathway is in the OFF-state. Under these conditions the "destruction complex" (GSK3β, AXIN, APC, PP2A, CK1 to name a few) induces site-specific phosphorylation of β-Catenin, leading to its ubiquitination and proteasomal degradation. When the pathway is OFF the LEF/TCFs recruit repressors to silence expression of WNT target genes. In the presence of canonical WNT ligands on the other hand, the WNT binds to FRIZZLED and LRP receptors, disrupting the destruction complex and stabilizing β-Catenin. Stabilized β-Catenin then translocates into the nucleus, binds to the LEF/TCFs, displaces the interacting repressors and recruits transcription activating co-regulators to induce WNT target gene transcription.

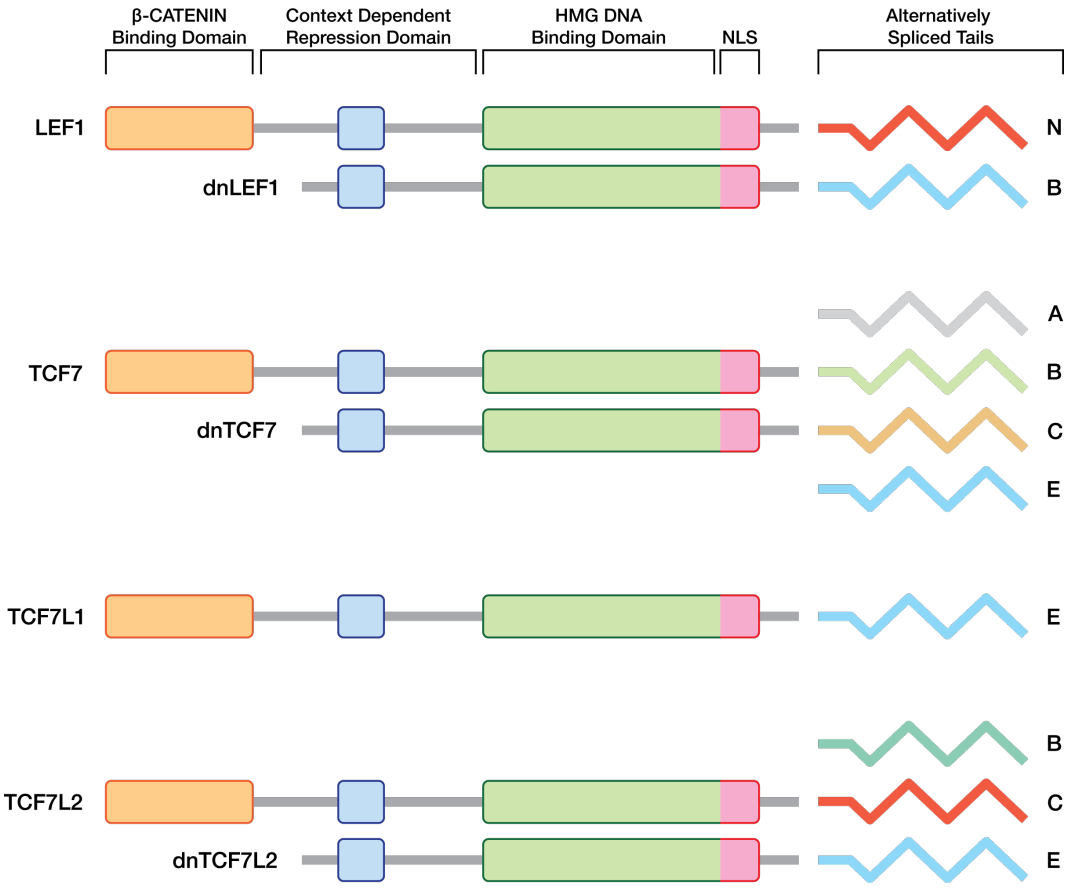


Figure 1.7: **The LEF/TCF transcription factors**. The LEF/TCF transcription factors share a very similar domain structure. All LEF/TCFs have a highly conserved N-terminal β-catenin binding domain ($\sim$60% sequence identity) and HMG DNA-binding domain ($\sim$95-99% sequence identity) which contains a nuclear localization signal (NLS) [1]. Variability exists within the context dependent repression domain and it is involved in the recruitment of repressors (i.e., Groucho/TLE) and interacting transcription factors. Interestingly, alternative promoter usage can generate dominant negative isoforms of LEF1 (dnLEF1), TCF7 (dnTCF7) and TCF7L2 (dnTCF7L1) which lack the β-catein binding domain. Furthermore, alternative splicing events produce proteins with different C-terminal tails; the functions of these different tails has not been fully appreciated, but they can confer additional DNA binding specificity and recruit additional repressor complexes [50].

1.3 Properties of hESCs

Pluripotency and self-renewal are the defining properties of hESCs (**Figure 1.8**). Pluripotency describes the potential of a cell to differentiate into cell types of all three germ layers of the embryo: endoderm, mesoderm and ectoderm. Like hESCs, mESCs, cells of the inner cell mass and epiblast cells are all considered pluripotent. Self-renewal is the ability of hESCs to indefinitely produce identical pluripotent copies of themselves through symmetric cell divisions. With proper culture technique and conditions, hESCs can be grown indefinitely *in vitro* while stably maintaining these key properties.

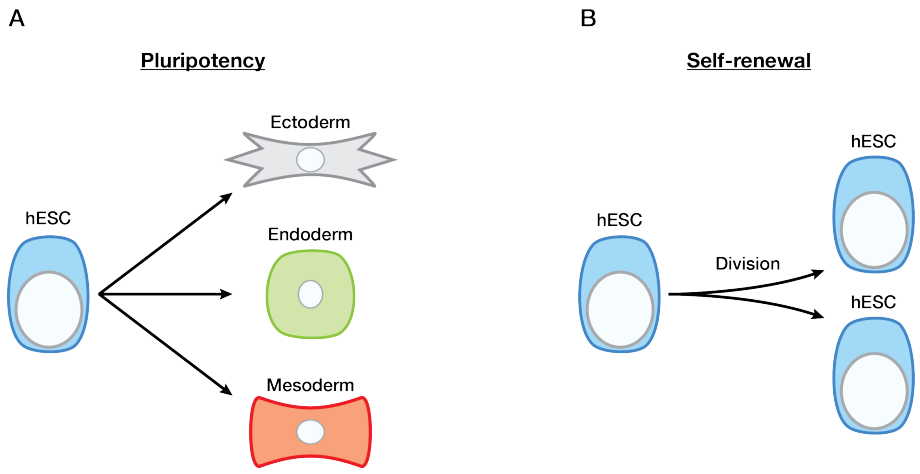


Figure 1.8: **Pluripotency and self-renewal**. **A**. Pluripotent hESCs are capable of multi-lineage differentiation into cell types of all three germ lineages: ectoderm, endoderm and mesoderm. **B**. Self-renewal defines the ability of hESCs to indefinitely propagate themselves *in vitro* while maintaining their pluripotent-state. This is achieved through symmetric cell divisions (approximately every 24 hours) generating two identical hESCs.

Human ESCs are commonly cultured on mitotically inactivated mouse embryonic fibroblast (MEFs) and grow as tight, compact, colonies consisting of thousands of cells (**Figure 1.9**). Within colonies, individual hESCs are firmly held together by cell-cell adhesion complexes resulting in well defined boundaries visible by phase microscopy [121]. Maintenance of

hESCs self-renewal and pluripotency *in vitro* depends on activation of the PI3K/Akt and SMAD2/3 signaling pathways, features that are covered in more detail in **Section 1.4**. Alternatively, hESCs can be grown under feeder-free conditions to prevent unwanted growth factor influences or genetic contamination from the MEFs. Under these conditions, hESCs are grown on a reconstituted basement membrane extract known as Geltrex and supported by a chemically defined culture media called mTeSR1 (**Figure 1.9**) [64].

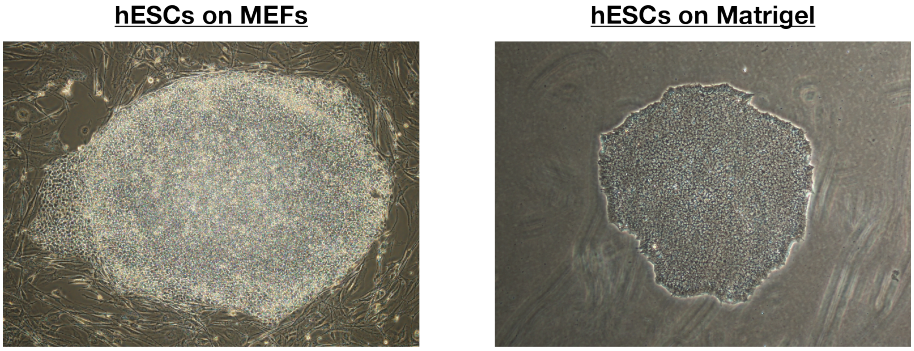


Figure 1.9: **Human ESC culture methods**. The phase contrast image on the left shows a single colony of hESCs cultured on mouse embryonic fibroblasts (MEF) feeder cells. The hESC colony is the dense circular population of cells in the center of the image and the long filamentous cells around the periphery are the MEFs. The phase contrast image on the right shows a single hESC colony grown under feeder-free conditions on Geltrex with chemically defined medium. In feeder-free conditions hESCs still maintain their characteristic tight compact colony morphology.

Pluripotent hESCs can be identified by markers specific to the undifferentiated-state. Since hESCs have a unique transcriptome, they express high levels of transcription factors that are necessary for pluripotency and self-renewal; namely, *OCT4*, *SOX2*, *NANOG* as well as *TERT* (a telomerase subunit). *OCT4*, *SOX2* and *NANOG* are commonly referred to as the "core transcriptional regulatory network" (discussed in **Section 1.5**). Human ESCs can also be identified by their expression of distinct cell surface antigens: stage-specific embryonic antigen (SSEA)3, SSEA4, TRA-1-60, TRA-1-81 and alkaline phosphatase. These transcription factors and cell-surface markers are useful for assessing pluripotency and

monitoring differentiation using mRNA and protein analysis assays.

A key hallmark of hESC pluripotency is the capacity for multi-lineage differentiation. *In vitro*, these cells can be directed to differentiate into cell types of interest using specific growth factor administration procedures. Conversely, hESCs can be induced to undergo spontaneous multi-lineage differentiation with an embryoid body (EB) differentiation assay [93]. In this assay, hESC clumps are cultured in suspension and form spheroid structures with hollow fluid-filled cavities in the center (**Figure 1.10**). The formation of EBs is thought to mimic, to a certain degree, early gastrulation events and are comprised of cell types from all three germ layers [27]. Human ESC differentiation potential can also be tested *in vivo* using a teratoma assay. In this assay, hESCs are injected into immunocompromised mice where they form tumors called teratomas which contain cell types of all three germ layers (**Figure 1.10**). The *in vivo* conditions of the teratoma assay facilitates more diverse differentiated cell types than the *in vitro* EB assay; nevertheless, both assays are valuable methods for invoking multi-lineage hESC differentiation.

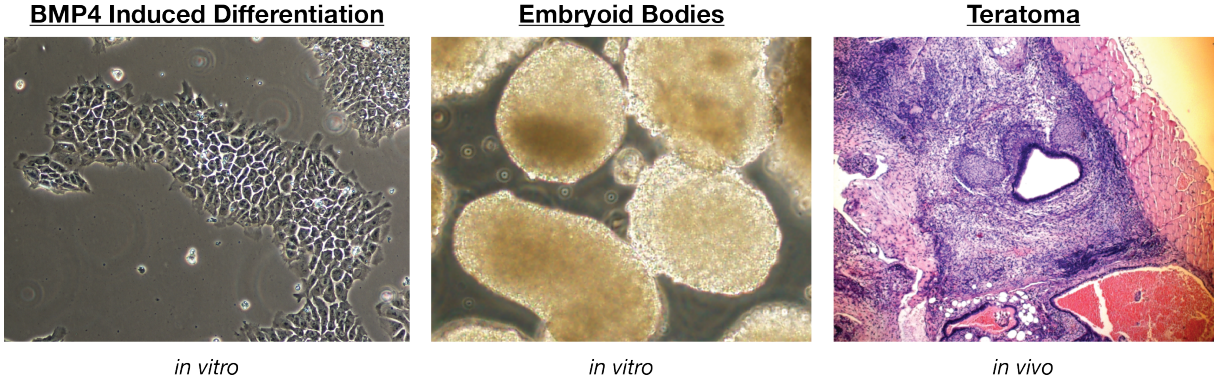


Figure 1.10: **Human ESC differentiation methods**. The phase contrast image (10X) on the left shows growth factor directed differentiation of hESCs under feeder-free conditions. BMP4 was used to induce hESC differentiation towards mesoderm *in vitro*. The center bright-field image is of 14 day old embryoid bodies (EBs) generated *in vitro*. Each spheroid structure is a single EB. The image on the right shows histological analysis of an *in vivo* generated teratoma tumor. The section was stained with hematoxylin and eosin to show the various cell types.

1.4 Human ESC Signaling Pathways Requirements

During development, the pluripotent status of cells only exists for a brief period of time leading up to germ layer specification. As such, pluripotency is not a fixed stable state, but a state in constant flux as the cells transition from the inner cell mass stage to the epiblast stage during embryogenesis. Due to the ephemeral nature of the pluripotent state *in vivo*, hESCs must be grown under very specific conditions in order to artificially stabilize and sustain an undifferentiated state *in vitro*. Specifically, *in vitro* propagation of hESCs requires extrinsic activation of two signaling pathways: the PI3K/Akt and TGFβ signaling pathways. Failure to properly stimulate these pathways results in hESC differentiation or even cell death.

The phosphatidylinositide 3-kinase (PI3K)/Akt pathway contributes to hESC pluripotency through inactivation of MEK/ERK kinases. ERK inhibits GSK3β (glycogen synthase kinase 3 beta), which targets the WNT pathway transcriptional activator β-catenin for proteasomal degradation, thus preventing β-catenin-mediated activation WNT target genes (**Figure 1.11**) [100]. The β-catenin-driven WNT signaling pathway is explained in greater detail in **Section 1.6.3**. Activation of Akt has roles in maintaining hESC survival through inhibition of apoptosis [90]. Most commonly, recombinant fibroblast growth factor-2 (FGF2) is used to stimulate the PI3K/Akt pathway, but recent studies have found that a combination of recombinant insulin-like growth factor -1 (IGF-1) and heregulin can more effectively activate the PI3K/Akt pathway [100][34][134][33]. Loss of PI3K/Akt signaling, either by small molecule inhibition (i.e., LY294002) or removal of PI3K/Akt-stimulating growth factors, such as FGF2, from hESC culture medium, can induce mesendoderm differentiation [90][69].

Whereas PI3K/Akt signaling acts to inhibit differentiation, the TGFβ pathway drives pro-self-renewal gene transcription in hESCs. As described in **Section 1.2.2**, TGFβ

signaling morphogens (i.e., TGFβ1, ACTIVIN A and NODAL) bind to cognate type I and type II serine/threonine kinase receptors to induce SMAD2/3 phosphorylation. This in turn leads to transcriptional activation of self-renewal and pluripotency genes such as *NANOG*, an essential member of the core transcriptional regulatory network in human and mouse ESCs (**Figure 1.11**) [15][54][130][122][97][135][116][72][78]. Inactivation of the SMAD2/3 signaling cascade induces a loss of hESC pluripotency and differentiation towards ectoderm, highlighting the importance of the TGFβ signaling pathway for maintaining an undifferentiated state [116].

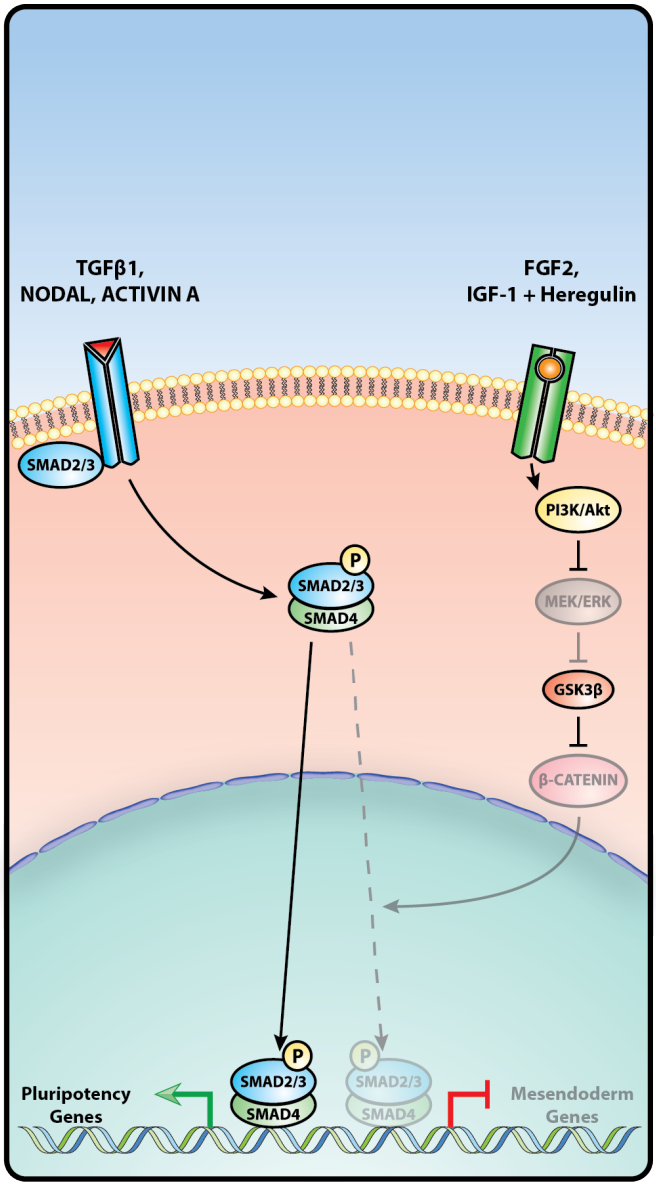


Figure 1.11: **Required hESC signaling pathways**. Human ESC self-renewal and pluripotency requires SMAD2/3 and PI3K/Alt signaling pathway activation. TGFβ1, NODAL and ACTIVIN A are TGFβ superfamily signaling morphogens capable of stimulating the phosphorylation of SMAD2/3. Once phosphorylated SMAD2/3 interacts co-SMAD4, translocates into the nucleus and activates pluripotency gene transcription. SMAD2/3 is also involved in mesendoderm differentiation in hESCs, but the PI3K/Akt signaling pathway (activated by FGF2 or IGF-1 + Heregulin) antagonizes this behvaior and prevents pro-differentiation SMAD2/3 events. PI3K/Akt maintains active GSK3β which degrades β-catenin and prevents the activation of mesendoderm genes.

1.5 The Core Transcriptional Regulatory Network

Downstream of the pro-pluripotency signaling pathways, lies the core transcriptional regulatory network: *OCT4*, *SOX2* and *NANOG* (**Figure 1.12**). In 2005, a landmark study showed that *OCT4*, *SOX2* and *NANOG* co-occupy the promoters of numerous (353) target genes and through collaborative regulation these transcription factors activate ESC genes and suppress differentiation genes [11]. More recently, a detailed study of how *OCT4*, *SOX2* and *NANOG* individually function in hESCs, found that each transcription factor uniquely represses formation of a particular germ lineage [124]. In addition, these three core transcription factors regulate their own, as well as each others, transcriptional expression setting up an auto-regulatory feedforward loop that sustains the pluripotent-state [11][23]. Together this triumvirate cooperatively regulates hESC pluripotency and self-renewal. Below, these transcription factors are discussed in greater detail.

OCT4 (also known as *POU5F1*) is a member of the POU (PIT/OCT/UNC) family of homeodomain proteins. In mice, *Oct4* is maternally provided in the oocyte and then transcriptionally expressed in the cells of the ICM (E3.5) and epiblast (E6.0). Through its POU DNA-binding domain, Oct4 recognizes and binds to the octamer sequence "ATGCAAAT" and recruits chromatin modifying complexes (NuRD and SWI/SNF) as well as transcription factors with known roles in ESC self-renewal (Sall4, Esrrb and Dax1) [87][120]. Importantly, OCT4 has the ability to homo- and hetero-dimerize with other transcription factors, such as SOX2, and co-regulate target genes. *Oct4* knockout (KO) mice fail to form a pluripotent ICM and die before implantation, reinforcing its' role in pluripotency. Recently, a study in hESCs determined that *OCT4* stabilizes the pluripotent state by repressing ectoderm differentiation [124].

SOX2 (short for Sex determining Region Y-box 2) is a member of the SOX-related HMG-box family of transcription factors and binds to the consensus DNA motif "ATTGTT".

Similar to Oct4, Sox2 interacts with chromatin modifying complexes, NuRD and SWI/SNF, as well as key transcription factors (Sall4, Essrb and Rex1) in mESCs to regulate target gene transcription [39]. In mice, *Sox2* expression is first detected at the morula stage (E2.5), becomes restricted to the ICM (E3.5) and then later is expressed in the epiblast (E6.0) [4]. *In vivo*, *Sox2*-/- mice fail to form an epiblast and in hESCs it represses mesoderm differentiation [124].

NANOG is a homeodomain family transcription factor that homo-dimerizes with itself and binds to the consensus sequence "TAAT(G/T)(G/T)" [76][51][80]. Sox2 has been shown to be a direct Nanog binding partner along with other mESC pluripotency transcription factors (e.g., Esrrb, Rex1 and Sall4) [38]. Similar to Oct4 and Sox2, Nanog also interacts with the NuRD chromatin modifying complex as well as various other complexes and proteins [38]. *In vivo* mouse studies show that *Nanog* is first expressed during the morula stage, becomes localized to the ICM of the blastocyst and then expressed throughout epiblast (E6.0) [19]. Along with *Oct4*, *Nanog* expression becomes restricted to pluripotent primordial germ cells (these cells eventually give rise to eggs and sperm) (E11.5-12.5) after the onset of gastrulation [45]. *Nanog* KO mice fail to develop beyond the blastocyst stage and in hESCs *NANOG* represses ectoderm differentiation [76][124].

Although these three transcription factors are widely regarded as the "master" regulators of mouse and human ESC pluripotency, it is likely that they are not the only transcription factors regulating pluripotency. The work described here suggests that the transcription factor *TCF7L1* should also be considered a critical regulator of the pluripotent state in hESCs.

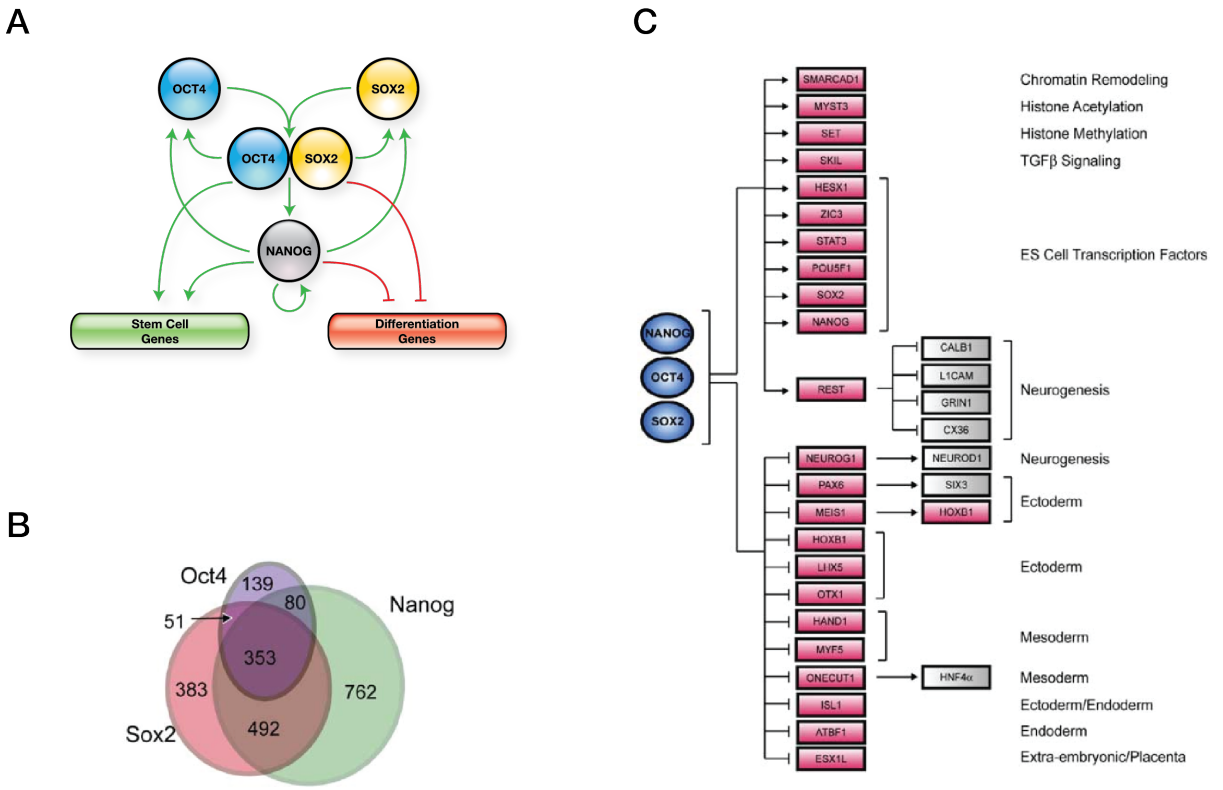


Figure 1.12: **The core transcriptional regulatory network**. **A**. *OCT4*, *SOX2* and *NANOG* comprise the core transcriptional regulatory network in hESCs and establish the molecular basis governing hESC pluripotency and self-renewal. Together they form a feedforward loop; meaning, they maintain each others expression as well as their own expression. Typically, OCT4 and SOX2 hetero-dimerize and co-regulate target genes such as *NANOG*. This transcriptional configuration helps ensure stabilization of the pluripotent-state transcriptome. **B**. Venn diagram of OCT4, SOX2 and NANOG ChIP-Chip data illustrating overlapping genomic binding sites in hESCs. **C**. *OCT4*, *SOX2* and *NANOG* collectively drive the expression of pluripotency and self-renewal genes, while blocking the expression of lineage-specific differentiation genes. *Figures **B** and **C** taken from [11] with permission.*

1.6 Human ESCs verses Mouse ESCs

Human and mouse ESCs are two mammalian models systems useful for studying pluripotency, but they are not equivalent. Human ESCs are more similar to later-stage post-implantation "primed" epiblast cells, whereas mESCs are representative of earlier pre-implantation "naive" ICM cells [35][113][110][85][88]. Differences in their respective developmental stages is evidenced by the different mechanisms governing their self-renewing pluripotent-states *in vivo* and *in vitro* [110]. For example, mESCs require LIF (leukemia inhibitory factor) to maintain pluripotency, whereas hESCs are not responsive to LIF. Also, Bmp4 and canonical Wnt signaling support mESC pluripotency, but conversely induce hESC differentiation [14][116]. Because mESCs and hESCs represent two unique developmental stages, they can be used to investigate the mechanisms regulating different cell type transitions during embryogenesis; for example, "naive" ICM-to-epiblast and "poised" epiblast-to-germ lineage fate decisions.

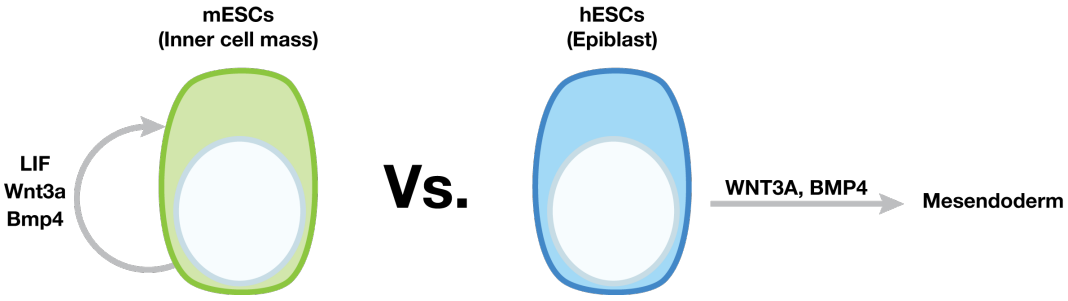


Figure 1.13: **Human and mouse ESC comparison**. Human and mouse ESCs represent different stages of development and have different growth factor responses. Mouse ESCs (green) represent "naive" pre-implantation ICM cells and their pluripotency is supported by LEF, Wnt3 and Bmp4. Human ESCs (blue) represent later-staged "poised" epiblast cells and undergo mesendoderm differentiation in response to WNT3A and BMP4 stimulation.

Chapter 2

Characterization of *TCF7L1* Expression in hESCs

2.1 Introduction

In this chapter, I describe studies in which I investigate LEF/TCF expression in hESCs. As downstream transcription factor effectors of the canonical WNT signaling pathway, I hypothesized that these transcription factors would perform an important function in the regulation of hESC pluripotency and differentiation. However, proper insights into their function largely depends on the activity of the canonical WNT signaling pathway, which was originally controversial. Two studies, published within close succession of each other, proposed opposing roles of the pathway in hESCs: Sato et al. concluded that the canonical WNT pathway supported pluripotency, whereas Dravid et al. found that activation of the canonical WNT pathway promoted hESC differentiation [96][31]. Considering the significance of this pathway in development, it is not only necessary to characterize LEF/TCF expression in hESCs, but also necessary to clarify the influence of the canonical WNT

signaling pathway on hESCs.

Here, I show that *TCF7L1* is the most highly expressed LEF/TCF in hESCs. I also find that hESCs are in a WNT-inactive state which suggests that *TCF7L1* behaves as a transcriptional repressor of WNT target genes. Lastly, I identify potential signaling pathway regulators of *TCF7L1* transcription in hESCs.

2.2 Results

2.2.1 The Canonical WNT Pathway is Inactive in hESCs

Initially, I investigated the activity of the canonical WNT signaling pathway in hESCs. Because sub-cellular β-catenin localization is indicative of canonical WNT signaling activity, immunofluorescence (IF) microscopy was used to determine β-catenin protein localization. In undifferentiated hESCs, β-catenin was located at the plasma membrane, but absent from the nucleus (**Figure 2.1A**). When hESCs were treated with the canonical WNT ligand WNT3A for 24 hours, the levels of nuclear localized β-catenin became more abundant (**Figure 2.1A**). Next, using a canonical WNT signaling reporter construct (TOPflash), the levels of endogenous WNT signaling was measured in hESCs. The TOPflash reporter contains multiple LEF/TCF DNA binding sites upstream of a minimal promoter driving bioluminescent luciferase expression. As a negative control, a mutant form of this construct (FOPflash) containing mutated LEF/TCF binding sites, was used to establish background levels of luciferase expression. If canonical WNT signaling is active, LEF/TCFs will recruit nuclear stabilized β-catenin and activate luciferase transcription resulting in a large TOPflash-to-FOPflash ratio. However, if the pathway is inactive luciferase will not be expressed and the TOPflash-to-FOPflash ratio will close to "1". The data I generated shows that hESCs have a TOPflash-to-FOPflash ratio of "1" under normal growth

conditions hESCs, yet are also highly responsive to canonical WNT stimulation when treated with WNT3A (**Figure 2.1B**). Furthermore, treatment of hESCs with WNT3A for 48 hours strongly induced the expression of the mesendoderm genes *T*, *MIXL1* and *FOXA2* (**Figure 2.1C**). Additionally, WNT ligand secretion was blocked to test whether any WNT ligands, canonical or non-canonical, were necessary for maintaining hESC pluripotency in culture. Treating the hESCs with IWP-2 (an inhibitor that prevents WNT ligand secretion from the Golgi) for 7 days had no effect on OCT4, SOX2 and NANOG expression and no effect on colony morphology (**Figure 2.2**) [20]. This result indicates that hESCs do require canonical or non-canonical autocrine WNT signaling for stabilization of the pluripotent state. This data lead to the conclusion that canonical WNT signaling does not support hESC pluripotency rather it induces hESC differentiation.

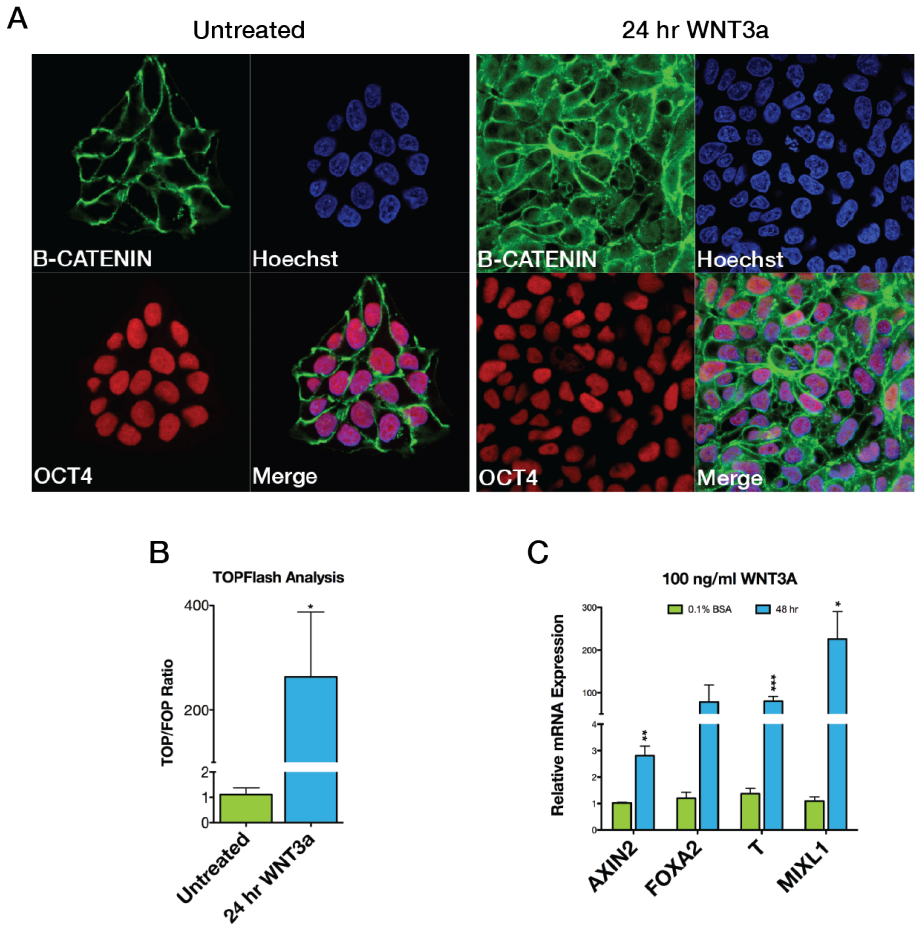


Figure 2.1: **Human ESCs are in a WNT-inactive state**. **A**. Immunofluorescence confocal microscopy analysis of β-catenin (green) and OCT4 (red) localization in hESCs. Under normal growth conditions (left panel: mTeSR1 medium) β-catenin is localized at the plasma membrane and is not present in the nucleus. When treated with WNT3A (100 ng/ml) for 24 hours (right panel), β-catenin levels increase in the nucleus and overlap with OCT4 indicating hESC are capable of responding to canonical WNT ligands. Nuclei were stained with Hoechst (blue). **B**. Transfected hESCs displayed equal levels of luciferase expression from a TOPFlash construct and the negative control FOPFlash construct. Stimulation of hESCs with WNT3A (100 ng/ml) for 24 hours induced high level of luciferase expression from TOPFlash transfected hESCs as compared to FOPFlash transfected hESCs. Data is for each condition is presented as the ratio of TOPflash luciferase activity to FOPFlash luciferase activity (n=3). **C**. Canonical WNT activation induces mesendoderm gene expression. Primitive streak and mesendoderm genes were analyzed by qPCR after 48 hours of WNT3A (100 ng/ml) treatment under feeder-free conditions (n=3). *AXIN2* is a classic WNT target gene and was used as a positive control. Statistical analysis was performed by Student's t-test and presented as mean +/-SEM (*p <0.05, **p <0.01, ***p <0.001).

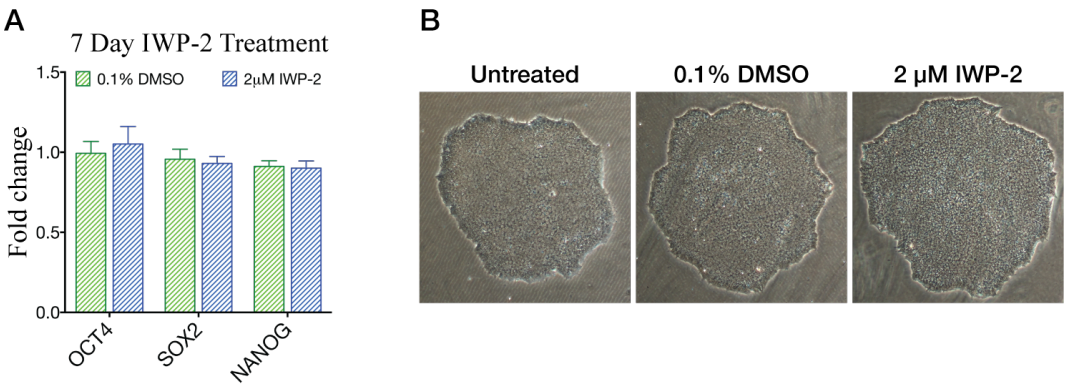


Figure 2.2: **Autocrine WNT signaling is not required for hESC maintenance**. **A**. Quantitative PCR analysis of pluripotent marker genes - *OCT4*, *SOX2* and *NANOG* - was performed 7 days after treating hESCs, under feeder-free conditions, with IWP-2 (2 μM) (n=3). Treatment had no affect as compared to DMSO vehicle control. **B**. Phase contrast images (10X) were taken after 7 days of IWP-2 treatment under feeder-free conditions. In all three conditions hESCs were morphologically indistinguishable from each other.

2.2.2 LEF/TCF Expression Analysis

LEF/TCF transcription factors mediate transcriptional activation and repression of WNT target genes depending on the activity of the pathway. To address the function of LEF/TCFs in hESCs, their mRNA and protein expression levels were analyzed. Quantitative PCR (qPCR) analysis using primers sets with 100% efficiencies for amplification, indicated that *TCF7L1* is the most highly expressed LEF/TCF family member in hESCs (**Figure 2.3**). Additionally, *TCF7L1* has similar mRNA expression levels in three other hESC-lines: H1, H9 and H14 (**Figure 2.3**). We also assessed LEF/TCF protein levels using SDS-PAGE and observed a strong correlation between mRNA and protein levels, further supporting the notion that *TCF7L1* is the most highly expressed LEF/TCF family member in hESCs (**Figure 2.3C**). Lastly, IF staining of TCF7L1 and the pluripotency marker OCT4, verified TCF7L1 expression and nuclear localization in hESCs (**Figure 2.3E**).

Our data corroborate the emerging view that hESC pluripotency depends on WNT pathway

inactivity and that WNT signaling triggers hESC differentiation. Our data also suggest that *TCF7L1* acts as a repressor in the WNT-inactive state, as it does in other cell types including mESCs [53][73][101][131]. I therefore next investigated how *TCF7L1* expression is regulated in hESCs.

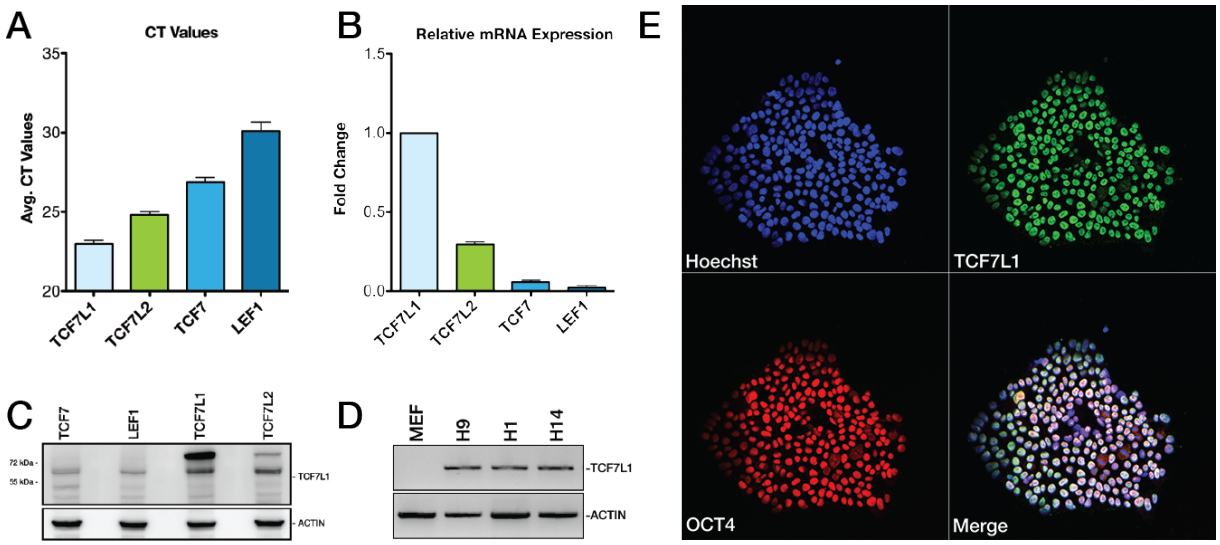


Figure 2.3: ***TCF7L1* is the most highly expressed LEF/TCF in hESCs**. **A**. LEF/TCF qPCR analysis in H9 hESCs showing TCF7L1 mRNA is the most abundant. Data presented as raw, averaged Ct-values (n=3). **B**. $\Delta\Delta$Ct-based analysis of LEF/TCF qPCR data relative to *TCF7L1* (n=3). **C**. Western blot analysis of LEF/TCF protein levels in H9 hESCs showing TCF7L1 is the most abundant protein. Blots were probed with individual antibodies separately, but imaged simultaneously. *β-ACTIN* was used as a loading control. **D**. Reverse transcription PCR (PCR) analysis of *TCF7L1* mRNA expression in H1, H9 and H14 hESCs. The MEF-only sample ensures species-specificity of the *TCF7L1* primers. *β-ACTIN* was used as a template loading control. **E**. Confocal IF analysis of OCT4 (red) and TCF7L1 (green) proteins in H9 hESCs under feeder-free conditions. Nuclei were stained with Hoechst. Merge is overlay of the three channels.

2.2.3 Analysis of TGFβ and PI3K/Akt Regulation of *TCF7L1*

To understand how *TCF7L1* expression is regulated in hESCs, I asked whether the pro-pluripotency signaling pathways TGFβ and PI3K/Akt regulate its' expression. FGF2-mediated PI3K/Akt signaling was inhibited using the small molecule inhibitor SU5402 (a fibroblast growth factor receptor (FGFR)-specific tyrosine kinase inhibitor). After twenty-four and forty-eight hours of inhibition with SU5402, there was no effect on *TCF7L1* mRNA levels (**Figure 2.4A**) [34]. In the same conditions, *OCT4* and *NANOG* were slightly down-regulated, whereas *SOX2* (a well known ectoderm marker) was modestly up-regulated [40]. In contrast, inhibition of TGFβ signaling (using the small molecule inhibitor SB431542, a type I ALK4/5/7 receptor kinase inhibitor) for 24 hours significantly down-regulated *TCF7L1* mRNA expression by nearly 50% (**Figure 2.4B**). As expected, *NANOG*, an established downstream target of the TGFβ pathway, was robustly down-regulated (**Figure 2.4B**) [135]. These results strongly suggest that *TCF7L1* is downstream of the TGFβ pathway, but not the PI3K/Akt pathway.

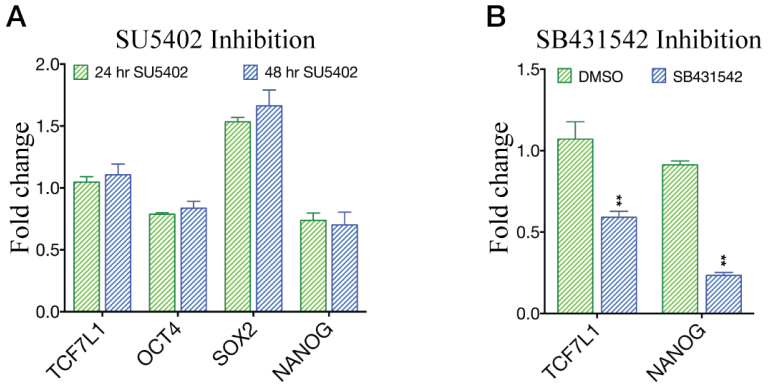


Figure 2.4: **TGFβ signaling activates *TCF7L1* expression**. **A**. *TCF7L1*, *OCT4*, *SOX2* and *NANOG* mRNA expression were assayed by qPCR following 24 and 48 hours of SU5402 (10 μM) treatment under feeder-free conditions (n=3). There was a slight down-regulation of *OCT4* and *NANOG*, no effect on *TCF7L1* and slight up-regulation of *SOX2*. **B**. *TCF7L1* and *NANOG* mRNA expression were down-regulated following 24 hours of SB431542 (10 μM) treatment under feeder-free conditions (n=3). Statistical analysis was performed by Student's t-test and presented as mean +/-SEM (*p <0.05, **p <0.01).

2.2.4 *TCF7L1* is Regulated by the Core Transcriptional Regulatory Network

The core transcriptional regulatory factors *OCT4*, *SOX2* and *NANOG* are central to hESC pluripotency. As previously described, these three transcription factors cooperatively maintain the pluripotency and self-renewal gene programs in hESCs, while preventing the expression of differentiation programs [11]. To test if these three transcription factors regulated *TCF7L1*, I used siRNA to knockdown *OCT4*, *SOX2* and *NANOG* then measured *TCF7L1* mRNA by qPCR. Of the three transcription factors tested, *OCT4* knockdown had the strongest affect, leading to a 50% reduction in *TCF7L1* mRNA levels (**Figure 2.5A**). *NANOG* had a significant, but much less substantial effect on *TCF7L1* mRNA expression: its expression was reduced by about 20% (**Figure 2.5A**). Unfortunately, the *SOX2* siRNA knockdown was ineffective and there was no observed effect on *TCF7L1* mRNA expression. This result could be attributed to either insufficient knockdown of *SOX2*, absence of regulatory control over *TCF7L1* expression or possible compensation from another member of the SOX transcription factor family such as *SOX3* [124]. I interpreted these data to suggest that *TCF7L1* is dependent on *OCT4*, and to a lesser extent *NANOG*, for expression in hESCs.

To further investigate the connection between the core transcriptional regulatory network and *TCF7L1*, published OCT4, SOX2 and NANOG chromatin immunoprecipitation sequencing (ChIP-seq) datasets from mouse and human ESCs were investigated to look for putative binding sites near the *TCF7L1* locus. Data from both species identified OCT4, SOX2 and NANOG binding sites at multiple locations across the *TCF7L1* locus. In particular, all three transcription factors co-bind a conserved region approximately 1.3 kb downstream of the *TCF7L1* transcription start site (**Figure 2.5B**)[11][65][22][79][15][60]. This region was particularly interesting because of its proximity to the *TCF7L1* promoter. Using ChIP-qPCR, I confirmed that OCT4, SOX2 and NANOG bound to this site in H9 hESCs

indicating that *TCF7L1* is directly targeted by these master pluripotency transcription factors (**Figure 2.5C**). These data are consistent with the report by Boyer et al. that *TCF7L1* is one of thirty eight actively expressed transcription factors co-bound by OCT4, SOX2 and NANOG in hESCs [11]. These results strongly suggest that *TCF7L1* is driven by the pro-pluripotency network and also likely to be involved in the regulation of hESC pluripotency.

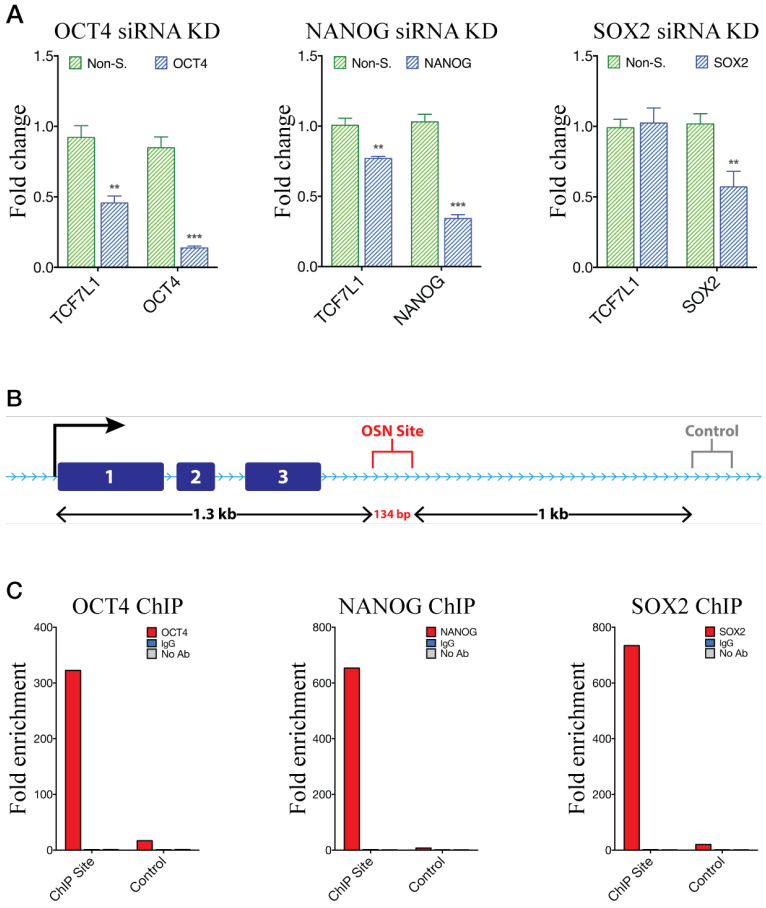


Figure 2.5: **Analysis of *TCF7L1* regulation by the core regulatory network**. **A**. *OCT4*, *SOX2* and *NANOG* siRNA knockdowns were analyzed by qPCR for *TCF7L1* mRNA expression 3 days post-knockdown. *OCT4* and *NANOG* siRNA knockdowns significantly down-regulated *TCF7L1* mRNA levels, but *SOX2* knockdown did not change TCF7L1 expression. Non-silencing (Non-S.), *OCT4*, *SOX2* and *NANOG* siRNAs were used at 100 nM. Results are shown as fold change relative to an untreated control (n=3). **B**. Diagram of the conserved OCT4, SOX2 and NANOG binding sites identified in human and mouse ESC ChIP-seq datasets. Exons are marked as blue boxes. The region where OCT4, SOX2 and NANOG co-bind the *TCF7L1* locus (OSN site; red) is 1.3 kb downstream of the start codon. A control primer set (Control; grey) used for ChIP-qPCR analysis is indicated 1 kb downstream of the "OSN Site". **C**. ChIP-qPCR analysis of OCT4, SOX2 and NANOG confirmed direct binding to the *TCF7L1* loci in hESCs. No-antibody and IgG pre-immune serum were used as controls. Data is presented as fold change in enrichment relative to the No-antibody condition. Statistical analysis was performed by Student's t-test and presented as mean +/-SEM (*p <0.05, **p <0.01, ***p <0.001).

2.2.5 *TCF7L1* Response to hESC Differentiation

To test if *TCF7L1* expression was specific to undifferentiated hESCs, its' expression was analyzed during hESC differentiation. First, hESCs were subjected to BMP4-induced differentiation which induces the cells to differentiate towards mesendoderm *in vitro* [8][124]. Human ESC cultures were treated with BMP4 (100 ng/ml) for 24 and 48 hours under feeder-free conditions [133]. Quantitative PCR analysis indicated that *TCF7L1* mRNA was robustly down-regulated such that by 24 and 48 hours mRNA levels had declined by 44% and 77% respectively (**Figure 2.6A**). Like *TCF7L1*, *SOX2* and *NANOG* were also down-regulated by BMP4, but there was no effect on *OCT4* (**Figure 2.6A**). Concomitantly, BMP4 treatment robustly down-regulated TCF7L1 protein levels within the first 12 hours (**Figure 2.6B**). Based on these results, I postulated that *TCF7L1* could be a downstream BMP4 signaling pathway target.

Because the canonical WNT pathway also induces mesendoderm differentiation in hESCs, I tested how canonical WNT signaling affected *TCF7L1* expression [29][100]. Human ESCS were treated with recombinant WNT3A (100 ng/ml) under feeder-free conditions for 24 and 48 hours then mRNA expression levels were analyzed by qPCR. The data showed a clear up-regulation of primitive streak/mesendoderm marker genes *BRACHYURY* (*T*), *MIXL1* and *FOXA2*, but there was no effect on *TCF7L1* mRNA expression (**Figure 2.6C**). Thus in hESCs, *TCF7L1* expression is responsive to BMP4-induced mesendoderm differentiation, but not WNT-induced mesendoderm differentiation, alluding to a potential cross-regulatory mechanism between these two signaling pathways.

Human ESCs were also directed to differentiate towards an ectodermal fate using retinoic acid (RA), a small lipophilic molecule derived from vitamin A [94][5]. Treatment of hESCs with RA (5 μM) for 12, 24 and 48 hours did not have any significant affect on *TCF7L1*, *OCT4* and *NANOG* mRNA, but slightly up-regulated *SOX2* ($\sim$1.6 fold) (**Figure 2.6D**).

Therefore, I do not consider *TCF7L1* to be a downstream target of the RA pathway.

In addition to the direct differentiation methods, *TCF7L1* expression was analyzed during embryoid body differentiation. As explained in the previous chapter, the embryoid body assay is thought to recapitulate early gastrulation differentiation events *in vitro*. This assay is useful for providing insights into whether a particular gene is uniquely expressed only in pluripotent hESCs, or if it might be expressed in subsequent differentiated cell types [27][126]. For example, *OCT4* and *NANOG* are specifically expressed in pluripotent hESCs and therefore permanently down-regulated during this assay; *SOX2*, on the other hand, is expressed in the ectoderm lineage, thus its' expression is maintained during this assay. For this experiment, embryoid bodies were grown and harvested at 2, 7 and 14 days then analyzed by qPCR for gene expression changes. Interestingly, *TCF7L1* expression was slightly up-regulated (∼2 fold) throughout the 14 day time course; *SOX2* expression was steadily maintained over the first 7 days, then sharply up-regulated (∼8 fold) at day 14 (**Figure 2.6E**). *OCT4* and *NANOG* were strongly down-regulated due to their specific expression in pluripotent hESCs (**Figure 2.6E**). Marker genes specific to endoderm, mesoderm and ectoderm were up-regulated as expected (data not shown). The results of this experiment indicated that *TCF7L1* expression is not specific to pluripotent hESCs, but that it is also expressed in the differentiated progeny of hESCs.

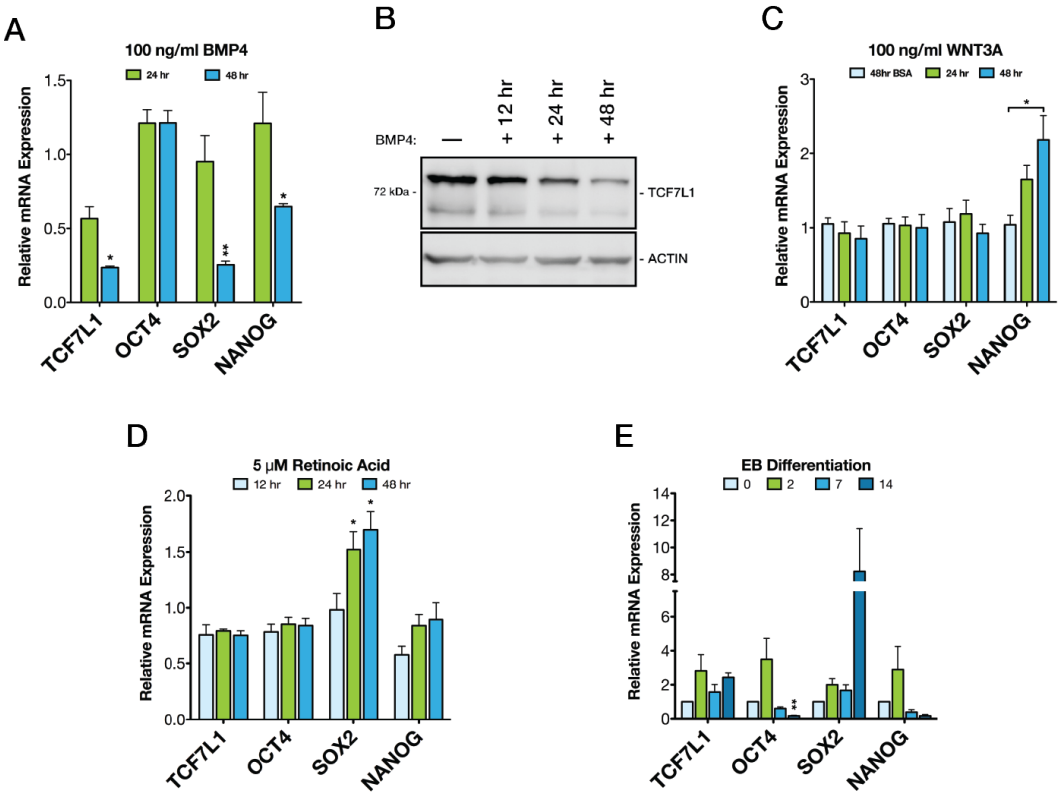


Figure 2.6: ***TCF7L1* analysis during hESC differentiation**. **A**. Quantitative PCR (n=3) analyses show steady reduction of *TCF7L1* and *SOX2* over 12, 24 and 48 hours of BMP4 (100 ng/ml) treatment. **B**. Western blot analysis shows that BMP4 (100 ng/ml) down-regulates TCF7L1 protein levels over 12, 24 and 48 hours of treatment. **C**. Quantitative PCR analysis of *TCF7L1* mRNA after 48 hours of recombinant WNT3A (100 ng/ml) exposure shows maintained expression levels (n=3). *NANOG* was slightly up-regulated in these conditions. **D**. RA (5 μM) does not affect *TCF7L1* expression as analyzed by qPCR. SOX2 was slightly up-regulated. **E**. *TCF7L1* mRNA expression levels were slightly up-regulated during the 14 day embryoid body differentiation assay (n=3). *SOX2* was slightly up-regulated by RA treatment. Statistical analysis was performed by Student's t-test and presented as mean +/-SEM (*p <0.05, **p <0.01).

2.3 Discussion

The work presented here identifies *TCF7L1* as the predominant LEF/TCF in hESCs and takes the initial steps towards understanding the underlying regulatory mechanisms controlling its expression. I also determined that the canonical WNT pathway is inactive in hESCs, data which is supported by more recent studies [104][30][9][10]. Inactivity of this pathway and the absence of β-catenin from the nucleus, implies that *TCF7L1* is likely acting as a transcriptional repressor in hESCs.

I found that the *TCF7L1* is downstream of a TGFβ1/ACTIVIN A/NODAL-driven SMAD2/3 signaling pathway, but not of an FGF-mediated signaling pathway. Inhibition of the TGFβ1 pathway down-regulates *TCF7L1* expression, along with the well known downstream target, *NANOG* [135]. Our gene expression analyses are consistent with published SMAD2/3 ChIP-seq data in hESCs indicating that SMAD2/3, in fact, binds at numerous locations across the *TCF7L1* genomic locus, supporting the notion that *TCF7L1* is a downstream target of this pathway [22][15][79]. Interestingly, ChIP-seq results show that SMAD2/3 binds to the same region of *TCF7L1* that is co-bound by OCT4, SOX2 and NANOG (**Figure 2.5B**) suggesting SMAD2/3 might be directly involved in regulating *TCF7L1* expression. The relationship between the SMAD2/3 signaling and *TCF7L1* transcription supports the idea that its' expression is involved in maintaining the undifferentiated state of hESCs.

I also determined that *TCF7L1* is downstream of the core transcriptional regulatory network. More specifically, the data shows that *OCT4* acts as a strong transcriptional activator of *TCF7L1* expression. *NANOG* had a significant, but minor regulatory influence over *TCF7L1*, which could be due to the inadequate knockdown efficiency. It is equally possible that sustained *OCT4* expression (data not shown) was capable of supporting *TCF7L1* expression during the *NANOG* knockdown and therefore masked its effects. Unfortunately,

the *SOX2* siRNA knockdown was not effective, thus, I cannot make any conclusions about its regulatory influence over *TCF7L1* expression. It is likely that this was the result of inadequate SOX2-targeting siRNAs as SOX2 was only reduced by ∼50% and will need to be optimized for any future experiments. After the completion of this experiment, a group published a *SOX2* shRNA knockdown microarray dataset in hESCs that indicated *TCF7L1* mRNA expression was not affected by the loss of *SOX2* [124]. Nevertheless, published ChIP-seq and our own ChIP-qPCR results indicated that all three master transcription factors physically bind to the *TCF7L1* locus; hence, I propose that *TCF7L1* is a core transcriptional regulatory network target gene, which is supported by OCT4, SOX2 and NANOG ChIP-chip data published by Boyer and colleagues [11]. It would be interesting to perform a simultaneous *OCT4*, *SOX2* and *NANOG* knockdown to test for synergistic co-regulation *TCF7L1* expression. In the future, promoter-reporter mutants could be generated to assay for the functionality of these binding sites. Because the TGFβ1 signaling pathway supports *TCF7L1* expression, the core regulatory network factors bind to the *TCF7L1* locus and OCT4 and NANOG positively regulate *TCF7L1*'s expression, I propose that *TCF7L1* expression is driven by pro-pluripotency mechanism in hESCs.

To characterize *TCF7L1* expression in hESCs, I asked how its expression was affected by hESC differentiation. To test this, I used a combination of directed and spontaneous differentiation assays and discovered that: *TCF7L1* does not ubiquitously respond to differentiation, rather it seems to be specifically targeted for down-regulation by BMP4 - a known inducer of mesendoderm differentiation *in vitro* and primitive streak formation *in vivo*. This is particularly interesting because, like SMAD2/3, BMP4-activated SMAD1/5/8 also binds with co-SMAD4 for translocation into the nucleus. Mouse ESC ChIP-seq data for Smad1 shows binding to the same conserved Oct4, Sox2 and Nanog co-bound site in the *Tcf7l1* locus. I propose three possible mechanisms of BMP4-induced suppression of *TCF7L1*: i) direct SMAD1/5/8-mediated repression of *TCF7L1*, ii) sequestration of co-SMAD4 away from SMAD2/3 (and therefore its target genes) and redirected to SMAD1/5/8 target genes;

or iii) by some indirect regulatory mechanism. More in-depth analysis must be carried out to uncover the mechanism of BMP4-directed down-regulation of *TCF7L1*, which could have exciting implications for the regulation of early development.

Despite strong evidence supporting canonical WNT signaling as an inducer of hESC differentiation, controversy still remains. Using clonal hESC lines carrying the TOPflash reporter driving GFP, Blauwkamp and colleagues recently found that cultured hESCs show heterogeneous levels of canonical WNT activity [9]. Interestingly, when they isolated WNT-high and WNT-low expressing single cells by fluorescence-activated cell sorting (FACS), they discovered that the WNT-high cell population showed an increased survival and formed more OCT4-pluripotent colonies when re-cultured under pluripotent conditions as compared to the WNT-low cells. On the surface, this result contradicts my prediction about canonical WNT signaling inducing hESC differentiation. However, I argue that in culture hESCs are likely to experience heterogeneous levels of canonical WNT signaling, but that low levels of endogenous WNT signaling are permissible to the pluripotent state; it is only when a certain threshold of WNT signaling is achieved that hESCs fully commit to differentiation. In addition, pluripotency-stabilizing conditions are likely to prevent autocrine WNT signaling from reaching this threshold and reestablish an optimal pluripotent state. Furthermore, Blauwkamp et al.'s gene expression and differentiation assays support the rationale that canonical WNT signaling stimulates differentiation. A comparison of WNT-high and WNT-low sorted cells indicated that the WNT-high cells expressed tenfold higher levels of primitive streak marker genes (*BRACHYURY*, *MIXL1* and *GSC*) than the WNT-low population. Also, the WNT-high hESCs cells were much more efficient at mesoderm and endoderm differentiation than the WNT-low cells, which were more efficient at ectoderm differentiation, again implying canonical WNT signaling directs differentiation towards mesendoderm lineages. While Blauwkamp's data irrefutably shows endogenous canonical WNT signaling occurs within pluripotent hESCs in culture, I believe that low-level fluctuations in endogenous canonical WNT activity demonstrate the precarious nature

of *in vitro* cultured hESCs, but I contest that these fluctuations are dispensable for hESC pluripotency.

In all, my data suggests that *TCF7L1* is the LEF/TCF responsible for regulating WNT target genes in hESCs and likely acting as a transcriptional repressor. As such, I propose a model where pro-pluripotency mechanisms drive *TCF7L1* expression to ensure repression of WNT target genes (**Figure 2.7**). In ensuing chapters, I will investigate *TCF7L1* target genes and their relationship to the hESC pluripotency.

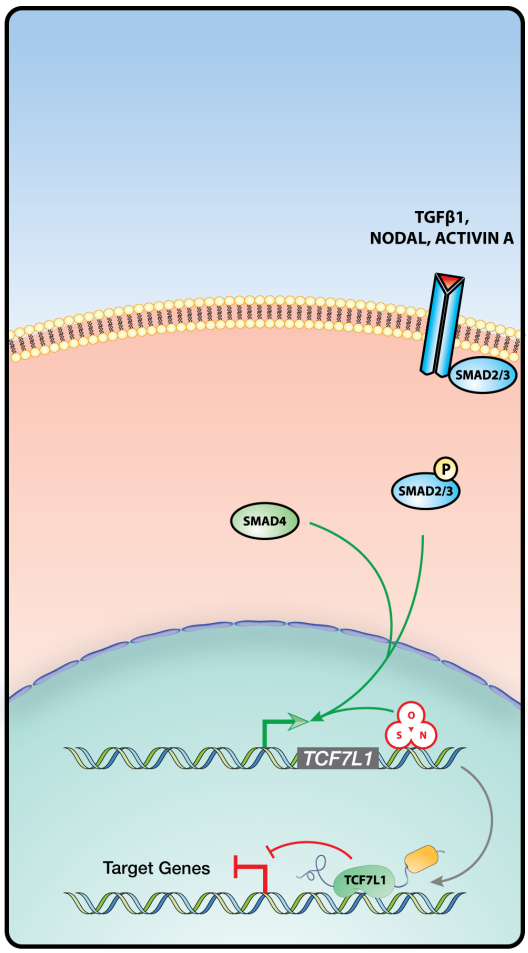


Figure 2.7: **Model of *TCF7L1* regulation in hESCs**. Illustration of the mechanisms governing *TCF7L1* expression in hESCs. In the pluripotent state, SMAD2/3-mediated signaling and the core transcriptional regulatory network (OCT4:O; SOX2:S; NANOG:N), particularly OCT4, support the transcriptional activation of *TCF7L1*. In turn, *TCF7L1* likely aids in the maintenance of hESC pluripotency by repressing its' target genes.

Chapter 3

Genome-wide Analysis of *TCF7L1* Target Genes

3.1 Introduction

Transcription factors regulate expression of their gene targets by facilitating recruitment of transcriptional activation or repression machinery. Collectively, transcription factors work in concert with each other to orchestrate proper gene expression patterns required for specific developmental states and cell lineages during embryogenesis. As such, identifying the genome-wide binding location of these factor's allows one to make informed predictions of putative gene targets and ultimately provides insights into its' function. Performing a large-scale genome-wide analysis like this is challenging, however, and requires coupling chromatin immunoprecipitation with massively parallel sequencing (ChIP-seq) [37]. If done properly under the appropriate conditions, ChIP-seq can identify all the binding sites of a specific transcription factor across the genome in a particular cell type with high resolution.

In this chapter, I describe studies in which I use ChIP-seq to survey *TCF7L1* binding sites

(and by extension target genes) in pluripotent hESCs. Specifically, I wanted to know whether *TCF7L1* targets pluripotency genes or genes involved in differentiation. I hypothesized that data from this experiment would enable me to develop a more focused hypothesis of *TCF7L1* function in hESCs.

3.2 Results

3.2.1 ChIP-seq Optimization

Reliable ChIP-seq data is highly dependent upon the quality of the antibody: it must be able to specifically recognize and immunoprecipitate the protein of interest. Thus, identifying a suitable antibody was critical for this experiment. Fortunately, previous ChIP-chip and ChIP-seq studies in mESCs had identified a suitable ChIP-grade antibody capable of recognizing and immunoprecipitating human TCF7L1 (**Figure 3.1**) [26][65].

Initial attempts at TCF7L1 ChIP-seq were unsuccessful because the necessary amount of DNA (10 ng) required for sequencing was not recovered after immunoprecipitation. All procedural parameters were carefully tested: buffer compositions, incubation times, wash stringencies, cross-linking methodologies and various sonication methods (Bioruptor, Covaris, QSonica with micro tip probe), including micrococcal nuclease (MNase) digestion to enzymatically fragment the chromatin. This latter modification was the most important, enabling recovery of enough DNA for ChIP-seq (∼40 - 60 ng). Interestingly, every aspect of the MNase procedure was identical to the sonication ChIP protocol except for this single enzymatic step; leading us to conclude that sonication was adversely affecting the ChIP procedure. Figure 3.2 compares the enrichment of TCF7L1 at the *NODAL* enhancer using the MNase or sonication methods of DNA cleavage and shows that MNase fragmentation resulted in a 10-fold greater enrichment of TCF7L1 (**Figure 3.2**). It is possible that the

harsh condition of sonication is detrimental to the protein-antibody interaction [103]. Based on these results, ChIP-seq analysis was performed using the MNase fragmentation method.

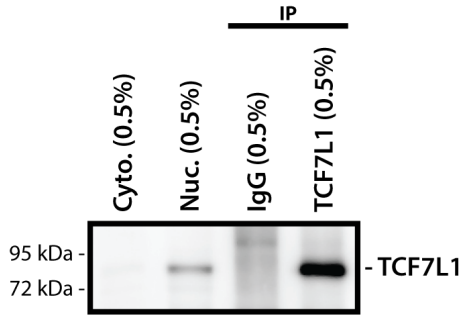


Figure 3.1: **Identification of ChIP-grade TCF7L1 antibody**. Cellular extracts of hESCs were fractionated into cytoplasmic and nuclear sub-fractions. The nuclear fraction was used for TCF7L1 immunoprecipitation using the Santa Cruz M-20 antibody (goat). Equal percentages (0.5%) of the lysates were analyzed by Western blot to test the effectiveness of the immunoprecipitation. IgG sera was used as a negative control. The blot was probed with a TCF7L1 antibody from a different species (rabbit) and the results show enrichment of TCF7L1 after immunoprecipitation.

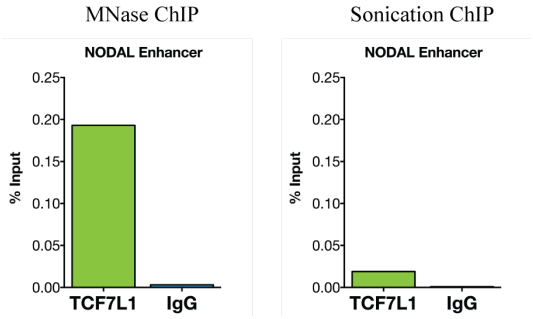


Figure 3.2: **MNase fragmentation is the most effective TCF7L1 ChIP method**. ChIP-qPCR was used to assess the efficiency of TCF7L1 immunoprecipitation using the sonication or MNase fragmentation methods. MNase fragmentation increased TCF7L1 enrichment at a known binding site in the NODAL enhancer by 10 fold. For both methods, 20 million cells were used for the TCF7L1 ChIP and 20 million cells were used for the IgG ChIP; 5 μg of IgG and TCF7L1 antisera were used for each ChIP.

3.2.2 ChIP-seq Analysis of TCF7L1 Binding Locations

Human ESCs were grown under feeder-free conditions (to avoid MEF contamination) and TCF7L1 immunoprecipitated DNA from H9 hESCs was sequenced using the Illumina HiSeq 2500 platform. Two biological ChIP replicates were analyzed by single-end (50 bp) sequencing. Approximately 54 million reads were mapped for each replicate (53,797,735 & 54,273,005) and there was a high degree of overlap between both samples (**Figure 3.3**). An input, non-immunoprecipitated sample was sequenced as a control for the ChIP-seq peak calling software MACS2. High-confidence TCF7L1 peaks were determined by irreproducibility discovery rate (IDR) analysis - a quality control method advocated by the ENCODE (Encyclopedia of DNA Elements) project - which reports significant peaks by measuring the consistency between the replicates [58]. IDR analysis rests on the assumption that true ChIP-seq peaks should have a similar rank order between the two replicates and false-positive, yet reproducible ChIP-seq peaks will have different rank orders between the two replicates. Thus, the IDR does not rely on simple reproducibility of ChIP-seq peaks, but instead requires the reproducible peaks to have a similar rank orders between the two replicates. A peak with a very high IDR (akin to a high false discovery rate) has a different rank order strength between the two replicates and a peak with a low IDR has a similar rank order strength between the replicates. Using an IDR cutoff of 0.013249, we identified 9,760 significant TCF7L1 peaks across the genome of undifferentiated hESCs (**Figure 3.3**).

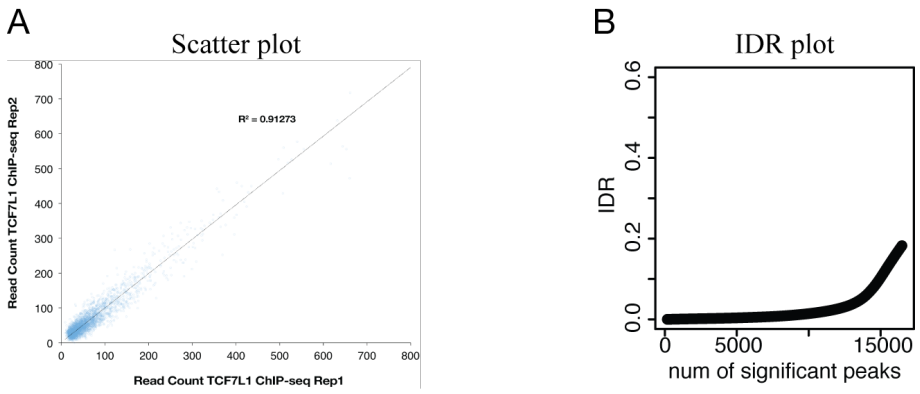


Figure 3.3: **Analysis of TCF7L1 ChIP-seq replicates**. **A**. Scatter plot illustrating the high degree of reproducibility of the two TCF7L1 ChIP-seq replicates. The number of reads that mapped to the top 3518 peaks (IDR <0.002406) were determined for each TCF7L1 ChIP-seq biological replicate. The replicate with a greater number of mapped reads was scaled to the replicate with less mapped reads. **B**. IDR plot showing a high degree of reproducibility in the two ChIP-seq biological replicates. The IDR is plotted as an increasing number of reproducible ChIP-seq peaks are found. IDR analysis reveals the appropriate cutoff for the number of ChIP-seq peaks to be included in all subsequent analyses. An IDR cutoff of 0.013249 (top 9760) were used for the analysis, which is well within the range suggested by ENCODE.

Next, we analyzed our ChIP-seq data using Regulatory Sequence Analysis Tools (RSAT) and HOMER (Hypergeometric Optimization of Motif EnRichment) next generation sequence software for motif analysis as well as peak annotation [46][112]. As expected, the canonical LEF/TCF DNA-binding sequence "CTTTGA" was enriched in the TCF7L1 peaks as the top scoring motif (**Figure 3.4**). Annotation of the peaks indicated that 41% (4,021) of TCF7L1 binding sites were within +/-5 kb of annotated RefSeq transcription start sites (TSS), 22% (2,133) were located in genic regions and 37% (3,606) occurred in distal intergenic regions (**Figure 3.4**). Figure 3.4 shows the overlapping enrichment of both replicates +/-5 kb from the TSS relative to the input (**Figure 3.4**). These results demonstrated the quality of our ChIP-seq data and indicated a prevalence of TCF7L1 binding near TSS's.

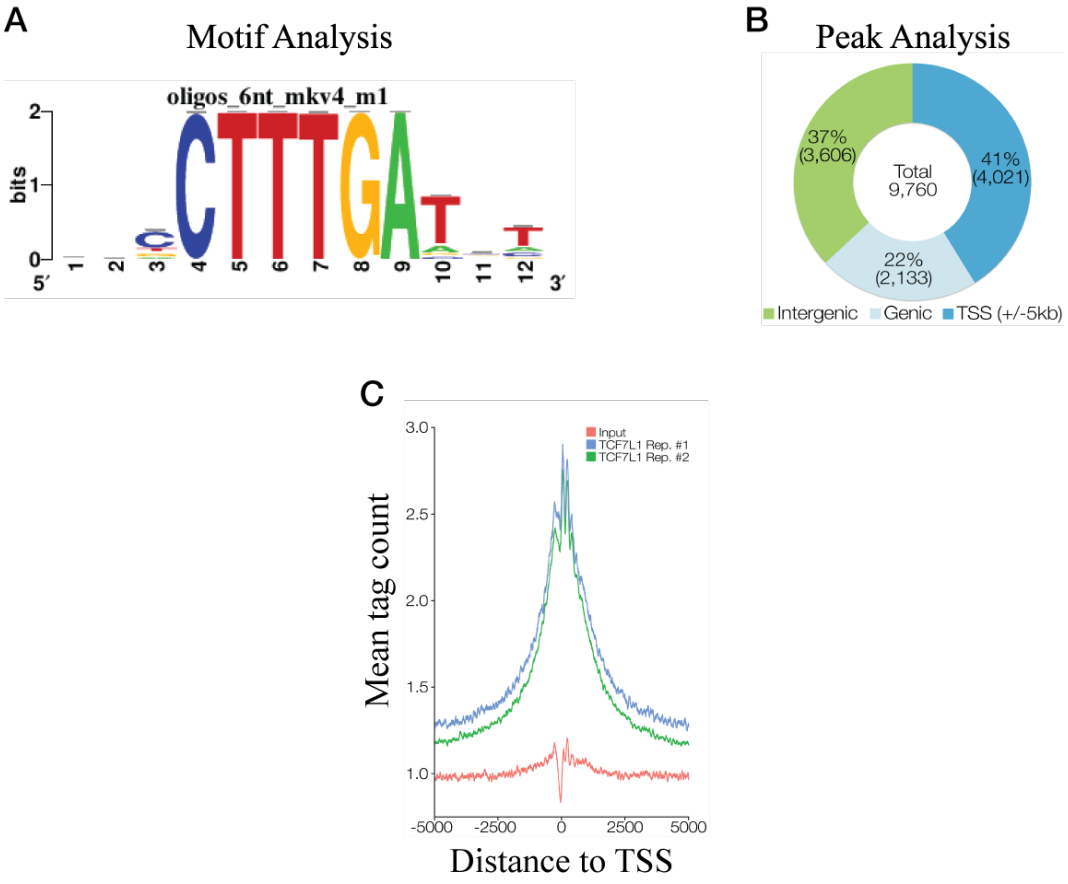


Figure 3.4: **Characterization of TCF7L1 binding sites in hESCs**. **A**. Regulatory Sequence Analysis Tools (RSAT) de novo motif analysis of the top 1400 ChIP-seq peaks (IDR <0.001). The indicated motif is the known TCF7L1 binding motif, and was the top motif found in the RSAT oligos algorithm. This motif was also found using two other de novo motif finders: BETA and HOMER (not shown). **B**. Location of the top 9760 peaks (IDR <0.013249) relative to the nearest gene as determined by HOMER annotation. **C**. CoverageBed was used to determine the number of reads that map to each base-pair position +/-5 kb of the TSS of all human hg19 RefSeq genes. The number of mapped reads used for this analysis was the same for each sample (approximately 54 million). R was used to calculate and plot the mean tag count at each position, which represents the total number of reads mapped to each base divided by the number of RefSeq genes.

On their own, the 9,760 peaks did little to inform us about the function of *TCF7L1*. We used GREAT (Genomic Regions Enrichment of Annotations Tool) analysis to look for the predominant functional categories of our TCF7L1-bound genes [70]. Interestingly, Gene Ontology analysis showed that biological processes involved in gastrulation were highly represented among TCF7L1 gene targets; for example, Gastrulation, Anterior/Posterior Pattern Specification, Formation of the Primary Germ Layers and Mesoderm Formation (**Figure 3.5**). Additionally, GREAT provided mouse phenotype analysis as well; TCF7L1-bound genes were important for germ layer formation and accompanying morphology processes (i.e., Abnormal Primitive Streak Morphology, Abnormal Endoderm, Ectoderm, and Mesoderm Development) - corroborating the Gene Ontology data (**Figure 3.5**). Because TCF7L1 bound to genes involved in primitive streak formation, I postulated that *TCF7L1* might regulate epiblast differentiation during the initiation of gastrulation.

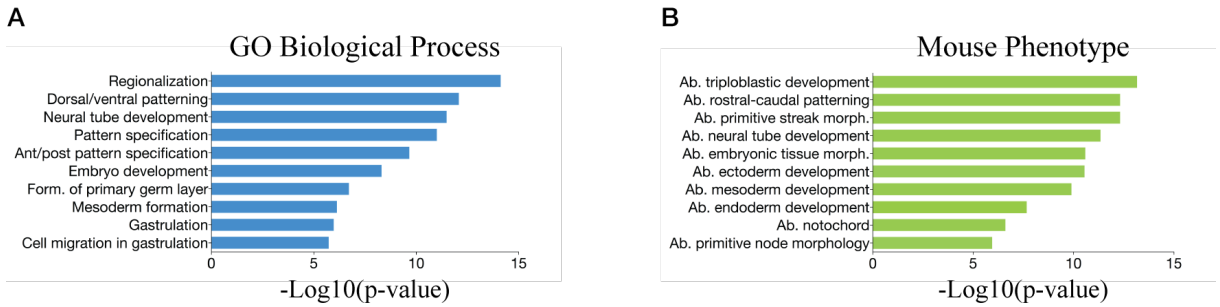


Figure 3.5: **Gene ontology analysis of *TCF7L1* target genes**. Genomic Regions Enrichment of Annotations Tool (GREAT) was used to analyze underlying gene ontology themes represented among *TCF7l* target genes. The top 1000 peaks were analyzed using the "Basal plus extension" default settings. **A**. Bar graph displaying the top ten Biological Process categories identified by GREAT. **B**. Bar graph displaying the top ten Mouse Phenotype categories associated with TCF7L1 target genes. Data is presented as the negative log-base 10 of their respective p-values for convenience.

Because the very first event of gastrulation is formation of the primitive streak, the ChIP-seq data were analyzed to predict whether *TCF7L1* targeted genes were involved in this process. Mouse Genome Informatics (MGI) provides a repository of mutant genotypes and their

consequential phenotypes called Mammalian Phenotype (MP) Browser. The MP database contains a category termed "failure of primitive streak formation" which is defined as the "inability to form the epiblast ridge from which arises the germ layers of the embryo". Of the 35 genes included in this list, 16 were bound by TCF7L1 within +/-5 kb of their TSS. Validation of TCF7L1 binding was performed by ChIP-qPCR on ten of genes required for primitive streak formation (**Figure 3.6**). These results strongly implicate *TCF7L1* as a regulator of genes involved in primitive streak differentiation.

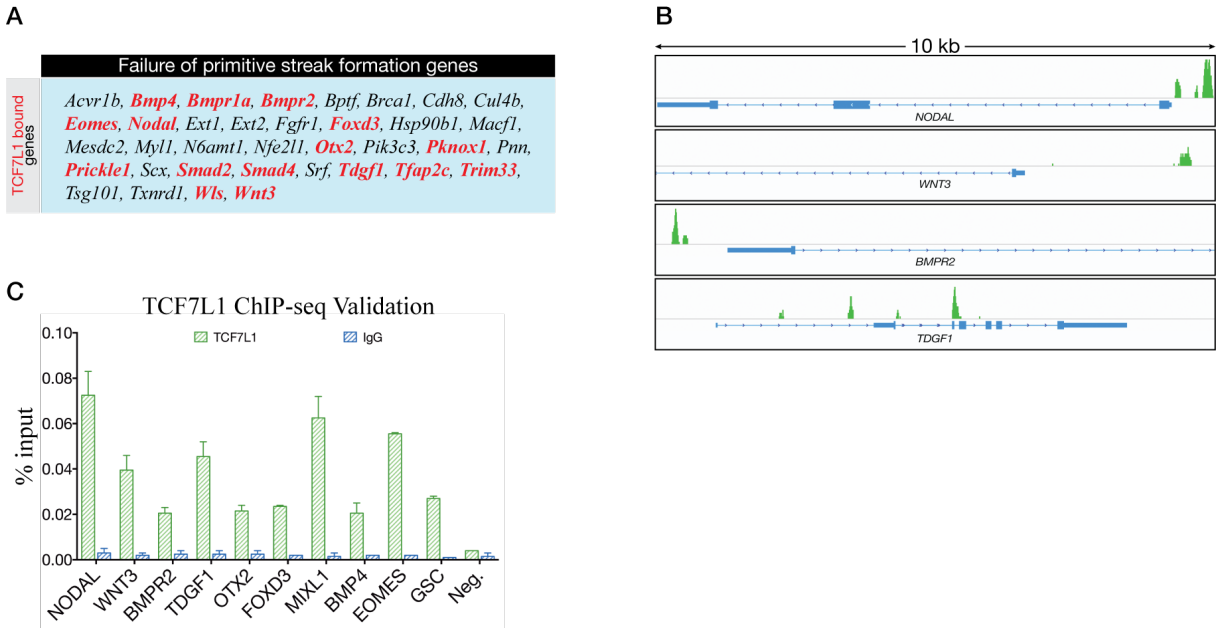


Figure 3.6: ***TCF7L1* binds genes necessary for primitive streak formation**. **A**. TCF7L1 bound genes (marked in red) represented in the Mammalian Phenotype category "Failure of primitive streak formation". Genes with TCF7L1 binding within +/-5 kb were used for this comparison. **B**. Genome Browser representation of TCF7L1 enrichment relative to some of the candidate primitive streak genes. **C**. ChIP-qPCR analysis of TCF7L1 binding to critical primitive streak genes. Enrichment analysis is presented as percent input recovery of TCF7L1 (green) and IgG (blue). The data is derived from two technical replicates.

3.3 Discussion

Next-generation sequencing analysis of TCF7L1 binding sites in hESCs suggests that it is a regulator of early gastrulation differentiation. What is particularly striking is that numerous genes bound by TCF7L1 close to their TSSs (+/-5 kb) are required for primitive streak formation; for example, the classical primitive streak regulators *NODAL*, *WNT3* and *BMP4* [28][61][127][106]. Guided by these findings, I hypothesize that *TCF7L1* regulates the first phase of the epiblast differentiation program, an event that occurs as nascent primitive streak is formed.

This is not the first time genome-wide binding analysis has been performed for TCF7L1. Previously, Tcf7l1 binding was mapped in mESCs by ChIP-chip (Cole et al.) and by ChIP-seq (Marson et al.) [26][65]. Despite alternative methodologies, both groups reported a high degree of overlap between Tcf7l1 and the core regulatory network members - Oct4, Sox2 and Nanog - across the genome, which were frequently positioned near their transcription start sites suggesting overlapping regulation [26][65]. Interestingly, both groups shared a common observation suggesting that the cohort of overlapping genes fell into two classes: actively transcribed self-renewal genes and inactive differentiation genes. These studies were not the first to discover that the core regulatory network has targeted paradoxical gene sets. The prescribed underlying theme was that *Tcf7l1* limits activation intensity of the pluripotency gene network while preventing the core regulatory network genes from prematurely activating "poised" differentiation genes.

While, we were unable to directly test for overlap of TCF7L1, OCT4, SOX2 and NANOG binding in hESCs, ChIP-seq datasets are available for OCT4, SOX2 and NANOG in hESCs [60][79][15]. However, the different procedural and analytical methods made this meta-analysis too complicated. However, we were able to ask whether or not TCF7L1 peaks were located near *OCT4*, *SOX2* and *NANOG* promoters. Like mESCs, we find that TCF7L1

binds to *SOX2* and *OCT4* within +/-5 kb of their promoters, but does not bind to the *NANOG gene*. Although this diverges from the data presented by Cole and Marson, it is possible that TCF7L1 regulates *NANOG* from a distal enhancer [26][65]. Functional assays are needed to determine if *TCF7L1* directly regulates these core regulatory transcription factors in hESCs.

Further support for *TCF7L1* acting as a regulator of primitive streak formation was recently shown by Hoffman and colleagues [47]. Using precisely staged mouse embryos, they surveyed Tcf7l1 protein expression within the epiblast during the progression of primitive streak formation. They found that as gastrulation proceeds, Tcf7l1 is downregulated in epiblast cells in the posterior region of the embryo where the primitive streak is formed. This spaciotemporal pattern of expression, supports the notion that *Tcf7l1* acts as repressor of a primitive streak gene expression program in undifferentiated epiblast cells. Interestingly, Sox2 - a known transcriptional regulator of the neuroectoderm lineage - had a similar protein expression pattern. These data led me to propose that perhaps these two transcription factors have complementary roles during gastrulation.

In attempting to interpret this data, it is important to remember that mouse and human ESCs are not equivalent and represent two different cell types [117][44][110][14][41][92][16][17]. Mouse ESCs represent the earlier "naive" state, whereas hESCs represent the ensuing "poised" state. As such, these two cell types have unique signaling pathway requirements, transcriptomes and cell fate choice capabilities. These differences between human and mouse ESCs could mean that TCF7L1 has a unique function in each of these cell types *TCF7L1*. In the next chapter, I test the hypothesis that a primary role of *TCF7L1* in hESCs is to suppress a primitive streak gene program.

Chapter 4

Functional Analysis of *TCF7L1* in hESCs

4.1 Introduction

Pluripotent cells are defined by their unrestricted developmental potential. During embryogenesis, pluripotent cells serve as the precursors of all germ layer tissues (endoderm, mesoderm and ectoderm). In order to study the mechanisms that define the pluripotent state, two embryonic pluripotent cell models have been developed from early stages of development and maintained *in vitro*: mouse embryonic stem cells (mESCs), derived from the pluripotent cells of the inner cell mass (ICM) of pre-implantation blastocysts, and human ESCs (hESCs), which are also derived from cells of the pre-implantation blastocyst ICM. Despite the fact that they have similar methods of derivation and are both defined as being pluripotent, hESCs are understood to be more developmentally mature than "naive" mouse ICM cells and as such hESCs are more similar to later-stage post-implantation "primed" mouse epiblast cells which are descendants of the ICM [35][113][110][85][88]. Species-specific differences

between mouse and human ESCs are further evident by the different mechanisms governing their self-renewing pluripotent-states *in vivo* and *in vitro*. Because mESCs and hESCs represent two unique developmental stages, they can be used to investigate the mechanisms regulating different transitions during embryogenesis; for example, "naive" ICM-to-epiblast and "poised" epiblast-to-germ lineage. Although many studies have examined the role of the "core transcriptional regulators" (*OCT4*, *SOX2* and *NANOG*) in mESCs and hESCs, and although their respective pluripotency-promoting signaling pathways have also been examined, fewer studies have examined the role of the downstream effectors of the WNT signaling pathway in hESCs. This is perhaps surprising because the WNT signaling pathway is a key regulatory pathway in metazoan development [119].

Canonical WNT signaling is important for the formation of the primary germ layers during gastrulation [119]. A hallmark of gastrulation initiation is the formation of a structure called the primitive streak. This structure arises as epiblast cells differentiate into a bi-potent mesendoderm progenitor, which will either become mesoderm or definitive endoderm depending on their location within the forming embryo [106]. Concomitantly, as the epiblast cells begin to differentiate, they migrate away from the epiblast epithelium in a process known as an epithelial-mesenchymal transition (EMT) [2]. Cells undergoing EMT create a morphologically identifiable "furrow" called the primitive streak, which expands distally as the zone of differentiating epiblast cells expand. In mice, initiation of epiblast differentiation into mesendoderm depends on specific actions of morphogen-stimulated signaling pathways. Previous studies of gastrulation mice have identified *Wnt3*, *Bmp4* and *Nodal* as key morphogens that initiate epiblast differentiation at the onset of gastrulation [127][61][18]. Interestingly, hESCs can be directed to differentiate towards mesendoderm via stimulation of the TGFβ/ACTIVIN/NODAL, BMP4 and WNT signaling pathways, indicating that they respond to morphogens in the same way as the pluripotent epiblast cells from which they are derived [29][116][100][10][104].

As previously described, mESCs and hESCs are thought to represent two distinct developmental stages. The differences between these two cell types are highlighted by their responses to signaling pathway morphogens. For example, mESCs require LIF to maintain pluripotency, whereas hESCs are not responsive to LIF. Similarly, Bmp4 and Wnt signaling maintain mESC pluripotency, but each factor can induce hESC differentiation [14][116]. To better define the molecular basis for the opposing responses of human and mouse ESCs to canonical WNT signaling, I focused on understanding the role of the *TCF7L1* in hESCs. In mESCs, *Tcf7l1* - also the most highly expressed LEF/TCF - is known to repress expression of pluripotency genes [89][26][65][107][137]. Therefore, I asked how *TCF7L1* functioned in "primed" hESCs. I hypothesized that *TCF7L1* would have a distinct function in "primed" hESCs.

In this chapter, I describe studies in which I investigated how *TCF7L1* functions in hESCs. Previously, I showed that hESCs do not possess active canonical WNT signaling, yet are highly responsive to exogenous WNT3A [31][104][10][30]. This suggests that hESCs, like "primed" epiblast cells of the mouse embryo, are capable of responding to WNT ligands [12]. In contrast, to its pluripotency-limiting role in mESCs, I find that *TCF7L1* preserves hESC pluripotency through suppression of a differentiation program equivalent to that which normally acts within the developing primitive streak. Together, these results fill vital gaps in our understanding of how hESC pluripotency is maintained and, by extension, provides novel insights into how "primed" epiblast differentiation is regulated at the onset of gastrulation.

4.2 Results

4.2.1 *TCF7L1* siRNA Knockdown

To dissect the function of TCF7L1 in hESCs, I began by performing a loss-of-function assay to deplete *TCF7L1* mRNA by transient siRNA knockdown (KD) (**Figure 4.1**). After three days of siRNA treatment, *TCF7L1* mRNA and protein levels were analyzed and found to be significantly reduced by approximately 90% (**Figure 4.1B,C**). Morphologically, loss of *TCF7L1* produced "spiky" colonies, as if the peripheral cells of colonies were migratory (**Figure 4.1C**). Based on these observations, I surmised that *TCF7L1* inactivation might be inducing an EMT response, yet there was no change in key EMT marker gene mRNAs (*SNAIL1*, *CDH1*) (**Figure 4.2A**) [25][111].

Previous studies have demonstrated that individual LEF/TCF members can behave very differently, even antagonistically, with regard to WNT target gene regulation [136]. Given their different activation and repression abilities, it was necessary to investigate the possibility that different LEF/TCF members compensate for each other during the *TCF7L1* knockdown and therefore affected the results. Quantitative PCR analysis showed that *TCF7* and *LEF1* were up-regulated by 1.2 and 1.8 fold three days post *TCF7L1* knockdown, while *TCF7L2* was unaffected (**Figure 4.2B**). Although *TCF7* and *LEF1* showed low levels of upregulation, protein signals were undetectable by Western blot (data not shown) consistent with the idea that hESCs express little or no *TCF7* and *LEF1*. Therefore, the effect of *TCF7L1* knockdown on *TCF7* and *LEF1* was considered to be inconsequential relative to the outcome of *TCF7L1* inactivation within this timeframe.

Variation in the expression of the core transcriptional regulatory network factors (*OCT4*, *SOX2* and *NANOG*) has been widely used to inform about the status of the pluripotent state. Therefore, I asked how loss of *TCF7L1* might influence the expression of these factors.

Based on *Tcf7l1* studies in mESCs, one would predict that these factors would be up-regulated - especially *OCT4* and *NANOG*. Interestingly, qPCR analysis corroborated the mESC data as *OCT4* (1.2 fold) and *NANOG* (1.4 fold) were up-regulated to nearly the same extent as in mESCs (**Figure 4.2C**) [89][26]. There was no affect on *SOX2* expression (**Figure 4.2C**). Based on the fact that *OCT4*, *SOX2* and *NANOG* levels either remained the same or were slightly increased, I tentatively concluded that *TCF7L1* is not necessary to maintain pluripotency in poised hESCs.

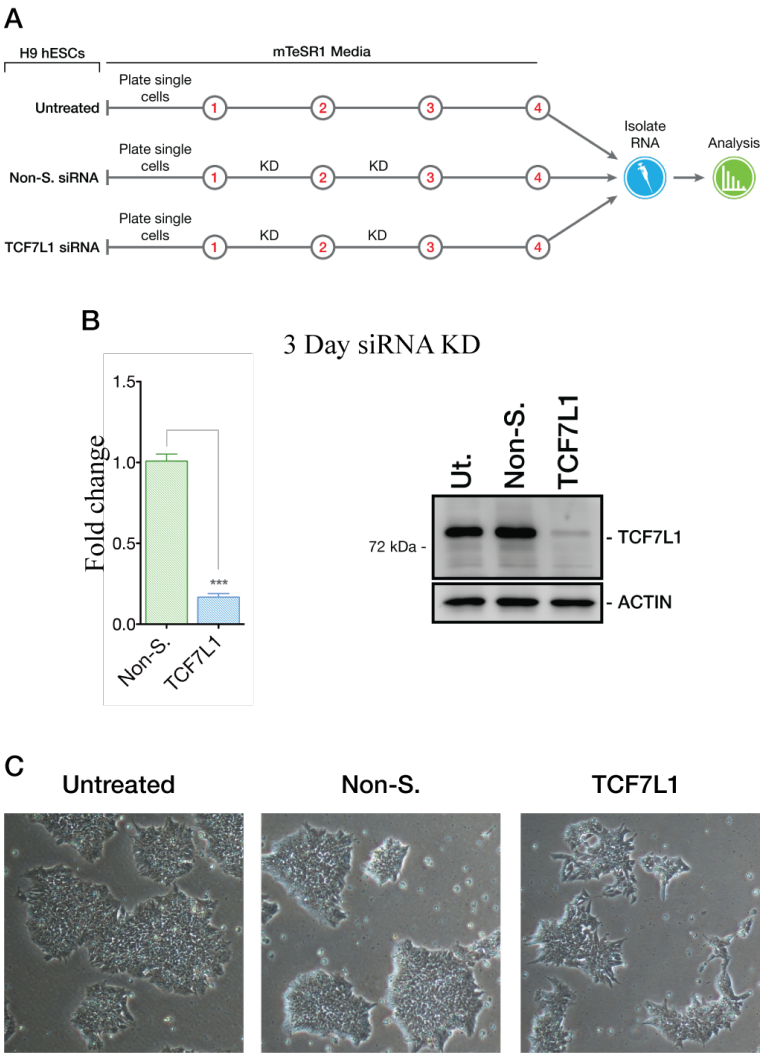


Figure 4.1: ***TCF7L1* siRNA knockdown**. **A**. Schematic of the *TCF7L1* siRNA knockdown experiments. Red numbers indicate days (24 hour intervals) of the procedure. Cells were harvested on the fourth day, 72 hours after the siRNA knockdown was initiated. **B**. *TCF7L1* siRNA knockdown significantly decreases *TCF7L1* levels after 3 days as analyzed by qPCR and Western blot. Non-silencing (Non-S.) and TCF7L1 siRNAs were used at 50 nM. Experiments were carried out in triplicate and results are shown as fold change relative to the Untreated (Ut.) control. **C**. Images of hESC colonies show that loss of *TCF7L1* induces morphological changes in hESCs. Phase contrast images (10X) of *TCF7L1* KD hESCs displayed smaller, less-dense colonies, which appeared to be a result of hESCs extending away from the periphery of the colonies. Data and statistics: performed by t test, presented as means ± SEM (***p <0.001).

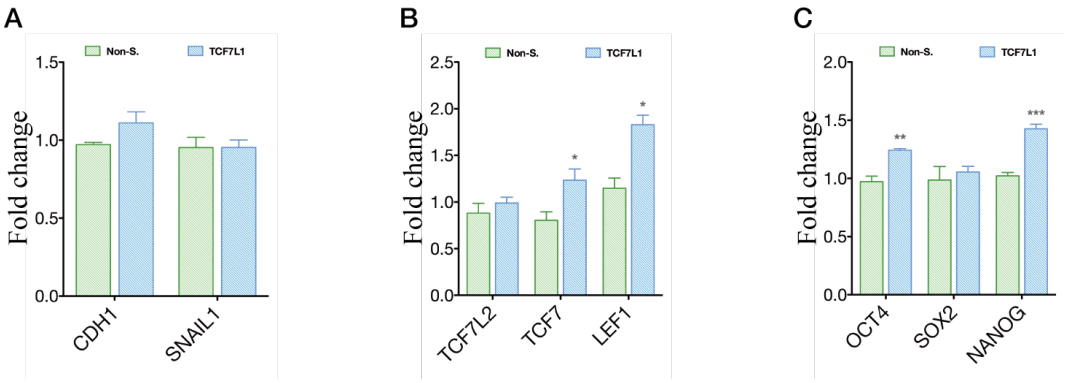


Figure 4.2: ***TCF7L1* siRNA knockdown analysis**. **A**. EMT marker genes *CDH1* and *SNAIL1* are unaffected by *TCF7L1* siRNA as measured by qPCR (n=3). **B**. *LEF1* and *TCF7* are minimally affected by *TCF7L1* KD. Quantitative PCR analysis of LEF/TCF mRNAs 3 days post-siRNA KD of *TCF7L1* show minimal upregulation of *LEF1* (Avg. Ct = 34.1) and *TCF7* (Avg. Ct = 29.7) (n=3). **C**. Core transcriptional regulatory transcription factors *OCT4* and *NANOG* are slightly up-regulated by *TCF7L1* knockdown, but not *SOX2* (n=3). Data and statistics: performed by t test, presented as means ± SEM (*p <0.05, **p <0.01, ***p <0.001).

To assess global changes in gene expression following TCF7L1 knockdown, I performed microarray analysis using mRNA from *TCF7L1* siRNA knockdown hESCs. Untreated hESCs and hESCs transfected with a non-silencing siRNA were used as controls. Surprisingly, very few mRNAs were significantly altered in expression level based on a p-value cutoff of 0.01. In fact, only five mRNAs (*DACT1*, *THBS2*, *FST*, *NODAL*, *FAT4*) registered significant changes and even these had modest fold-changes in expression, ranging from 3.1 to -5.1 fold. In an effort to extract more substantive information from the data, the cutoff criteria were relaxed to include a p-value cutoff of ≤ 0.05 and a fold-change cutoff of 1.5 and -1.5. Even with these relaxed parameters, very few mRNAs were noted as significantly altered (49 up-regulated genes and 37 down-regulated genes) (**Figure 4.3A,B**). Using qPCR, a panel of up-regulated candidate genes were validated and despite the modest nature of gene expression changes, I confirmed that the microarray data were accurate (**Figure 4.4A,B**). These results were also confirmed using a second *TCF7L1* targeting siRNA (**Figure 4.3C,D**).

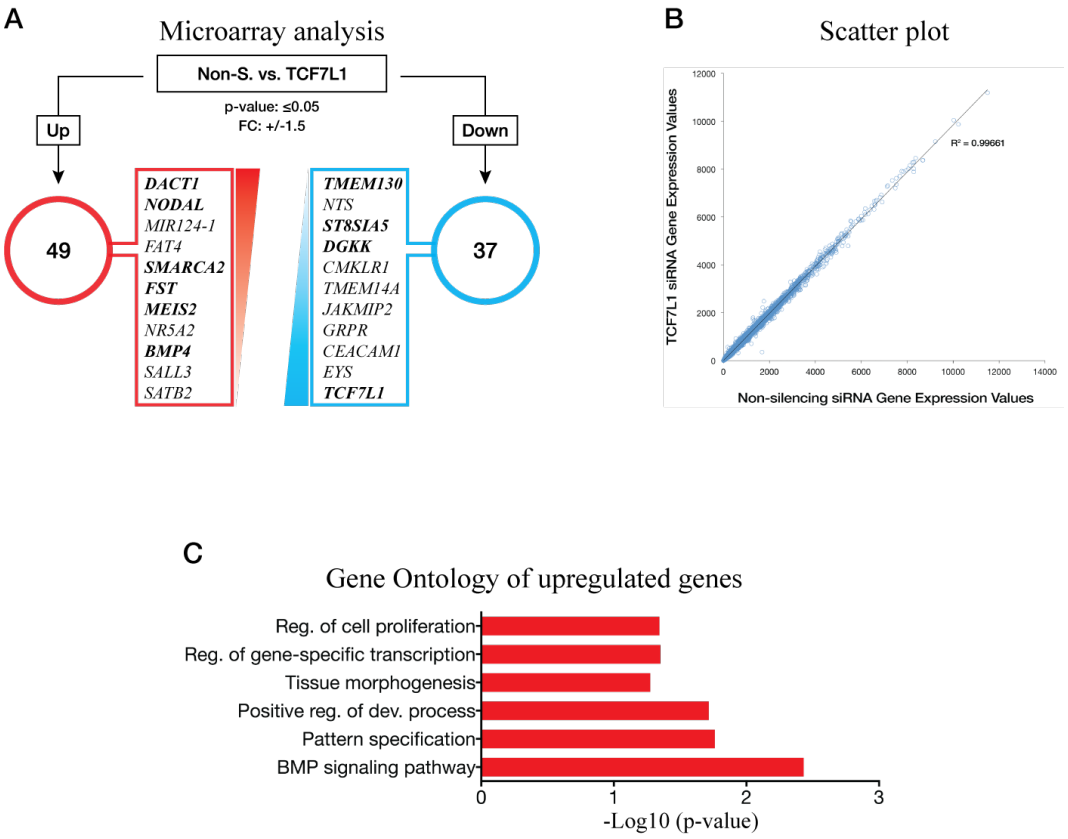


Figure 4.3: **Microarray analysis of *TCF7L1* siRNA knockdown**. **A**. Diagram illustrating the results of the 3 day TCF7L1 siRNA knockdown microarray analysis. Analysis was performed using Cyber-T (http://cybert.ics.uci.edu) software and gene candidates were selected based on the following criteria: p-value of $\leq$0.05; fold changes of $\geq$ 1.5 and $\leq$ -1.5. **B**. Scatter plot of microarray mean gene expression values from the *TCF7L1* siRNA condition plotted against Non-silencing (Non-S.) siRNA condition. Uniform linear distribution signifies highly similar expression values in both conditions. **C**. DAVID (www.david.abcc.ncifcrf.gov) was used for gene ontology analysis of genes up-regulated by *TCF7L1* siRNA knockdown. Only Biological Process categories with a p-value of $\leq$0.05 were considered significant. No Biological Process categories were significantly enriched in the down-regulated gene cohort. X-axis shows -Log base 10 conversion of the p-values.

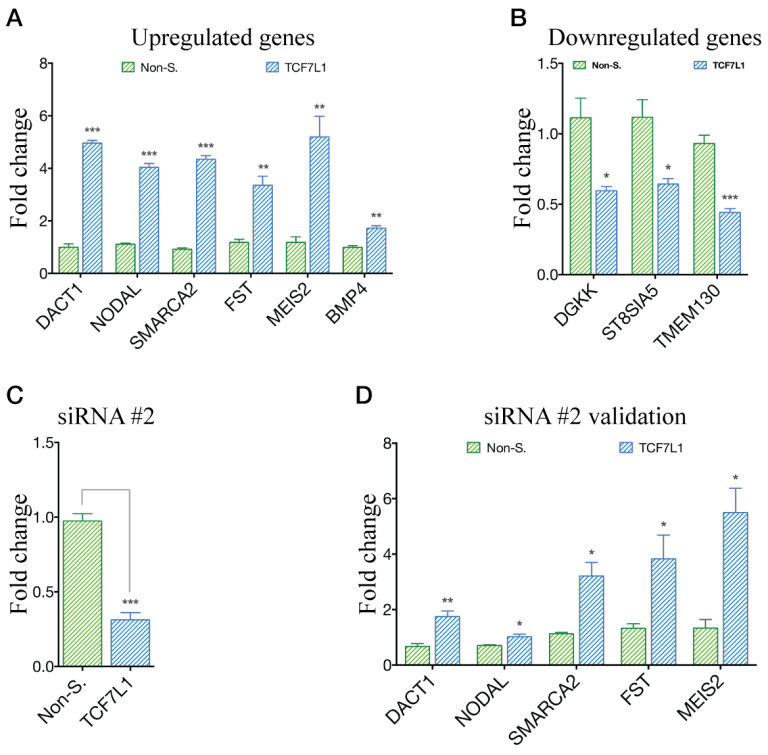


Figure 4.4: **Validation of microarray results**. **A**. Candidate up-regulated genes identified by microarray analysis were validated by qPCR. **B**. Quantitative PCR validation of candidate down-regulated genes. **C**. *TCF7L1* knockdown using a second siRNA (siRNA #2) after 3 days. Non-silencing (Non-S.) and *TCF7L1* siRNA were used at 50 nM. This second *TCF7L1* siRNA is not as effective as the siRNA used for the original microarray analysis. **D**. Confirmation of *TCF7L1*-regulated genes identified in the microarray study using siRNA #2 by qPCR (n=3). Data are presented as fold changes in expression relative to the No siRNA condition. Statistical analysis was performed by Student's t-test and presented as mean +/- SEM (*p <0.05, **p <0.01, ***p <0.001).

While lack of a *TCF7L1* loss-of-function phenotype begged the question of whether or not *TCF7L1* performed any substantive function at all, the microarray data provided insightful clues into its function. Specifically, up-regulated genes identified in the microarray, (*NODAL* and *BMP4*, had known, critical, regulatory functions in mesendoderm formation during gastrulation. For example, *Nodal* and *Bmp4* knockout mice do not properly initiate mesendoderm differentiation and fail to form the primitive streak, leading to an embryonic lethal phenotype [138][28][127]. Thus, I speculated that *TCF7L1* might repress the expression of genes involved in mesendoderm specification of epiblast cells at the onset

of gastrulation.

Furthermore, I reasoned that even though *TCF7L1* appeared to be inconsequential to the pluripotent state of "poised" hESCs, it was also possible that a highly optimized, pluripotency-stabilizing, defined medium (mTeSR1) could be masking the effects of *TCF7L1* knockdown on gene expression and pluripotency. To address this possibility, I chose to use a *TCF7L1* gain-of-function assay while inducing hESC differentiation.

4.2.2 Over-expression of *TCF7L1*

Gain-of-function experiments were performed using the TET-ON doxycycline-inducible lentiviral system. I generated clonal hESC lines to ensure consistent uniform induction of a *TCF7L1* transgene. I developed one clonal cell line to over-express full-length *TCF7L1* with an N-terminal FLAG-tag and a C-terminal HTBH-tag (H9-TCF7L1) and a control cell line to over-express GFP (H9-GFP) [105]. Both cell lines robustly expressed their transgene when treated with doxycycline (**Figure 4.5**). Next, I developed a differentiation protocol capable of inducing primitive streak-like mesendoderm gene expression (e.g., *BRACHYURY*, *EOMES*, *MIXL1*, *GSC*). This protocol did not require a complex cocktail of growth factor administration but instead was an abridged version (48 hours) of the definitive endoderm differentiation procedure described by Teo et al., [109]. This modified procedure was termed primitive streak-like (PS) differentiation and is outlined in **Figure 4.6A**. Importantly, PS differentiation was performed in feeder-free conditions with a fully defined medium containing recombinant BMP4 (50 ng/mL) and ACTIVIN A (50 ng/mL), affording precise control over hESC differentiation.

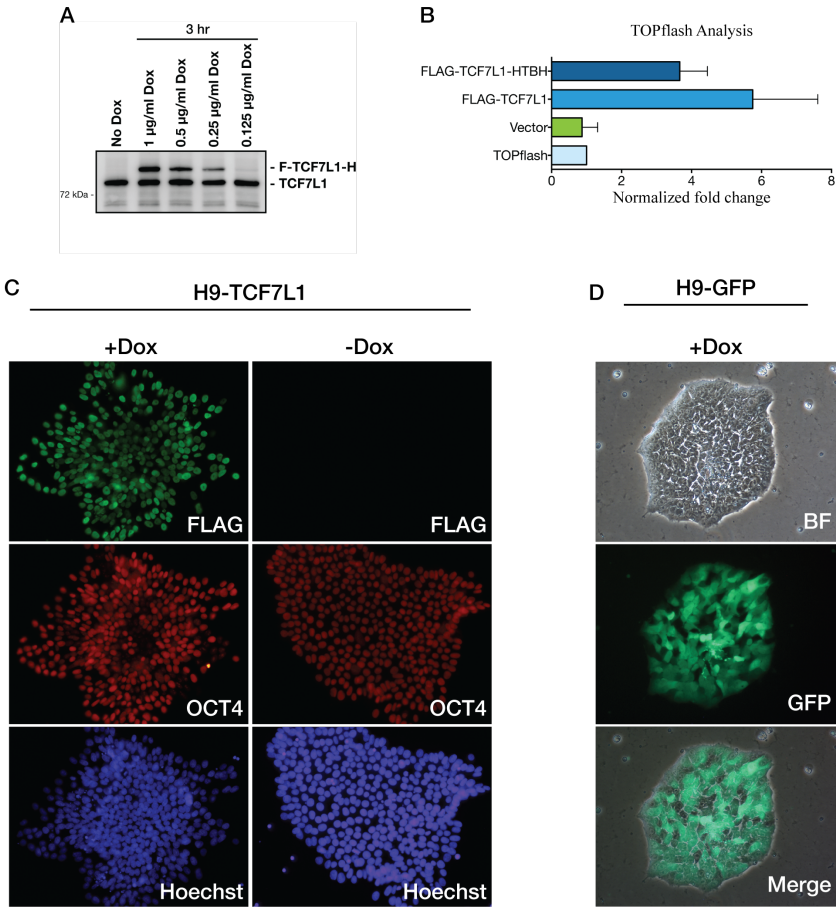
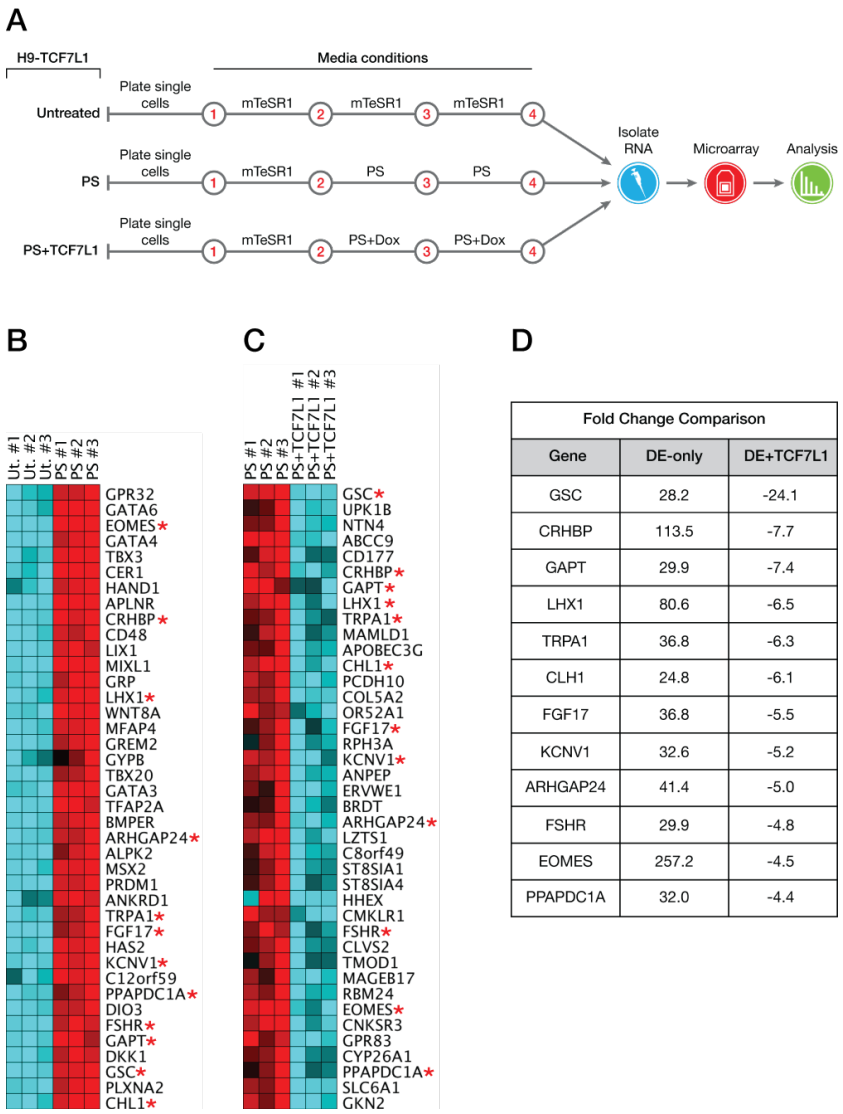


Figure 4.5: **Characterization of an H9 cell line over-expressing *TCF7L1*.** **A**. FLAG-TCF7L1-HTBH (F-TCF7L1-H) over-expression is robustly controlled by doxycycline. H9-TCF7L1 cell line was induced for 3 hours with the indicated concentrations of doxycycline and then whole cell extracts were harvested for Western blot analysis. The higher molecular weight band is F-TCF7L1-H and the second, slightly lower, band is endogenous TCF7L1 which serves as a loading control. The data show rapid responses to doxycycline and no observable leaky expression in the absence of doxycycline. **B**. F-TCF7L1-H does not significantly affect its ability to activate a TOPflash reporter in the presence of β-CATENIN. COS-1 cells were transfected with either: F-TCF7L1-H, FLAG-TCF7L1, empty vector (Vector) or TOPflash alone. In addition, each condition included a co-transfected β-CATENIN expression plasmid. For this assay there must be a functional LEF/TCF and β-CATENIN in order to activate the TOPflash reporter. After 24 hours, cells transferred in each condition were harvested and their luciferase expression was measured (n=3). F-TCF7L1-H was slightly, but not significantly, less active than FLAG-TCF7L1. **C**. F-TCF7L1-H is localized to the nucleus in H9-TCF7L1 line. Immunofluorescence analysis of OCT4 (red) and FLAG (green) in H9-TCF7L1 under feeder-free conditions with or without 24 hours of doxycycline induction. Nuclei were stained with Hoechst. **D**. The H9-GFP inducible cell line robustly and uniformly expresses GFP in the presence of doxycycline. Live cells were imaged for GFP expression after 24 hours of doxycycline (1 μg/ml) induction.

First, the PS differentiation gene expression signature was determine using microarray analysis by comparing untreated H9-TCF7L1 cells to PS differentiated H9-TCF7L1 cells. After 48 hours, 1409 genes were up-regulated and 1596 genes were down-regulated. The top 40 up-regulated genes are shown in **Figure 4.6B**. As expected, there was a strong induction of primitive streak/mesendoderm genes (*EOMES*, *MIXL1*, *LHX1*, *BRACHYURY*, *GSC*). Gene ontology (GO) analysis showed that genes involved in development and differentiation (i.e., gastrulation, primary germ layer formation and mesoderm development) were significantly represented within the up-regulated gene set (**Figure 4.7A**). These results confirmed that the PS differentiation procedure induced a primitive streak-like gene expression program in hESCs.



Fold Change Comparison		
Gene	DE-only	DE+TCF7L1
GSC	28.2	-24.1
CRHBP	113.5	-7.7
GAPT	29.9	-7.4
LHX1	80.6	-6.5
TRPA1	36.8	-6.3
CLH1	24.8	-6.1
FGF17	36.8	-5.5
KCNV1	32.6	-5.2
ARHGAP24	41.4	-5.0
FSHR	29.9	-4.8
EOMES	257.2	-4.5
PPAPDC1A	32.0	-4.4

Figure 4.6: **H9-TCF7L1 over-expression analysis**. **A**. Schematic of *TCF7L1* over-expression microarray experiment. H9-TCF7L1 cells were treated with the indicated conditions under feeder-free conditions. Doxycycline was used at a concentration of 1 μg/ml to induce *TCF7L1* expression. Red numbers indicate days of the procedure (24 hr intervals). Cells were harvested for analysis on the fourth day. **B**. PS differentiation induces the expression of genes associated with primitive streak and mesendoderm. Genome-wide expression analysis was performed after 48 hours of PS differentiation by comparing Untreated and PS conditions. Heat map represents Log2 transformed microarray expression values for each replicate and shows the top 40 up-regulated genes. **C**. *TCF7L1* over-expression during PS differentiation suppresses the expression of genes associated with primitive streak/mesendoderm. Genome-wide expression analysis was performed after 48 hours of PS differentiation + TCF7L1 induction by comparing PS and PS+TCF7L1 conditions. Heat map and results are presented as described in (B). **D**. *TCF7L1* represses genes induced by PS-only differentiation. Table listing fold change expression values for overlapping genes from (B) and (D). Red asterisk indicates genes identified in both datasets.

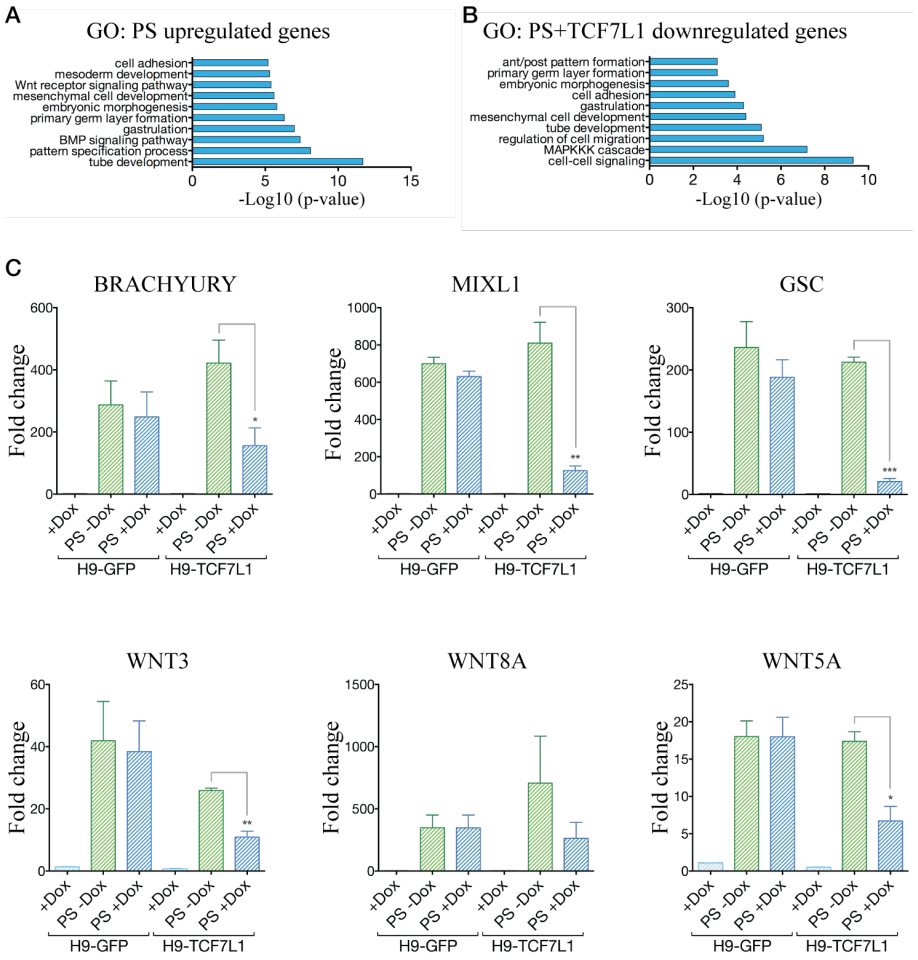


Figure 4.7: ***TCF7L1* over-expression opposes primitive streak gene expression**. **A**. Gene ontology analysis (GO) of genes induced by PS differentiation conditions. **B**. GO analysis of genes repressed by *TCF7L1* over-expression during PS differentiation condition (PS+TCF7L1). **C**. Quantitative PCR validation of candidate *TCF7L1*-repressed genes identified by microarray analysis. Analysis was performed with H9-TCF7L1 and control H9-GFP cell lines (n=3). Data is presented as fold change in expression relative to the Untreated -Dox condition. Statistics performed by Students t-test and presented as means ± SEM (*p <0.05, **p <0.01, ***p <0.001).

Next, I tested whether *TCF7L1* over-expression during PS differentiation affected the expression of primitive streak/mesendoderm genes induced by PS differentiation alone. For this, the "PS-treated H9-TCF7L1 condition" was compared to the "Dox-treated PS-treated H9-TCF7L1 condition" (PS+TCF7L1). Based on this comparison, 917 mRNAs were more highly expressed in the PS+TCF7L1 condition and 970 mRNAs were less expressed in the PS+TCF7L1 condition (**Figure 4.8A**). Interestingly, genes associated with the primitive streak and mesendoderm (e.g., *BRACHYURY*, *EOMES*, *MIXL1*, *GSC*, *WNT3*) were significantly suppressed by induced expression of *TCF7L1* during PS differentiation (the top 40 down-regulated genes are shown in **Figure 4.6C**). Importantly, there was a notable overlap between the genes most highly up-regulated by PS differentiation and the genes most highly down-regulated by *TCF7L1* over-expression during PS induced differentiation (PS+TCF7L1), implying a strong inverse correlation between these two datasets (**Figure 4.6D**). Furthermore, GO analysis indicated overlapping biological processes between PS+TCF7L1 down-regulated genes and PS up-regulated genes; for example, gastrulation, embryonic morphogenesis, and primary germ layer formation (**Figure 4.7B**). Using control H9-GFP mRNA from identical treatment conditions, *TCF7L1*-specific repression of candidate primitive streak/mesendoderm markers *BRACHYURY*, *GSC*, *MIXL1* and *WNT3*, as well as *WNT8A* and *WNT5A* were validated using qPCR (**Figure 4.7C**). These data support the hypothesis that *TCF7L1* represses primitive streak/mesendoderm genes in hESCs.

I postulated that the genes suppressed by *TCF7L1* might be involved in mesendoderm/primitive streak differentiation in hESCs. Indeed, a comparison of the set of genes increased by PS conditions to the set of PS+TCF7L1 down-regulated genes, revealed a highly significant overlap of 544 genes (56% of PS+TCF7L1 up-regulated genes) (**Figure 4.8A**). Similarly, the genes down-regulated by PS differentiation conditions versus PS+TCF7L1 up-regulated genes had an overlap of 399 genes (43.5% of the PS+TCF7L1 up-regulated genes) (**Figure 4.8A**). In comparison, only 11.3% of PS+TCF7L1 down-regulated genes

overlapped in both down-regulated gene sets and only 9.8% of PS+TCF7L1 up-regulated genes overlapped in both up-regulated gene sets (**Figure 4.8A**). These findings, therefore, mark *TCF7L1* as a key repressor of genes involved in primitive streak differentiation.

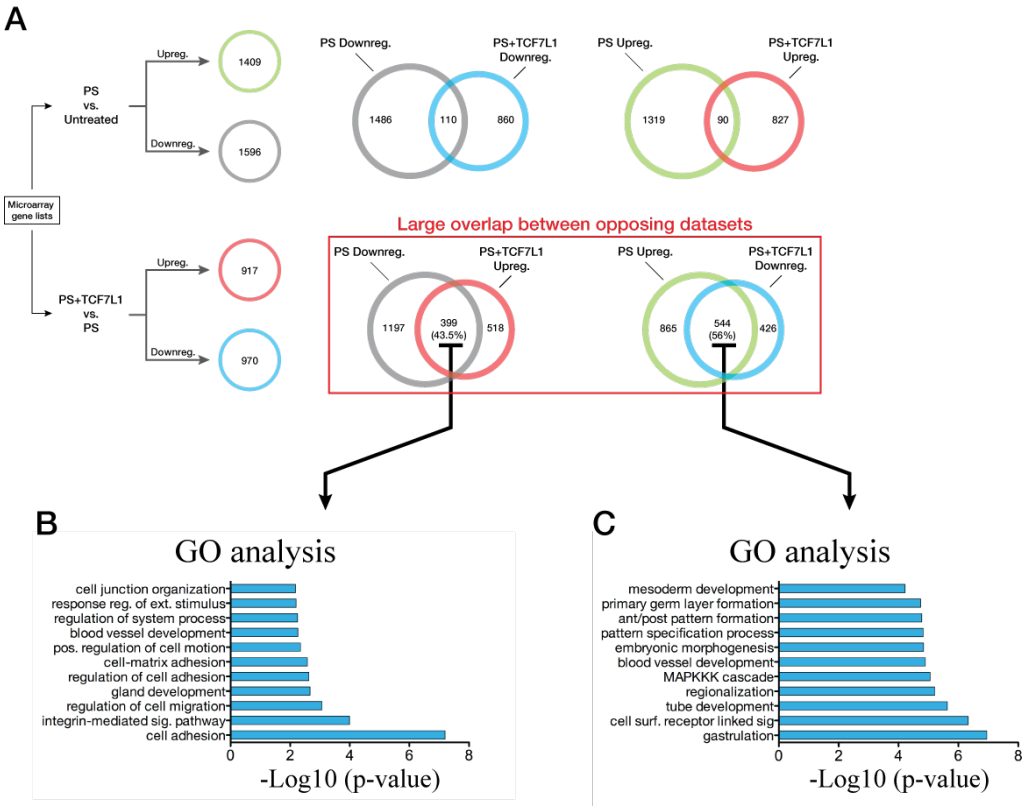


Figure 4.8: ***TCF7L1* suppresses genes induced during gastrulation**. **A**. Venn diagram comparing overlap between PS and PS+TCF7L1 gene sets. PS treatment (as compared to the untreated condition) resulted in 1409 up-regulated genes and 1596 down-regulated genes. PS+TCF7L1 treatment (as compared to the PS condition) resulted in 917 up-regulated genes and 970 down-regulated genes. Our analysis shows a large degree of overlap between opposing datasets. PS down-regulated genes verses PS+TCF7L1 up-regulated genes had an overlap of 43.5% and PS up-regulated verses PS+TCF7L1 down-regulated genes had an overlap of 56%. **B**. GO analysis of "PS Downreg. vs. PS+TCF7L1 Upreg." overlapping genes showing high incidence of genes involved in adhesion and migration. **C**. GO analysis of shared "PS+TCF7L1 Upreg. vs. PS Downreg." genes indicating enrichment of genes involved in differentiation, embryonic patterning and embryonic morphogenesis genes.

Using GO analysis, I investigated which gene categories were enriched in these highly overlapping datasets. GO analysis of the shared PS down-regulated and PS+TCF7L1 up-regulated genes showed enrichment of cell adhesion and cell migration genes (**Figure 4.8B**). In contrast, GO analysis of shared PS up-regulated and PS+TCF7L1 down-regulated genes, indicated enrichment of genes involved in differentiation and development (i.e., gastrulation, mesoderm development, primary germ layer formation) suggesting that TCF7L1 suppresses genes involved in primitive streak formation (**Figure 4.8C**). Focusing specifically on the shared "PS up-regulated and PS+TCF7L1 down-regulated" dataset, I plotted gene expression fold changes for a subset of genes from relevant developmental biological categories (gastrulation, EMT, WNT signaling and lineage genes expressed during primitive streak formation) to highlight the repressive activity of *TCF7L1* over-expression (**Figure 4.9**). In each case, genes that were up-regulated by PS differentiation were weakly induced under conditions of *TCF7L1* over-expression. Together, these data support a model in which *TCF7L1* preserves pluripotency by blocking active transcription of genes poised for expression as epiblast cells respond to differentiation signals, differentiate into mesendoderm and migrate through the primitive streak.

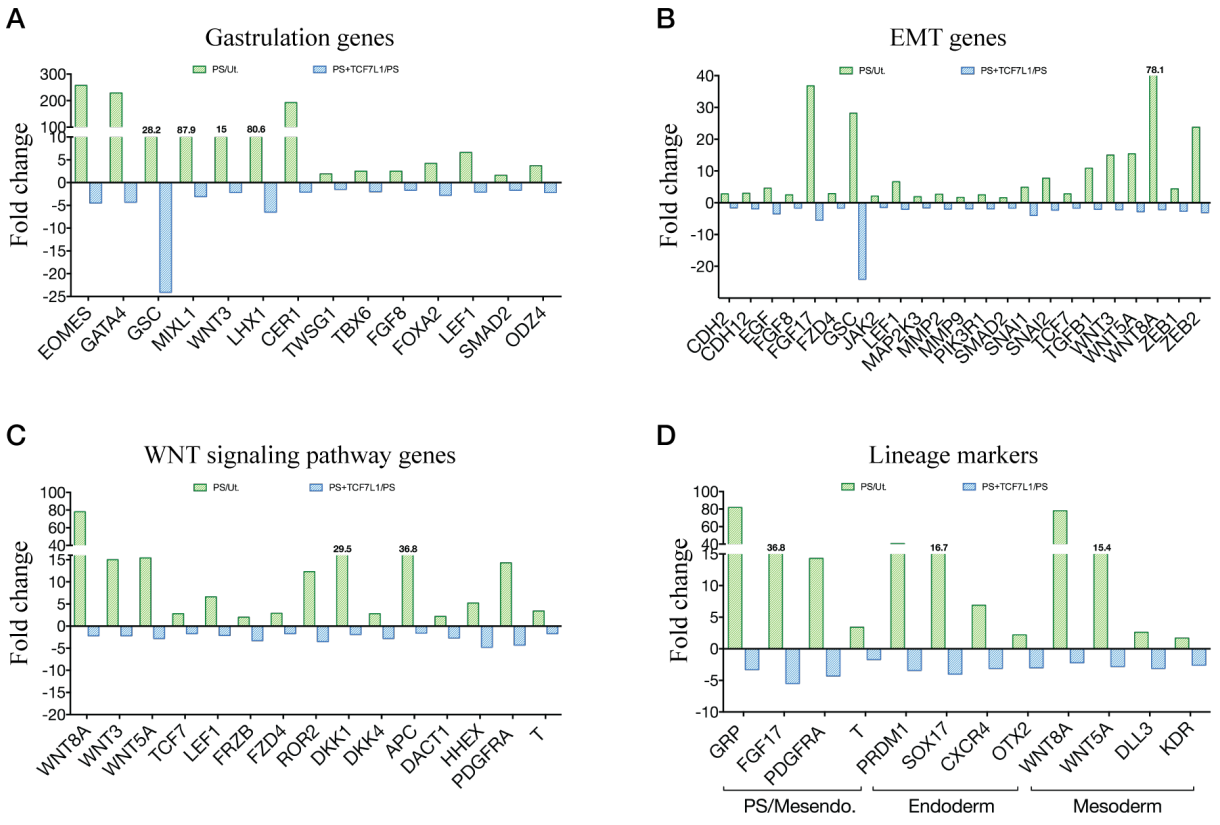


Figure 4.9: **Gene ontology analysis of overlapping *TCF7L1* repressed genes**. Gene ontology analysis of genes shared between the PS up-regulated gene set and the *TCF7L1* down-regulated gene set. Fold changes are shown for specific genes in relevant biological categories. The results show *TCF7L1*-mediated suppression of key genes involved in: gastrulation (A), EMT (B), WNT signaling (C) and mesoderm and endoderm lineages (D).

4.3 Discussion

Using gain- and loss-of-function experimental assays, I have discovered how *TCF7L1* is involved in the regulation of hESC pluripotency. Key insights came from *TCF7L1* siRNA knockdowns where genes involved in mesendoderm differentiation were up-regulated (i.e., *BMP4* and *NODAL*). Moreover, *TCF7L1* over-expression suppressed genes involved in gastrulation, embryonic morphogenesis and mesendoderm specification (i.e., *BRACHYURY*, *GSC*, *EOMES*, *MIXL1*). Together, the functional analyses described in this chapter indicate

that *TCF7L1* represses primitive streak/mesendoderm differentiation in hESCs.

It is worth discussing the somewhat perplexing results of the *TCF7L1* siRNA knockdown because it provided a valuable lesson about experimental design when using hESCs. Loss of *TCF7L1* had a minimal effect on hESCs, which made it difficult to decipher *TCF7L1*'s function. Despite the subtle morphological evidence of hESC differentiation (**Figure 4.1C**), the microarray data implied that *TCF7L1* regulates very few genes in hESCs: 49 genes were up-regulated and 37 genes were down-regulated when *TCF7L1* was knocked down (**Figure 4.3A**). In hindsight, performing the siRNA knockdown experiment in feeder-free conditions with mTeSR1 medium, a condition that that strongly stabilizes the pluripotent state, was counter productive as it likely masked the function of *TCF7L1*. Instead of allowing *TCF7L1* target genes to influence hESC gene expression, high concentrations of FGF2 and TGFβ1 in the medium continuously forced the expression of the pluripotency gene network. For example, if a key role of *TCF7L1* is to repress *BMP4* expression, then high levels of FGF2 in the medium could antagonize autocrine BMP4 signaling since FGF2 is known to antagonize phospho-SMAD1 translocation into the nucleus [82][43]. Another possibility is that *TCF7L2*, which is also expressed in hESCs (albeit at much lower levels than *TCF7L1* (**Figure 2.3B**)), could act redundantly to *TCF7L1* and repress its target genes. In the future, I believe loss-of-function experiments might be better tested under growth factor free conditions in order to circumvent mTeSR1 medium interference.

Going forward, we plan to test the role of *TCF7L1* as a repressor of primitive streak differentiation *in vivo*, using transgenic mouse models. For these experiments, two different transgenic constructs were created to conditionally over-express *TCF7L1* during the initial stages of mouse gastrulation, a developmental stage of embryogenesis that can be isolated and analyzed for abnormal phenotypes. We chose to design and create two different constructs as complementary backup approaches. The first construct uses the TET-ON system to specifically over-express TCF7L1 in epiblast cells undergoing primitive streak differentiation

(**Figure 4.10A**). This is accomplished by placing reverse tetracycline transactivator 3 (rtTA3) under the control of the Brachyury promoter (a gene specifically expressed in the primitive streak) then feeding pregnant mice doxycycline (Dox) containing feed. If the hypothesis proposed here is correct, +Dox embryos over-expressing *TCF7L1* should fail to form a primitive streak because *TCF7L1* would suppress mesendoderm gene expression, conversely -Dox embryos should proceed through this stage. The second construct I created is a construct targeted to the ROSA26 locus and uses the Cre-lox system to conditionally induce *TCF7L1* expression (**Figure 4.10B**). The benefit of ROSA26 target transgene integration is that this locus acts as a "safe harbor" and prevents epigenetic silencing of the construct [36]. This approach is more versatile since it allows one to over-express *TCF7L1* in any lineage, so long as there are gene-specific Cre-expressing mice available for mating. For our purposes, we will specifically induce *TCF7L1* over-expression in cells of the epiblast by mating ROSA-TCF7L1 mice with Sox2-Cre mice. *Sox2* is expressed in epiblast cells, so a *Sox2* promoter driving Cre will activate *TCF7L1* expression before gastrulation begins. Again, the prediction is that embryos induced to over-express *TCF7L1* would fail to form a proper primitive streak. A positive result from either of these mouse models would solidify the finding that *TCF7L1* acts as a suppressor of primitive streak/mesendoderm differentiation.

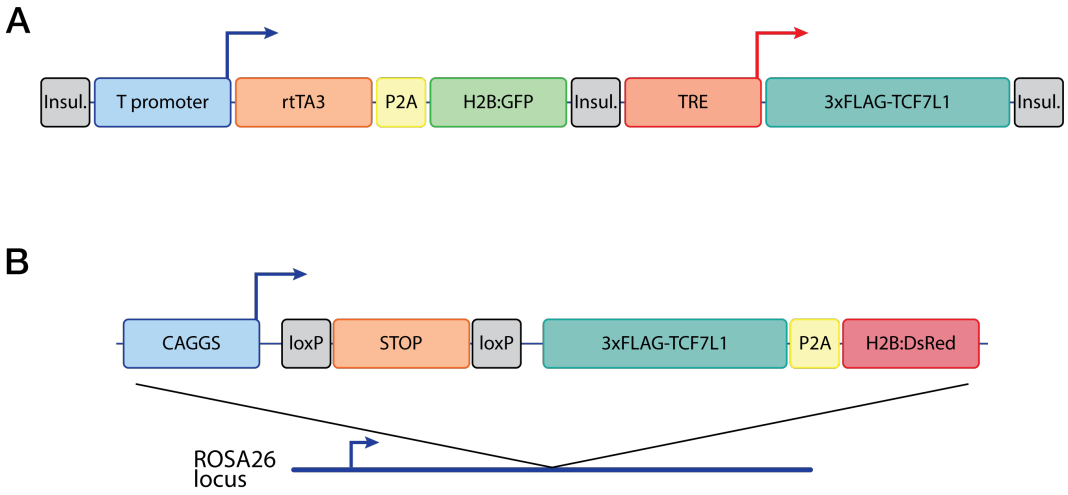


Figure 4.10: **Mouse transgenic *TCF7L1* over-expression constructs**. **A**. Illustration of the TET-On transgenic construct. This construct has the *T* (*Brachyury*) promoter fragment (blue) that is specifically transcriptionally active in cells undergoing mesendoderm differentiation. Once active, the *T* promoter will drive expression of reverse tetracycline transactivator 3 (rtTA3 - orange) and histone H2B tagged GFP (green) for visualization, which are separated by the self-cleaving peptide sequence P2A. Downstream of these components is the tetracycline responsive promoter, which contains multiple rtTA3 binding sites - or tetracycline responsive elements (TRE - red). In the presence of doxycycline (a tetracycline analogue), rtTA3 is able to bind to these TREs and activate the expression of FLAG-TCF7L1 (dark green). Importantly, the different components of this construct are separated by insulators (grey) to prevent both epigenetic silencing and leaky expression. **B**. Illustration of the ROSA26-targeted construct. This construct contains the high activity promoter CAGGS and a series of loxP-flanked (grey) Stop codons (orange) that separate the promoter and the FLAG-TCF7L1 transgene (dark green) from each other. Under normal circumstance, transcription initiated by the CAGGS promoter is inhibited by the intervening stop codons and, thus, FLAG-TCF7L1 is not expressed. However, when mated to a Cre-expressing mouse, Cre removes the loxP-flanked stop codons and allows for transcription of the FLAG-TCF7L1 transgene. Also, in the presence of Cre histone H2B fused to DsRed (red) would be simultaneously expressed because it is placed downstream of FLAG-TCF7L1 and separated by the self-cleaving P2A peptide sequence. DsRed fluorescence allows us to visualize FLAG-TCF7L1 over-expressing cells. Because this transgene is targeted to specifically insert into the ROSA26 locus, it is protected from epigenetic silencing.

Chapter 5

Regulation of *TCF7L1* by *BMP4*

5.1 Introduction

Data presented in the preceding chapters mark *TCF7L1* as a suppressor of the primitive streak differentiation program. Besides understanding how *TCF7L1* functions during gastrulation, it is also important to understand how its' expression is regulated during this process. Since *TCF7L1* is a known transcriptional repressor, one would predict that *TCF7L1* mRNA and/or protein is subject to down-regulation in epiblast cells as they differentiate within the primitive streak. Recently, this prediction was shown to be true by Hoffman and colleagues [47]. Analyzing Tcf7l1 protein levels in pre- to late-streak mouse embryos, they found that Tcf7l1 is progressively down-regulated in posterior epiblast cells undergoing mesendoderm differentiation at the site of the primitive streak. These findings in combination with my studies described in this thesis suggest that *TCF7L1* suppression of mesendoderm genes must be released for proper epiblast differentiation. My experiments further suggest that the signal for TCF7L1 suppression is bone morphogenetic protein 4 (BMP4).

Experiments described in Chapter 2 provided insightful clues as how *TCF7L1* expression

might be regulated. *TCF7L1* mRNA and protein levels were robustly down-regulated in hESCs when treated with BMP4 (**Figure 2.6**). This was particularly interesting because of BMP4's role in driving epiblast differentiation *in vivo* during gastrulation, leading to formation of the primitive streak [127]. In an effort to better understand gastrulation and the regulatory mechanisms that connect *TCF7L1* to this developmental stage, I investigated the regulatory connection between the BMP4 pathway and *TCF7L1*.

5.2 Results

5.2.1 BMP4 induces *TCF7L1* down-regulation

Human ESCs can be directed to differentiate towards mesendoderm using primitive streak-like (PS) differentiation medium, which only contains BMP4 and ACTIVIN A ligands. To test how *TCF7L1* expression was affected by this condition, mRNA and protein levels were analyzed after 24 and 48 hours of treatment. Following the PS differentiation protocol, TCF7L1 protein levels were down-regulated by ∼50% after 48 hours (**Figure 5.1A**). Quantitative PCR was then used to measure the effect of each individual growth factor (BMP4 and ACTIVIN A) on *TCF7L1* mRNA after 24 hours. The data showed that BMP4 caused *TCF7L1* down-regulation by ∼50%, closely mirroring the complete PS differentiation medium condition (**Figure 5.1B**). In contrast, ACTIVIN A and basal medium lacking both growth factors (Blank) did not affect *TCF7L1* expression (**Figure 5.1B**). In addition, a time course analysis of BMP4-treated hESCs showed robust and continuous reduction of TCF7L1 protein (**Figure 5.1C**). Together, these findings demonstrate that BMP4 negatively influences *TCF7L1* expression.

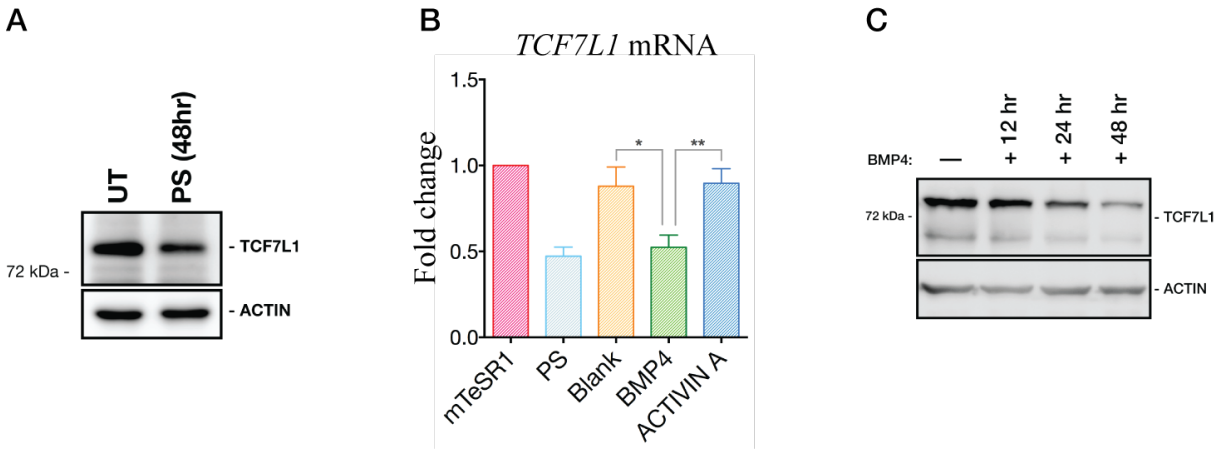


Figure 5.1: ***TCF7L1* is down-regulated by BMP4**. **A**. Western blot comparing TCF7L1 protein levels in hESCs grown in mTeSR1 (UT) and PS differentiation medium (48 hours) containing BMP4 (50 ng/ml) and ACTIVIN A (50 ng/ml). ACTIN serves as the loading control. **B**. Quantitative PCR analysis of *TCF7L1* mRNA levels after 24 hours in each condition. The blank condition is PS differentiation basal medium lacking BMP4 and ACTIVIN A. Data presented as fold change in mRNA relative to untreated mTeSR1 condition (n=3). **C**. Western blot analysis of TCF7L1 protein levels at 12, 24 and 48 hours of BMP4 induced differentiation. ACTIN serves as the loading control. Statistics performed by Student's t-test and presented as mean +/-SEM (*p <0.05, **p <0.01, ***p <0.001).

5.2.2 *TCF7L1* antagonizes BMP4 signaling

Because stimulation of hESCs with BMP4 strongly inactivated *TCF7L1* expression, I postulated that *TCF7L1* might suppress the BMP4-induced gene expression program. To test this hypothesis, *TCF7L1* was depleted in hESCs using siRNA knockdown and then subsequently induced to differentiate with BMP4; the prediction being that loss of *TCF7L1* should enhance the up-regulation of primitive streak/mesendoderm genes (**Figure 5.2A**). To ensure that the gene expression changes were within a discernible dynamic range, a five-fold lower BMP4 concentration (10 ng/ml) was used. This low concentration prevented overwhelming level of induction of primitive streak/mesendoderm genes. After 24 hours of BMP4 stimulation, colonies in the *TCF7L1* knockdown condition exhibited a more exaggerated, differentiated appearance: colony borders were less defined and cells had a

flattened, cobblestone morphology (**Figure 5.2B**). In addition, suppression of *TCF7L1* expression by siRNA knockdown resulted in a dramatic up-regulation of the primitive streak/mesendoderm genes *BRACHYURY* (1585 fold vs. 182 fold), *GSC* (194 fold vs. 14 fold) and *MIXL1* (542 fold vs. 55 fold) (**Figure 5.2C**). This result suggested that *TCF7L1* could be suppressing BMP4 target genes and thus be acting as a key negative regulator of the primitive streak differentiation gene expression program. Moreover, these results pose the question that BMP4-induced down-regulation of *TCF7L1* might be an important, primary, event in the sequence of BMP4-induced differentiation of hESCs.

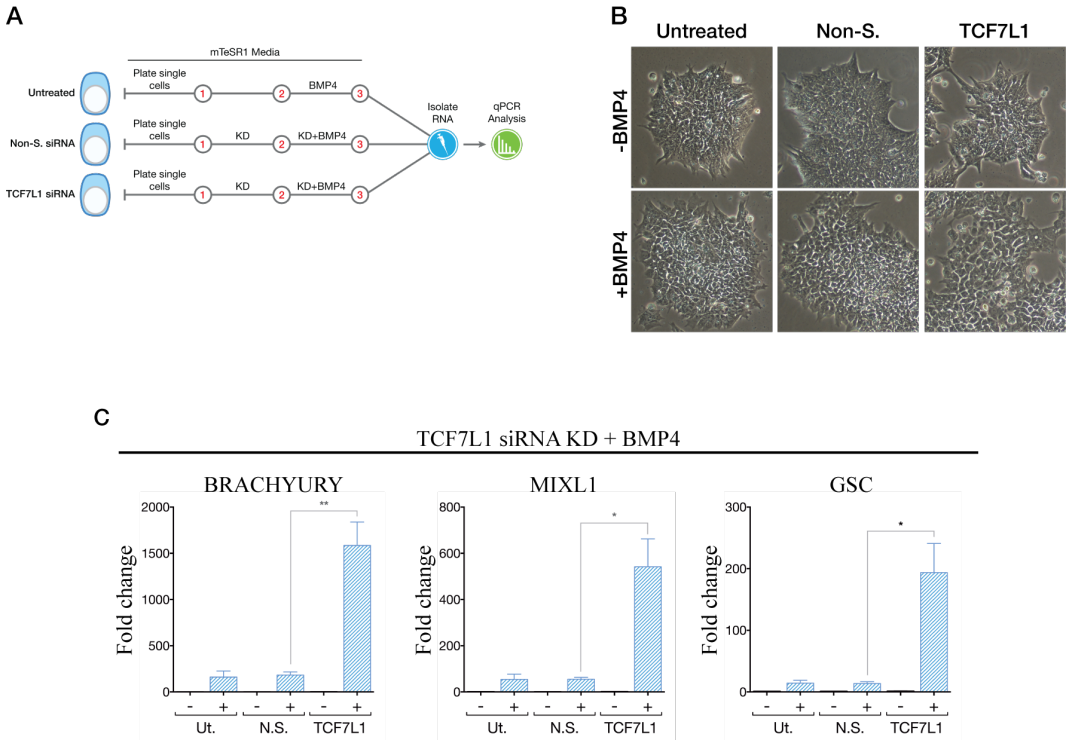


Figure 5.2: ***TCF7L1* knockdown enhances BMP4-induced gene expression**. **A**. Schematic of experiment. Human ESCs were treated with BMP4 (10 ng/ml) while performing *TCF7L1* siRNA (50 nM) knockdown under feeder-free conditions. Cells were harvested after 48 hours of siRNA knockdown and 24 hours of BMP4 treatment. Although absent from the schematic, we performed control experiments without BMP4 treatment for each condition. Red numbers indicate days (24 hour intervals) of the procedure. Cells were harvested for analysis on the third day. **B**. Simultaneous loss of *TCF7L1* and treatment with BMP4 causes pronounced, morphological changes in colonies. Phase contrast images (10X) show *TCF7L1* siRNA + BMP4-treated hESC colonies appear more flattened and differentiated than the controls. **C**. Quantitative PCR analysis showing synergistic up-regulation of primitive streak/mesendoderm markers when BMP4 treatment is combined with *TCF7L1* knockdown (n=3). Data is presented as fold change in expression relative to the untreated mTeSR1 condition. Statistics performed by Student's t-test and presented as mean +/-SEM (*p <0.05, **p <0.01, ***p <0.001).

A gain-of-function approach was taken to look more closely at a possible antagonistic relationship between *TCF7L1* and BMP4. Using the dox-inducible H9-TCF7L1 and H9-GFP hESC lines described previously, *TCF7L1* or GFP was over-expressed during a 24 hour period of BMP4-induced differentiation. The hypothesis was that if *TCF7L1* antagonizes BMP4-induced differentiation, then *TCF7L1* over-expression should impede hESC differentiation. As predicted, simultaneous *TCF7L1* over-expression and BMP4 treatment had the opposite effect on colony morphology. Cells over-expressing *TCF7L1* displayed tighter, compact, colony morphology, with well-defined boundaries; no effect was observed in the control GFP over-expressing cell line (**Figure 5.3**). Collectively, these results imply that *TCF7L1* functions to suppress genes induced by BMP4. Since *TCF7L1* is known to act as a transcriptional repressor, this suppression might occur at the level of BMP4-induction of of target gene transcription.

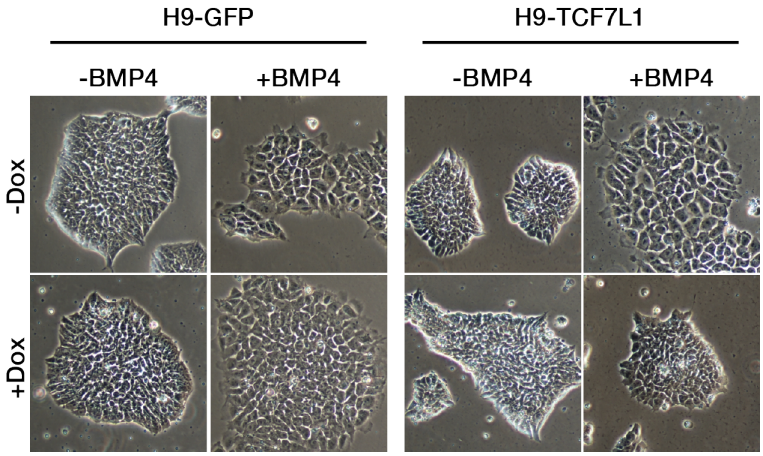


Figure 5.3: ***TCF7L1* over-expression opposes BMP4 differentiation**. *TCF7L1* over-expression stabilizes hESC colony morphology during BMP4-induced differentiation. H9-GFP and H9-TCF7L1 cells were grown under feeder-free conditions and treated with BMP4 for 24 hours while inducing GFP or *TCF7L1* with doxycycline (1 μg/ml). Phase contrast images (10X) shows *TCF7L1* over-expressing cells (right hand panel, bottom right image) have a far less differentiated morphology as compared to the control (left hand panel, bottom right image).

5.2.3 Analysis of *TCF7L1* regulation by *BMP4 in vivo*

If *Bmp4* down-regulates *Tcf7l1* in epiblast cells that are progressing through primitive streak differentiation, then there could be spatiotemporal evidence linking this regulatory interaction in embryos. To ascertain whether such a connection exists, our colleague Dr. Jackson Hoffman sectioned and stained embryonic day 6.5 (e6.5) mouse embryos for Bmp4-activated phospho-Smad1/5/8 (indicative of active BMP signaling) and Tcf7l1. Interestingly, confocal microscopy analysis showed an inverse staining pattern of phospho-Smad1/5/8 and Tcf7l1. Specifically, phospho-Smad1/5/8 staining was increased in the region of the primitive streak, whereas Tcf7l1 expression was lowest in this region (**Figure 5.4**) (Jackson Hoffman, Ph.D. personal communication associated with [47]). Furthermore, the primitive streak marker Brachyury was up-regulated in this region, affirming that these cells were in fact undergoing mesendoderm differentiation within the primitive streak (**Figure 5.4**). This experiment corroborates our *in vitro* data and further solidifies the model that Bmp4 is an upstream regulator of *Tcf7l1* during initiation of gastrulation.

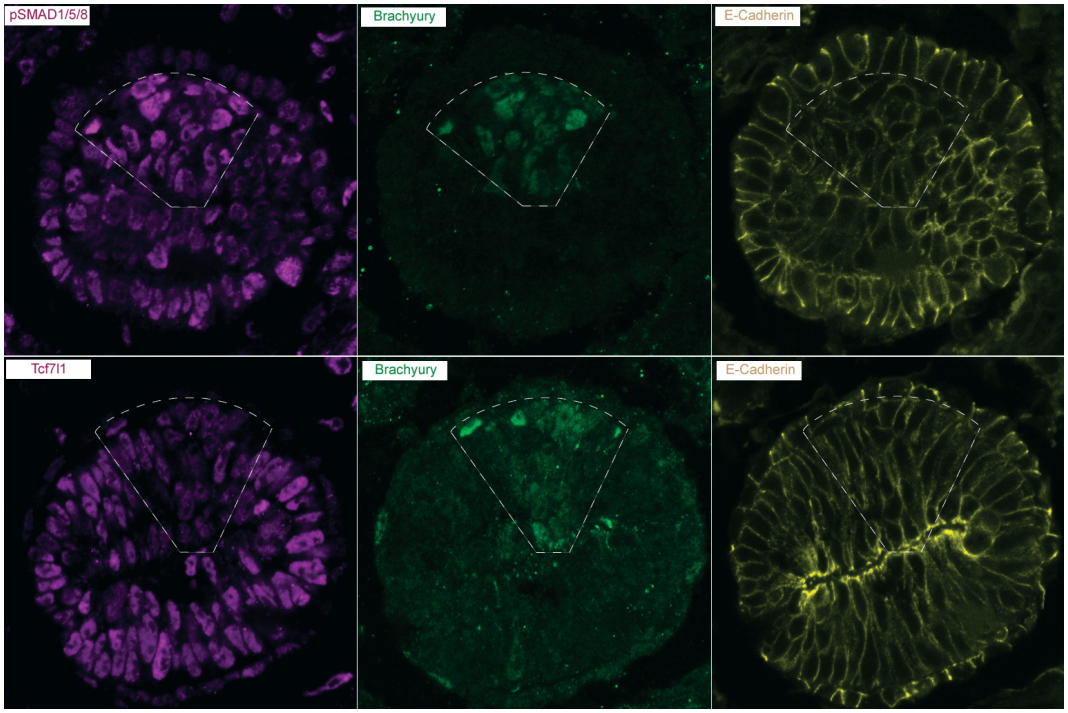


Figure 5.4: **Increased phospho-Smad1/5/8 and decreased Tcf7l1 within the nascent primitive streak**. Confocal microscopy images of stained sections of e6.5 embryos that are beginning to form the primitive streak. These are two transverse sections from the same embryo separated by about about 15 μm. The dashed line surrounds the nascent primitive streak region on the posterior side of the embryo where Tcf7l1 (purple - bottom left panel) expression is reduced and phospho-Smad1/5/8 (purple - top left panel) is increased. Brachyury (green - top and bottom middle panels) co-staining shows up-regulated expression in the primitive streak region. Plasma membranes are marked by E-cadherin (yellow - top and bottom right hand panels). *Image provided by Jackson Hoffman, Ph.D. as part of a personal communication associated with [47].*

5.3 Discussion

The data presented in this chapter strongly supports the idea that *TCF7L1* expression is responsive to *BMP4* signaling in hESCs. Specifically, I find that *TCF7L1* expression is robustly down-regulated during BMP4-induced differentiation (**Figure 5.1**). Gain- and loss-of-function assays demonstrated that *TCF7L1* antagonizes downstream BMP4-induced gene expression changes, suggesting that inactivation of *TCF7L1* is a critical

event during BMP4 differentiation of hESCs (**Figure 5.2**) (**Figure 5.3**). Moreover, *in vivo* immunofluorescence analysis shows the predicted inverse correlation between increased phospho-Smad1/5/8 and decreased Tcf7l1 within the domain of the developing primitive streak. These data support the hypothesis that BMP4 negatively regulates *TCF7L1* expression (**Figure 5.4**). Furthermore, these findings link *TCF7L1* to the overarching regulatory mechanisms governing the initiation of gastrulation during development.

It is important to note that these results are correlative and do not provide information on mechanism, thus I cannot conclude whether BMP4 signaling directly or indirectly suppresses *TCF7L1* expression. It is possible for phospho-SMAD1/5/8 transcription factors to be involved in repression of target genes, but a phospho-SMAD1/5/8 ChIP-seq dataset does not exist in hESCs, so it is not known whether these SMADs directly occupy regulatory sites in the *TCF7L1* locus. One possibility for indirect regulation is that activation of the BMP4 signaling cascade somehow prevents OCT4 from transcriptionally activating *TCF7L1*. This is a reasonable hypothesis because OCT4 binds 1.3 kb downstream of the *TCF7L1* transcription start site and OCT4 siRNA knockdown reduces *TCF7L1* mRNA expression by $\sim$50% (**Figure 2.5**). One problem with this scenario, however, is that *OCT4* expression is not affected by BMP4-induced differentiation in hESCs (**Figure 2.6A**). Furthermore, OCT4 protein expression is maintained in epiblast cells undergoing primitive streak differentiation [47]. Despite constant OCT4 expression, it is possible for BMP4 to trigger a post-translational mechanism that could prevent OCT4 activation of *TCF7L1* expression. A prime candidate for such a mechanism is SOX2, which heterodimerizes with OCT4 to co-regulate target genes. Like *TCF7L1*, BMP4 strongly down-regulates of *SOX2* in hESCs (**Figure 2.6A**) - which also has the same expression pattern as *TCF7L1* within the gastrulating mouse embryo [47]. It has been shown that loss of SOX2 binding to genomic targets decreases the transcriptional activity of OCT4/SOX2 target genes; therefore, a potential explanation is that BMP4 down-regulates SOX2, which then decreases the affinity of OCT4 for its *TCF7L1* binding site, resulting in transcriptional inactivation [95].

Support for this hypothesis comes from overlapping OCT4/SOX2 binding site within the same 1.3 kb region downstream of the *TCF7L1* transcription start site in hESCs and mESCs (**Figure 2.5C**) [65][26][60]. Sorting out whether BMP4 directly or indirectly down-regulates *TCF7L1* would likely be a challenging endeavor, but the hypothesis outlined here could be a solid starting point for future studies.

What also remains to be answered is how TCF7L1 antagonizes BMP4-directed differentiation. One intriguing possibility is that TCF7L1 directly binds to and prevents BMP4-activated phospho-SMAD1/5/8 from transcriptionally activating their target genes. This regulatory interaction is not unfounded as LEF/TCFs and SMADs have been shown to bind within close proximity of each other on target genes and to co-regulate gene expression [115][81][59]. Unfortunately, phospho-SMAD1/5/8 ChIP-seq has not been performed on BMP4 treated hESCs, thus this hypothesis has not yet been tested. When such a dataset becomes available in the future, a comparison between the TCF7L1 ChIP-seq data presented here and the phospho-SMAD1/5/8 ChIP-seq data should shed light on whether or not TCF7L1 directly antagonizes phospho-SMAD1/5/8 regulated transcription. Of course, it is equally conceivable that *TCF7L1* indirectly impedes BMP4 signaling. For example, *TCF7L1* could transcriptionally repress BMP4 antagonists like *NOGGIN* and *CER1*, which were identified as TCF7L1 targets by our ChIP-seq analysis. Nevertheless, defining how TCF7L1 regulates downstream BMP4 signaling events will be essential for understanding the underlying mechanisms of primitive streak differentiation *in vivo*.

Chapter 6

Overall Conclusions and Biological Implications

The studies I performed and describe here in this thesis represents the first in-depth study of LEF/TCFs in hESCs and identifies *TCF7L1* as the most highly expressed LEF/TCF transcription factor in hESCs. I found that the WNT signaling pathway is inactive in hESCs, corroborating more recent studies which mark WNT as an inducer of hESC differentiation[10][104][9][9][30]. Furthermore, because β-CATENIN is post-transcriptionally degraded and absent from the nucleus, WNT-target genes are presumably silent; thus, in accordance with previous reports, *TCF7L1* is likely recruiting repressor complexes (i.e., HDACs and TLE's) to silence these gene targets [24].

To date, the role of LEF/TCFs in the regulation of pluripotency mostly derives from studies done in mice [73][74][83][131][47]. In "naive" mESCs, *Tcf7l1* suppresses the pluripotency gene program (i.e., *Esrrb*, *Oct4* and *Nanog*) to ensure that cells are properly responsive to differentiation cues [66][89][137]. Genetic ablation of *Tcf7l1* in mESCs relieves repression of the "pro-pluripotency" program, thus delaying differentiation when the cells are placed

into differentiation conditions [89][42][47]. *In vivo* studies performed with *Tcf7l1* knockout (KO) mice indicate that *Tcf7l1* is required to coordinate axis specification with mesoderm differentiation and loss of *Tcf7l1* results in early embryonic lethality shortly after gastrulation [74][47]. These phenotypes imply that *Tcf7l1* has unique roles in pre-implantation ICM cells (i.e., mESCs) and post-implantation epiblast cells (i.e., hESCs). A major discovery of the work presented here is that *TCF7L1* acts to maintain hESCs pluripotency by suppressing primitive streak differentiation, a significant difference in function that either distinguishes mouse from human or "naive" from "poised" stem cell states.

At first glance our results seem to conflict with published reports of *Tcf7l1* function in mESCs. If *Tcf7l1* limits pluripotency in mESCs, then how can it stabilize the pluripotent state in hESCs? I propose that the results presented here favors the idea that *TCF7L1* actually functions similarly in both mESCs and hESCs. As described in **Section 1.6**, mESCs and hESCs represent two different embryonic stages and therefore have different growth factor requirements for pluripotency and differentiation [117][44][110][14][41][92][16][17]. Human ESC pluripotency depends on SMAD2/3 and PI3K/Akt signaling pathways, whereas mESC pluripotency is supported by activation of the LIF, Bmp4 and canonical Wnt signaling pathways [91][123][52][108]. Interestingly, loss of *Tcf7l1* in mESCs up-regulates *Bmp4* and the canonical WNT ligand *Wnt3a*; two ligands that act to maintain the pluripotent state of mESCs [26]. In my experiments, *BMP4* is one of the most highly up-regulated genes during *TCF7L1* siRNA knockdown suggesting that it suppresses pro-mesendoderm differentiation genes. The canonical WNT ligands *WNT3* and *WNT8A* were also up-regulated by *TCF7L1* knockdown in hESCs as determined by qPCR analysis (data not shown), but since they failed to pass the statistical tests of the microarray analysis, they were excluded from the results. I suspect that performing the siRNA knockdown in growth factor free conditions would allow for a more substantial up-regulation of these canonical WNT ligands. That being said, however, our TCF7L1 ChIP-seq data indicates that *WNT3* and *WNT8A* are both bound by TCF7L1 and they are also suppressed by *TCF7L1* over-expression during

PS differentiation (**Figure 4.7C**). Hence, I propose that *TCF7L1* regulates overlapping genes in mESCs and hESCs, but due to their different representative stages of development (ICM vs. epiblast) these two cell types respond to *TCF7L1* inactivation differently.

Based on these observations, I propose a unifying model of *TCF7L1* function during early development. I propose that *TCF7L1* regulates the transition of pluripotent cells to their next state during the pre- to post-implantation stages; i.e., ICM to epiblast transition, followed by the epiblast-to-mesendoderm transition during primitive streak formation. I hypothesize that *TCF7L1* is a key regulator of these transitions acting, in part, through suppression of autocrine BMP4 and canonical WNT signaling. This helps explain how "naive" and "primed" pluripotent cell types are differentially affected by *TCF7L1* inactivation. *In vivo*, "naive" pluripotent ICM cells express Bmp4 and have active canonical Wnt signaling [123][52][108]. Therefore, in "naive" mESCs, my model predicts that *Tcf7l1* would limit Wnt- and Bmp4-driven pluripotency signaling and that ablation of *Tcf7l1* would incidentally stabilize this state. It has been established that loss of *Tcf7l1* impedes mESC differentiation, but whether or not this is due to increased autocrine Bmp4 and Wnt signaling remains to be tested [26][47][107][89][129]. Later stage "primed" epiblast stem cells, however, respond very differently to canonical Wnt and Bmp4 stimulation, which stimulate mesendoderm differentiation and primitive streak formation. On the other hand, in "primed" hESCs, my data shows that *TCF7L1* represses BMP4 and canonical WNT ligand (*WNT8A* and *WNT3*) expression, thereby suppressing mesendoderm differentiation promoting autocrine signaling. I believe that this finding has important implications for understanding early human development, as well as the transition from ICM to mesendoderm.

If *TCF7L1* serves as a transcriptional barrier to primitive streak/mesendoderm differentiation, then one would predict that its expression would be diminished in the forming primitive streak. Indeed, recent *in vivo* mouse studies support this rationale. Using careful

spatio-temporal analysis, Hoffman and colleagues analyzed *Tcf7l1* expression *in vivo* at sequential stages of pre-streak, early-streak and mid-streak stages leading up to gastrulation in mice [47]. They found that Tcf7l1 was uniformly expressed throughout the pre-streak epiblast before the onset of gastrulation; yet as embryogenesis progressed from the early-streak to the mid-streak stage Tcf7l1 expression was down-regulated in the epiblast cells destined to give rise to the primitive streak. While *Tcf7l1* expression in wild-type mouse embryos corroborated our prediction, our finding that *TCF7L1* represses mesendoderm differentiation differs with the functional analyses performed using *Tcf7l1* KO embryos. Data from *Tcf7l1* knockout embryos suggests that *Tcf7l1* is not necessary for mesendoderm gene expression; rather it regulates the transition of ICM to epiblast and ensures proper coupling of gastrulation and timing of mesoderm differentiation. In contrast, I find that *TCF7L1* represses mesendoderm differentiation in epiblast-like hESCs and I predict *TCF7L1* is needed in the post-implantation mouse epiblast to stabilize pluripotency and inhibit differentiation. I propose that the basis for the apparent discrepancy between the knockout analyses and the data presented here, tracks with the previously described role of *Tcf7l1* function in mESCs. That is, loss of *Tcf7l1* in "naive" mESCs delays their ability to differentiate. I reason that differentiation is delayed because of an increase in *Wnt3* expression and subsequent stimulation of autocrine Wnt signaling; a signal that, along with Nodal and Bmp4, acts to maintain "naive" ICM cells and mESCs in an early, pluripotent state. The propensity of mESCs lacking *Tcf7l1* to resist differentiation could imply that, in *Tcf7l1* knockout mice, cells of the ICM are slow to transition from ICM-to-epiblast because of sustained activation of the Wnt pathway. As a consequence, *Tcf7l1*-/- ICM cells may be unsynchronized and may not have fully committed to the epiblast-state. Partial or uncoordinated transitioning from the ICM to epiblast may result in pluripotent cells incapable of undergoing a timely or coordinated differentiation response to mesendoderm inductive signals (i.e., Wnt3, Bmp4 and Nodal). Support for *TCF7L1* as a repressor of mesendoderm differentiation also comes from studies analyzing post-gastrulation *Tcf7l1* knockout mouse embryo's [74]. These

embryos show ectopic and expanded expression of primitive streak/mesendoderm marker genes *Brachyury* and *Foxa2* and in some cases even duplications of the primitive streak [74]. This suggests that *Tcf7l1* is involved primitive streak differentiation events as proposed here and emphasizes the importance of *Tcf7l1* in the progression of the ICM to epiblast to mesendoderm.

In addition to proposing a unifying function of *TCF7L1* in pluripotent stem cells, this work also identified *TCF7L1* as potential novel downstream regulator of the *BMP4* signaling pathway. Because there are over 30 other members of the BMP family, I propose that *TCF7L1* might also modulate other BMP-directed processes during embryogenesis, organogenesis and tissue homeostasis [48][49][32]. For instance, during development BMPs direct multipotent neural crest cell differentiation into cells of the peripheral nervous system [71][98]. They also instruct mesenchymal stem cell differentiation into chondrocytes, adipocytes, fibroblasts, myoblasts and osteoblasts [49]. Moreover, BMP signaling morphogens function in organogenesis of lung, heart, teeth, skin, gut and kidney [49][57]. In all, this discovery could have important implications for understanding how these other BMP-regulated processes are controlled and merit further investigation.

The research presented in this thesis was an remarkable challenge, but also a incredible privilege. Not many young scientists get an opportunity to study, discover and present new truths about one of the most pivotal processes in all of mammalian development, gastrulation. It comes with great pride and honor to propose a model of *TCF7L1* function during the initial stages of gastrulation (**Figure 6.1**). I would also like to highlight an important distinction that came to light over the course of this research project. Human and mouse ESCs should not be treated as two exclusive research models, rather two complementary models representing two ends of the pluripotency spectrum: the "naive" inner cell mass stage of pluripotency and the later "primed" epiblast stage of pluripotency. In this regard, both cell types can be used to study the mechanisms regulating the transition from the inner

cell mass stage to the epiblast stage of development. In closing, it is my hope that this work will make a meaningful contribution to our understanding of pluripotency and early development.

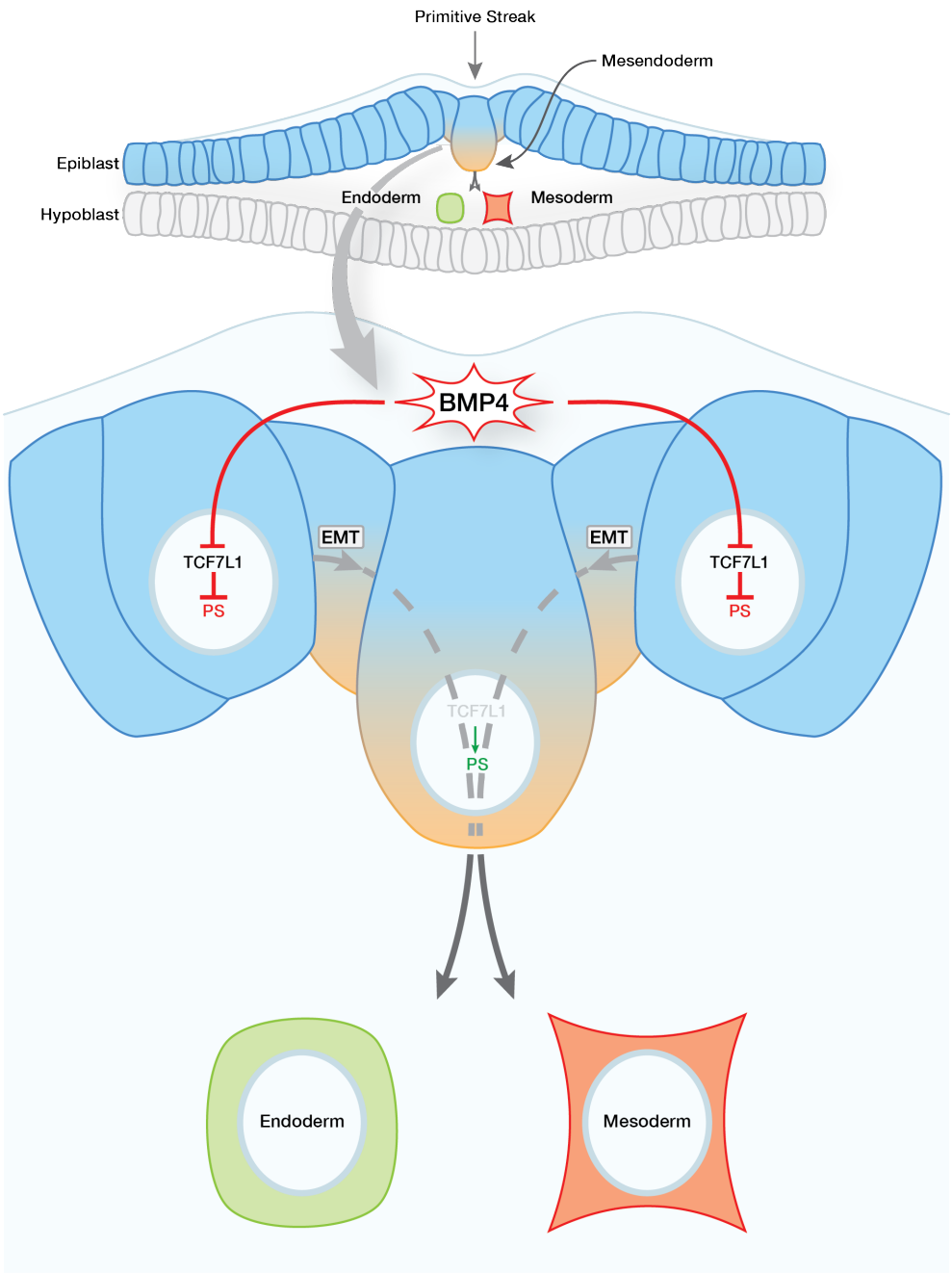


Figure 6.1: **Model of *TCF7L1* function in pluripotency and differentiation of epiblast stem cells**. In this model, epiblast cells (blue) *in vivo* - which can be thought of as synonymous with hESCs *in vitro* - remain in an undifferentiated state until the necessary inductive signals are received. In the undifferentiated-state, *TCF7L1* represses genes involved in primitive streak/mesendoderm (blue-orange gradient) differentiation. At the onset of gastrulation, as BMP4 triggers *TCF7L1* down-regulation and stimulates primitive streak/mesendoderm genes expression. The illustrated effect of BMP4 on *TCF7L1* should not be taken literally rather correlatively, since it is unknown whether this is a direct or indirect regulatory mechanism.

Bibliography

[1] L. Arce, N. N. Yokoyama, and M. L. Waterman. Diversity of LEF/TCF action in development and disease. *Oncogene*, 25(57):7492–7504, Dec 2006.

[2] S. J. Arnold and E. J. Robertson. Making a commitment: cell lineage allocation and axis patterning in the early mouse embryo. *Nat. Rev. Mol. Cell Biol.*, 10(2):91–103, Feb 2009.

[3] F. A. Atcha, A. Syed, B. Wu, N. P. Hoverter, N. N. Yokoyama, J. H. Ting, J. E. Munguia, H. J. Mangalam, J. L. Marsh, and M. L. Waterman. A unique DNA binding domain converts T-cell factors into strong Wnt effectors. *Mol. Cell. Biol.*, 27(23):8352–8363, Dec 2007.

[4] A. A. Avilion, S. K. Nicolis, L. H. Pevny, L. Perez, N. Vivian, and R. Lovell-Badge. Multipotent cell lineages in early mouse development depend on SOX2 function. *Genes Dev.*, 17(1):126–140, Jan 2003.

[5] H. Baharvand, N. Z. Mehrjardi, M. Hatami, S. Kiani, M. Rao, and M. M. Haghighi. Neural differentiation from human embryonic stem cells in a defined adherent culture condition. *Int. J. Dev. Biol.*, 51(5):371–378, 2007.

[6] S. Beck, J. A. Le Good, M. Guzman, N. Ben Haim, K. Roy, F. Beermann, and D. B. Constam. Extraembryonic proteases regulate Nodal signalling during gastrulation. *Nat. Cell Biol.*, 4(12):981–985, Dec 2002.

[7] H. Beppu, M. Kawabata, T. Hamamoto, A. Chytil, O. Minowa, T. Noda, and K. Miyazono. BMP type II receptor is required for gastrulation and early development of mouse embryos. *Dev. Biol.*, 221(1):249–258, May 2000.

[8] A. S. Bernardo, T. Faial, L. Gardner, K. K. Niakan, D. Ortmann, C. E. Senner, E. M. Callery, M. W. Trotter, M. Hemberger, J. C. Smith, L. Bardwell, A. Moffett, and R. A. Pedersen. BRACHYURY and CDX2 mediate BMP-induced differentiation of human and mouse pluripotent stem cells into embryonic and extraembryonic lineages. *Cell Stem Cell*, 9(2):144–155, Aug 2011.

[9] T. A. Blauwkamp, S. Nigam, R. Ardehali, I. L. Weissman, and R. Nusse. Endogenous Wnt signalling in human embryonic stem cells generates an equilibrium of distinct lineage-specified progenitors. *Nat Commun*, 3:1070, 2012.

[10] H. K. Bone, A. S. Nelson, C. E. Goldring, D. Tosh, and M. J. Welham. A novel chemically directed route for the generation of definitive endoderm from human embryonic stem cells based on inhibition of GSK-3. *J. Cell. Sci.*, 124(Pt 12):1992–2000, Jun 2011.

[11] L. A. Boyer, T. I. Lee, M. F. Cole, S. E. Johnstone, S. S. Levine, J. P. Zucker, M. G. Guenther, R. M. Kumar, H. L. Murray, R. G. Jenner, D. K. Gifford, D. A. Melton, R. Jaenisch, and R. A. Young. Core transcriptional regulatory circuitry in human embryonic stem cells. *Cell*, 122(6):947–956, Sep 2005.

[12] R. Brandenberger, H. Wei, S. Zhang, S. Lei, J. Murage, G. J. Fisk, Y. Li, C. Xu, R. Fang, K. Guegler, M. S. Rao, R. Mandalam, J. Lebkowski, and L. W. Stanton. Transcriptome characterization elucidates signaling networks that control human ES cell growth and differentiation. *Nat. Biotechnol.*, 22(6):707–716, Jun 2004.

[13] J. Brennan, C. C. Lu, D. P. Norris, T. A. Rodriguez, R. S. Beddington, and E. J. Robertson. Nodal signalling in the epiblast patterns the early mouse embryo. *Nature*, 411(6840):965–969, Jun 2001.

[14] I. G. Brons, L. E. Smithers, M. W. Trotter, P. Rugg-Gunn, B. Sun, S. M. Chuva de Sousa Lopes, S. K. Howlett, A. Clarkson, L. Ahrlund-Richter, R. A. Pedersen, and L. Vallier. Derivation of pluripotent epiblast stem cells from mammalian embryos. *Nature*, 448(7150):191–195, Jul 2007.

[15] S. Brown, A. Teo, S. Pauklin, N. Hannan, C. H. Cho, B. Lim, L. Vardy, N. R. Dunn, M. Trotter, R. Pedersen, and L. Vallier. Activin/Nodal signaling controls divergent transcriptional networks in human embryonic stem cells and in endoderm progenitors. *Stem Cells*, 29(8):1176–1185, Aug 2011.

[16] C. Buecker, H. H. Chen, J. M. Polo, L. Daheron, L. Bu, T. S. Barakat, P. Okwieka, A. Porter, J. Gribnau, K. Hochedlinger, and N. Geijsen. A murine ESC-like state facilitates transgenesis and homologous recombination in human pluripotent stem cells. *Cell Stem Cell*, 6(6):535–546, Jun 2010.

[17] C. Buecker and N. Geijsen. Different flavors of pluripotency, molecular mechanisms, and practical implications. *Cell Stem Cell*, 7(5):559–564, Nov 2010.

[18] A. Camus, A. Perea-Gomez, A. Moreau, and J. Collignon. Absence of Nodal signaling promotes precocious neural differentiation in the mouse embryo. *Dev. Biol.*, 295(2):743–755, Jul 2006.

[19] I. Chambers, D. Colby, M. Robertson, J. Nichols, S. Lee, S. Tweedie, and A. Smith. Functional expression cloning of Nanog, a pluripotency sustaining factor in embryonic stem cells. *Cell*, 113(5):643–655, May 2003.

[20] B. Chen, M. E. Dodge, W. Tang, J. Lu, Z. Ma, C. W. Fan, S. Wei, W. Hao, J. Kilgore, N. S. Williams, M. G. Roth, J. F. Amatruda, C. Chen, and L. Lum. Small molecule-mediated disruption of Wnt-dependent signaling in tissue regeneration and cancer. *Nat. Chem. Biol.*, 5(2):100–107, Feb 2009.

[21] D. Chen, M. Zhao, and G. R. Mundy. Bone morphogenetic proteins. *Growth Factors*, 22(4):233–241, Dec 2004.

[22] X. Chen, H. Xu, P. Yuan, F. Fang, M. Huss, V. B. Vega, E. Wong, Y. L. Orlov, W. Zhang, J. Jiang, Y. H. Loh, H. C. Yeo, Z. X. Yeo, V. Narang, K. R. Govindarajan, B. Leong, A. Shahab, Y. Ruan, G. Bourque, W. K. Sung, N. D. Clarke, C. L. Wei, and H. H. Ng. Integration of external signaling pathways with the core transcriptional network in embryonic stem cells. *Cell*, 133(6):1106–1117, Jun 2008.

[23] V. Chickarmane, C. Troein, U. A. Nuber, H. M. Sauro, and C. Peterson. Transcriptional dynamics of the embryonic stem cell switch. *PLoS Comput. Biol.*, 2(9):e123, Sep 2006.

[24] J. V. Chodaparambil, K. T. Pate, M. R. Hepler, B. P. Tsai, U. M. Muthurajan, K. Luger, M. L. Waterman, and W. I. Weis. Molecular functions of the TLE tetramerization domain in Wnt target gene repression. *EMBO J.*, 33(7):719–731, Apr 2014.

[25] B. Ciruna and J. Rossant. FGF signaling regulates mesoderm cell fate specification and morphogenetic movement at the primitive streak. *Dev. Cell*, 1(1):37–49, Jul 2001.

[26] M. F. Cole, S. E. Johnstone, J. J. Newman, M. H. Kagey, and R. A. Young. Tcf3 is an integral component of the core regulatory circuitry of embryonic stem cells. *Genes Dev.*, 22(6):746–755, Mar 2008.

[27] B. J. Conley, J. C. Young, A. O. Trounson, and R. Mollard. Derivation, propagation and differentiation of human embryonic stem cells. *Int. J. Biochem. Cell Biol.*, 36(4):555–567, Apr 2004.

[28] F. L. Conlon, K. M. Lyons, N. Takaesu, K. S. Barth, A. Kispert, B. Herrmann, and E. J. Robertson. A primary requirement for nodal in the formation and maintenance of the primitive streak in the mouse. *Development*, 120(7):1919–1928, Jul 1994.

[29] K. A. D'Amour, A. D. Agulnick, S. Eliazer, O. G. Kelly, E. Kroon, and E. E. Baetge. Efficient differentiation of human embryonic stem cells to definitive endoderm. *Nat. Biotechnol.*, 23(12):1534–1541, Dec 2005.

[30] K. C. Davidson, A. M. Adams, J. M. Goodson, C. E. McDonald, J. C. Potter, J. D. Berndt, T. L. Biechele, R. J. Taylor, and R. T. Moon. Wnt/-catenin signaling promotes differentiation, not self-renewal, of human embryonic stem cells and is repressed by Oct4. *Proc. Natl. Acad. Sci. U.S.A.*, 109(12):4485–4490, Mar 2012.

[31] G. Dravid, Z. Ye, H. Hammond, G. Chen, A. Pyle, P. Donovan, X. Yu, and L. Cheng. Defining the role of Wnt/beta-catenin signaling in the survival, proliferation, and self-renewal of human embryonic stem cells. *Stem Cells*, 23(10):1489–1501, 2005.

[32] P. Ducy and G. Karsenty. The family of bone morphogenetic proteins. *Kidney Int.*, 57(6):2207–2214, Jun 2000.

[33] P. Dvorak and A. Hampl. Basic fibroblast growth factor and its receptors in human embryonic stem cells. *Folia Histochem. Cytobiol.*, 43(4):203–208, 2005.

[34] L. Eiselleova, K. Matulka, V. Kriz, M. Kunova, Z. Schmidtova, J. Neradil, B. Tichy, D. Dvorakova, S. Pospisilova, A. Hampl, and P. Dvorak. A complex role for FGF-2 in self-renewal, survival, and adhesion of human embryonic stem cells. *Stem Cells*, 27(8):1847–1857, Aug 2009.

[35] M. J. Evans and M. H. Kaufman. Establishment in culture of pluripotential cells from mouse embryos. *Nature*, 292(5819):154–156, Jul 1981.

[36] G. Friedrich and P. Soriano. Promoter traps in embryonic stem cells: a genetic screen to identify and mutate developmental genes in mice. *Genes Dev.*, 5(9):1513–1523, Sep 1991.

[37] T. S. Furey. ChIP-seq and beyond: new and improved methodologies to detect and characterize protein-DNA interactions. *Nat. Rev. Genet.*, 13(12):840–852, Dec 2012.

[38] A. Gagliardi, N. P. Mullin, Z. Ying Tan, D. Colby, A. I. Kousa, F. Halbritter, J. T. Weiss, A. Felker, K. Bezstarosti, R. Favaro, J. Demmers, S. K. Nicolis, S. R. Tomlinson, R. A. Poot, and I. Chambers. A direct physical interaction between Nanog and Sox2 regulates embryonic stem cell self-renewal. *EMBO J.*, 32(16):2231–2247, Aug 2013.

[39] Z. Gao, J. L. Cox, J. M. Gilmore, B. D. Ormsbee, S. K. Mallanna, M. P. Washburn, and A. Rizzino. Determination of protein interactome of transcription factor Sox2 in embryonic stem cells engineered for inducible expression of four reprogramming factors. *J. Biol. Chem.*, 287(14):11384–11397, Mar 2012.

[40] V. Graham, J. Khudyakov, P. Ellis, and L. Pevny. SOX2 functions to maintain neural progenitor identity. *Neuron*, 39(5):749–765, Aug 2003.

[41] B. Greber, G. Wu, C. Bernemann, J. Y. Joo, D. W. Han, K. Ko, N. Tapia, D. Sabour, J. Sterneckert, P. Tesar, and H. R. Scholer. Conserved and divergent roles of FGF signaling in mouse epiblast stem cells and human embryonic stem cells. *Cell Stem Cell*, 6(3):215–226, Mar 2010.

[42] G. Guo, Y. Huang, P. Humphreys, X. Wang, and A. Smith. A PiggyBac-based recessive screening method to identify pluripotency regulators. *PLoS ONE*, 6(4):e18189, 2011.

[43] X. Guo and X. F. Wang. Signaling cross-talk between TGF-beta/BMP and other pathways. *Cell Res.*, 19(1):71–88, Jan 2009.

[44] D. W. Han, N. Tapia, J. Y. Joo, B. Greber, M. J. Arauzo-Bravo, C. Bernemann, K. Ko, G. Wu, M. Stehling, J. T. Do, and H. R. Scholer. Epiblast stem cell subpopulations represent mouse embryos of distinct pregastrulation stages. *Cell*, 143(4):617–627, Nov 2010.

[45] A. H. Hart, L. Hartley, M. Ibrahim, and L. Robb. Identification, cloning and expression analysis of the pluripotency promoting Nanog genes in mouse and human. *Dev. Dyn.*, 230(1):187–198, May 2004.

[46] S. Heinz, C. Benner, N. Spann, E. Bertolino, Y. C. Lin, P. Laslo, J. X. Cheng, C. Murre, H. Singh, and C. K. Glass. Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. *Mol. Cell*, 38(4):576–589, May 2010.

[47] J. A. Hoffman, C. I. Wu, and B. J. Merrill. Tcf7l1 prepares epiblast cells in the gastrulating mouse embryo for lineage specification. *Development*, 140(8):1665–1675, Apr 2013.

[48] B. L. Hogan. Bone morphogenetic proteins in development. *Curr. Opin. Genet. Dev.*, 6(4):432–438, Aug 1996.

[49] B. L. Hogan. Bone morphogenetic proteins: multifunctional regulators of vertebrate development. *Genes Dev.*, 10(13):1580–1594, Jul 1996.

[50] N. P. Hoverter, J. H. Ting, S. Sundaresh, P. Baldi, and M. L. Waterman. A WNT/p21 circuit directed by the C-clamp, a sequence-specific DNA binding domain in TCFs. *Mol. Cell. Biol.*, 32(18):3648–3662, Sep 2012.

[51] R. Jauch, C. K. Ng, K. S. Saikatendu, R. C. Stevens, and P. R. Kolatkar. Crystal structure and DNA binding of the homeodomain of the stem cell transcription factor Nanog. *J. Mol. Biol.*, 376(3):758–770, Feb 2008.

[52] C. Kemp, E. Willems, S. Abdo, L. Lambiv, and L. Leyns. Expression of all Wnt genes and their secreted antagonists during mouse blastocyst and postimplantation development. *Dev. Dyn.*, 233(3):1064–1075, Jul 2005.

[53] C. H. Kim, T. Oda, M. Itoh, D. Jiang, K. B. Artinger, S. C. Chandrasekharappa, W. Driever, and A. B. Chitnis. Repressor activity of Headless/Tcf3 is essential for vertebrate head formation. *Nature*, 407(6806):913–916, Oct 2000.

[54] S. W. Kim, S. J. Yoon, E. Chuong, C. Oyolu, A. E. Wills, R. Gupta, and J. Baker. Chromatin and transcriptional signatures for Nodal signaling during endoderm formation in hESCs. *Dev. Biol.*, 357(2):492–504, Sep 2011.

[55] D. Kimelman and W. Xu. beta-catenin destruction complex: insights and questions from a structural perspective. *Oncogene*, 25(57):7482–7491, Dec 2006.

[56] J. A. Le Good, K. Joubin, A. J. Giraldez, N. Ben-Haim, S. Beck, Y. Chen, A. F. Schier, and D. B. Constam. Nodal stability determines signaling range. *Curr. Biol.*, 15(1):31–36, Jan 2005.

[57] J. H. Lee, D. H. Bhang, A. Beede, T. L. Huang, B. R. Stripp, K. D. Bloch, A. J. Wagers, Y. H. Tseng, S. Ryeom, and C. F. Kim. Lung stem cell differentiation in mice directed by endothelial cells via a BMP4-NFATc1-thrombospondin-1 axis. *Cell*, 156(3):440–455, Jan 2014.

[58] J. J. Li, C. R. Jiang, J. B. Brown, H. Huang, and P. J. Bickel. Sparse linear modeling of next-generation mRNA sequencing (RNA-Seq) data for isoform discovery and abundance estimation. *Proc. Natl. Acad. Sci. U.S.A.*, 108(50):19867–19872, Dec 2011.

[59] S. K. Lim and F. M. Hoffmann. Smad4 cooperates with lymphoid enhancer-binding factor 1/T cell-specific factor to increase c-myc expression in the absence of TGF-beta signaling. *Proc. Natl. Acad. Sci. U.S.A.*, 103(49):18580–18585, Dec 2006.

[60] R. Lister, M. Pelizzola, R. H. Dowen, R. D. Hawkins, G. Hon, J. Tonti-Filippini, J. R. Nery, L. Lee, Z. Ye, Q. M. Ngo, L. Edsall, J. Antosiewicz-Bourget, R. Stewart, V. Ruotti, A. H. Millar, J. A. Thomson, B. Ren, and J. R. Ecker. Human DNA methylomes at base resolution show widespread epigenomic differences. *Nature*, 462(7271):315–322, Nov 2009.

[61] P. Liu, M. Wakamiya, M. J. Shea, U. Albrecht, R. R. Behringer, and A. Bradley. Requirement for Wnt3 in vertebrate axis formation. *Nat. Genet.*, 22(4):361–365, Aug 1999.

[62] K. J. Livak and T. D. Schmittgen. Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method. *Methods*, 25(4):402–408, Dec 2001.

[63] C. Y. Logan and R. Nusse. The Wnt signaling pathway in development and disease. *Annu. Rev. Cell Dev. Biol.*, 20:781–810, 2004.

[64] T. E. Ludwig, M. E. Levenstein, J. M. Jones, W. T. Berggren, E. R. Mitchen, J. L. Frane, L. J. Crandall, C. A. Daigh, K. R. Conard, M. S. Piekarczyk, R. A. Llanas, and J. A. Thomson. Derivation of human embryonic stem cells in defined conditions. *Nat. Biotechnol.*, 24(2):185–187, Feb 2006.

[65] A. Marson, S. S. Levine, M. F. Cole, G. M. Frampton, T. Brambrink, S. Johnstone, M. G. Guenther, W. K. Johnston, M. Wernig, J. Newman, J. M. Calabrese, L. M. Dennis, T. L. Volkert, S. Gupta, J. Love, N. Hannett, P. A. Sharp, D. P. Bartel, R. Jaenisch, and R. A. Young. Connecting microRNA genes to the core transcriptional regulatory circuitry of embryonic stem cells. *Cell*, 134(3):521–533, Aug 2008.

[66] G. Martello, T. Sugimoto, E. Diamanti, A. Joshi, R. Hannah, S. Ohtsuka, B. Gottgens, H. Niwa, and A. Smith. Esrrb is a pivotal target of the Gsk3/Tcf3 axis regulating embryonic stem cell self-renewal. *Cell Stem Cell*, 11(4):491–504, Oct 2012.

[67] B. L. Martin and D. Kimelman. Canonical Wnt signaling dynamically controls multiple stem cell fate decisions during vertebrate body formation. *Dev. Cell*, 22(1):223–232, Jan 2012.

[68] G. R. Martin. Isolation of a pluripotent cell line from early mouse embryos cultured in medium conditioned by teratocarcinoma stem cells. *Proc. Natl. Acad. Sci. U.S.A.*, 78(12):7634–7638, Dec 1981.

[69] A. B. McLean, K. A. D'Amour, K. L. Jones, M. Krishnamoorthy, M. J. Kulik, D. M. Reynolds, A. M. Sheppard, H. Liu, Y. Xu, E. E. Baetge, and S. Dalton. Activin a efficiently specifies definitive endoderm from human embryonic stem cells only when phosphatidylinositol 3-kinase signaling is suppressed. *Stem Cells*, 25(1):29–38, Jan 2007.

[70] C. Y. McLean, D. Bristor, M. Hiller, S. L. Clarke, B. T. Schaar, C. B. Lowe, A. M. Wenger, and G. Bejerano. GREAT improves functional interpretation of cis-regulatory regions. *Nat. Biotechnol.*, 28(5):495–501, May 2010.

[71] M. F. Mehler, P. C. Mabie, D. Zhang, and J. A. Kessler. Bone morphogenetic proteins in the nervous system. *Trends Neurosci.*, 20(7):309–317, Jul 1997.

[72] A. Mehra and J. L. Wrana. TGF-beta and the Smad signal transduction pathway. *Biochem. Cell Biol.*, 80(5):605–622, 2002.

[73] B. J. Merrill, U. Gat, R. DasGupta, and E. Fuchs. Tcf3 and Lef1 regulate lineage differentiation of multipotent stem cells in skin. *Genes Dev.*, 15(13):1688–1705, Jul 2001.

[74] B. J. Merrill, H. A. Pasolli, L. Polak, M. Rendl, M. J. Garcia-Garcia, K. V. Anderson, and E. Fuchs. Tcf3: a transcriptional regulator of axis induction in the early embryo. *Development*, 131(2):263–274, Jan 2004.

[75] Y. Mishina, A. Suzuki, N. Ueno, and R. R. Behringer. Bmpr encodes a type I bone morphogenetic protein receptor that is essential for gastrulation during mouse embryogenesis. *Genes Dev.*, 9(24):3027–3037, Dec 1995.

[76] K. Mitsui, Y. Tokuzawa, H. Itoh, K. Segawa, M. Murakami, K. Takahashi, M. Maruyama, M. Maeda, and S. Yamanaka. The homeoprotein Nanog is required for maintenance of pluripotency in mouse epiblast and ES cells. *Cell*, 113(5):631–642, May 2003.

[77] K. Miyazono, Y. Kamiya, and M. Morikawa. Bone morphogenetic protein receptors and signal transduction. *J. Biochem.*, 147(1):35–51, Jan 2010.

[78] A. Moustakas and C. H. Heldin. The regulation of TGFbeta signal transduction. *Development*, 136(22):3699–3714, Nov 2009.

[79] A. C. Mullen, D. A. Orlando, J. J. Newman, J. Loven, R. M. Kumar, S. Bilodeau, J. Reddy, M. G. Guenther, R. P. DeKoter, and R. A. Young. Master transcription factors determine cell-type-specific responses to TGF- signaling. *Cell*, 147(3):565–576, Oct 2011.

[80] N. P. Mullin, A. Yates, A. J. Rowe, B. Nijmeijer, D. Colby, P. N. Barlow, M. D. Walkinshaw, and I. Chambers. The pluripotency rheostat Nanog functions as a dimer. *Biochem. J.*, 411(2):227–231, Apr 2008.

[81] N. Nakano, S. Itoh, Y. Watanabe, K. Maeyama, F. Itoh, and M. Kato. Requirement of TCF7L2 for TGF-beta-dependent transcriptional activation of the TMEPAI gene. *J. Biol. Chem.*, 285(49):38023–38033, Dec 2010.

[82] K. Nakayama, Y. Tamura, M. Suzawa, S. Harada, S. Fukumoto, M. Kato, K. Miyazono, G. A. Rodan, Y. Takeuchi, and T. Fujita. Receptor tyrosine kinases inhibit bone morphogenetic protein-Smad responsive promoter activity and differentiation of murine MC3T3-E1 osteoblast-like cells. *J. Bone Miner. Res.*, 18(5):827–835, May 2003.

[83] H. Nguyen, B. J. Merrill, L. Polak, M. Nikolova, M. Rendl, T. M. Shaver, H. A. Pasolli, and E. Fuchs. Tcf3 and Tcf4 are essential for long-term homeostasis of skin epithelia. *Nat. Genet.*, 41(10):1068–1075, Oct 2009.

[84] K. K. Niakan, J. Han, R. A. Pedersen, C. Simon, and R. A. Pera. Human pre-implantation embryo development. *Development*, 139(5):829–841, Mar 2012.

[85] J. Nichols and A. Smith. Naive and primed pluripotent states. *Cell Stem Cell*, 4(6):487–492, Jun 2009.

[86] C. Niehrs. The complex world of WNT receptor signalling. *Nat. Rev. Mol. Cell Biol.*, 13(12):767–779, Dec 2012.

[87] M. Pardo, B. Lang, L. Yu, H. Prosser, A. Bradley, M. M. Babu, and J. Choudhary. An expanded Oct4 interaction network: implications for stem cell biology, development, and disease. *Cell Stem Cell*, 6(4):382–395, Apr 2010.

[88] M. F. Pera and P. P. Tam. Extrinsic regulation of pluripotent stem cells. *Nature*, 465(7299):713–720, Jun 2010.

[89] L. Pereira, F. Yi, and B. J. Merrill. Repression of Nanog gene transcription by Tcf3 limits embryonic stem cell self-renewal. *Mol. Cell. Biol.*, 26(20):7479–7491, Oct 2006.

[90] A. D. Pyle, L. F. Lock, and P. J. Donovan. Neurotrophins mediate human embryonic stem cell survival. *Nat. Biotechnol.*, 24(3):344–350, Mar 2006.

[91] X. Qi, T. G. Li, J. Hao, J. Hu, J. Wang, H. Simmons, S. Miura, Y. Mishina, and G. Q. Zhao. BMP4 supports self-renewal of embryonic stem cells by inhibiting mitogen-activated protein kinase pathways. *Proc. Natl. Acad. Sci. U.S.A.*, 101(16):6027–6032, Apr 2004.

[92] M. Rao. Conserved and divergent paths that regulate self-renewal in mouse and human embryonic stem cells. *Dev. Biol.*, 275(2):269–286, Nov 2004.

[93] B. E. Reubinoff, M. F. Pera, C. Y. Fong, A. Trounson, and A. Bongso. Embryonic stem cell lines from human blastocysts: somatic differentiation in vitro. *Nat. Biotechnol.*, 18(4):399–404, Apr 2000.

[94] M. Rhinn and P. Dolle. Retinoic acid signalling during development. *Development*, 139(5):843–858, Mar 2012.

[95] D. J. Rodda, J. L. Chew, L. H. Lim, Y. H. Loh, B. Wang, H. H. Ng, and P. Robson. Transcriptional regulation of nanog by OCT4 and SOX2. *J. Biol. Chem.*, 280(26):24731–24737, Jul 2005.

[96] N. Sato, L. Meijer, L. Skaltsounis, P. Greengard, and A. H. Brivanlou. Maintenance of pluripotency in human and mouse embryonic stem cells through activation of Wnt signaling by a pharmacological GSK-3-specific inhibitor. *Nat. Med.*, 10(1):55–63, Jan 2004.

[97] B. Schmierer and C. S. Hill. TGFbeta-SMAD signal transduction: molecular specificity and functional flexibility. *Nat. Rev. Mol. Cell Biol.*, 8(12):970–982, Dec 2007.

[98] N. M. Shah, A. K. Groves, and D. J. Anderson. Alternative neural crest cell fates are instructively promoted by TGFbeta superfamily members. *Cell*, 85(3):331–343, May 1996.

[99] M. M. Shen. Nodal signaling: developmental roles and regulation. *Development*, 134(6):1023–1034, Mar 2007.

[100] A. M. Singh, D. Reynolds, T. Cliff, S. Ohtsuka, A. L. Mattheyses, Y. Sun, L. Menendez, M. Kulik, and S. Dalton. Signaling network crosstalk in human pluripotent cells: a Smad2/3-regulated switch that controls the balance between self-renewal and differentiation. *Cell Stem Cell*, 10(3):312–326, Mar 2012.

[101] N. Solberg, O. Machon, O. Machonova, and S. Krauss. Mouse Tcf3 represses canonical Wnt signaling by either competing for -catenin binding or through occupation of DNA-binding sites. *Mol. Cell. Biochem.*, 365(1-2):53–63, Jun 2012.

[102] L. Solnica-Krezel and D. S. Sepich. Gastrulation: making and shaping germ layers. *Annu. Rev. Cell Dev. Biol.*, 28:687–717, 2012.

[103] P. B. Stathopulos, G. A. Scholz, Y. M. Hwang, J. A. Rumfeldt, J. R. Lepock, and E. M. Meiering. Sonication of proteins causes formation of aggregates that resemble amyloid. *Protein Sci.*, 13(11):3017–3027, Nov 2004.

[104] T. Sumi, N. Tsuneyoshi, N. Nakatsuji, and H. Suemori. Defining early lineage specification of human embryonic stem cells by the orchestrated balance of canonical Wnt/beta-catenin, Activin/Nodal and BMP signaling. *Development*, 135(17):2969–2979, Sep 2008.

[105] C. Tagwerker, K. Flick, M. Cui, C. Guerrero, Y. Dou, B. Auer, P. Baldi, L. Huang, and P. Kaiser. A tandem affinity tag for two-step purification under fully denaturing conditions: application in ubiquitin profiling and protein complex identification combined with in vivocross-linking. *Mol. Cell Proteomics*, 5(4):737–748, Apr 2006.

[106] P. P. Tam and D. A. Loebel. Gene function in mouse embryogenesis: get set for gastrulation. *Nat. Rev. Genet.*, 8(5):368–381, May 2007.

[107] W. L. Tam, C. Y. Lim, J. Han, J. Zhang, Y. S. Ang, H. H. Ng, H. Yang, and B. Lim. T-cell factor 3 regulates embryonic stem cell pluripotency and self-renewal by the transcriptional control of multiple lineage pathways. *Stem Cells*, 26(8):2019–2031, Aug 2008.

[108] D. ten Berge, D. Kurek, T. Blauwkamp, W. Koole, A. Maas, E. Eroglu, R. K. Siu, and R. Nusse. Embryonic stem cells require Wnt proteins to prevent differentiation to epiblast stem cells. *Nat. Cell Biol.*, 13(9):1070–1075, Sep 2011.

[109] A. K. Teo, Y. Ali, K. Y. Wong, H. Chipperfield, A. Sadasivam, Y. Poobalan, E. K. Tan, S. T. Wang, S. Abraham, N. Tsuneyoshi, L. W. Stanton, and N. R. Dunn. Activin and BMP4 synergistically promote formation of definitive endoderm in human embryonic stem cells. *Stem Cells*, 30(4):631–642, Apr 2012.

[110] P. J. Tesar, J. G. Chenoweth, F. A. Brook, T. J. Davies, E. P. Evans, D. L. Mack, R. L. Gardner, and R. D. McKay. New cell lines from mouse epiblast share defining features with human embryonic stem cells. *Nature*, 448(7150):196–199, Jul 2007.

[111] J. P. Thiery, H. Acloque, R. Y. Huang, and M. A. Nieto. Epithelial-mesenchymal transitions in development and disease. *Cell*, 139(5):871–890, Nov 2009.

[112] M. Thomas-Chollier, M. Defrance, A. Medina-Rivera, O. Sand, C. Herrmann, D. Thieffry, and J. van Helden. RSAT 2011: regulatory sequence analysis tools. *Nucleic Acids Res.*, 39(Web Server issue):86–91, Jul 2011.

[113] J. A. Thomson, J. Itskovitz-Eldor, S. S. Shapiro, M. A. Waknitz, J. J. Swiergiel, V. S. Marshall, and J. M. Jones. Embryonic stem cell lines derived from human blastocysts. *Science*, 282(5391):1145–1147, Nov 1998.

[114] G. G. Tortelote, J. M. Hernandez-Hernandez, A. J. Quaresma, J. A. Nickerson, A. N. Imbalzano, and J. A. Rivera-Perez. Wnt3 function in the epiblast is required for the maintenance but not the initiation of gastrulation in mice. *Dev. Biol.*, 374(1):164–173, Feb 2013.

[115] E. Trompouki, T. V. Bowman, L. N. Lawton, Z. P. Fan, D. C. Wu, A. DiBiase, C. S. Martin, J. N. Cech, A. K. Sessa, J. L. Leblanc, P. Li, E. M. Durand, C. Mosimann, G. C. Heffner, G. Q. Daley, R. F. Paulson, R. A. Young, and L. I. Zon. Lineage regulators direct BMP and Wnt pathways to cell-specific programs during differentiation and regeneration. *Cell*, 147(3):577–589, Oct 2011.

[116] L. Vallier, S. Mendjan, S. Brown, Z. Chng, A. Teo, L. E. Smithers, M. W. Trotter, C. H. Cho, A. Martinez, P. Rugg-Gunn, G. Brons, and R. A. Pedersen. Activin/Nodal signalling maintains pluripotency by controlling Nanog expression. *Development*, 136(8):1339–1349, Apr 2009.

[117] L. Vallier, T. Touboul, Z. Chng, M. Brimpari, N. Hannan, E. Millan, L. E. Smithers, M. Trotter, P. Rugg-Gunn, A. Weber, and R. A. Pedersen. Early cell fate decisions of human embryonic stem cells and mouse epiblast stem cells are controlled by the same signalling pathways. *PLoS ONE*, 4(6):e6082, 2009.

[118] R. van Amerongen and A. Berns. Knockout mouse models to study Wnt signal transduction. *Trends Genet.*, 22(12):678–689, Dec 2006.

[119] R. van Amerongen and R. Nusse. Towards an integrated view of Wnt signaling in development. *Development*, 136(19):3205–3214, Oct 2009.

[120] D. L. van den Berg, T. Snoek, N. P. Mullin, A. Yates, K. Bezstarosti, J. Demmers, I. Chambers, and R. A. Poot. An Oct4-centered protein interaction network in embryonic stem cells. *Cell Stem Cell*, 6(4):369–381, Apr 2010.

[121] D. Van Hoof, S. R. Braam, W. Dormeyer, D. Ward-van Oostwaard, A. J. Heck, J. Krijgsveld, and C. L. Mummery. Feeder-free monolayer cultures of human embryonic stem cells express an epithelial plasma membrane protein profile. *Stem Cells*, 26(11):2777–2781, Nov 2008.

[122] X. Varelas, R. Sakuma, P. Samavarchi-Tehrani, R. Peerani, B. M. Rao, J. Dembowy, M. B. Yaffe, P. W. Zandstra, and J. L. Wrana. TAZ controls Smad nucleocytoplasmic shuttling and regulates human embryonic stem-cell self-renewal. *Nat. Cell Biol.*, 10(7):837–848, Jul 2008.

[123] Q. T. Wang, K. Piotrowska, M. A. Ciemerych, L. Milenkovic, M. P. Scott, R. W. Davis, and M. Zernicka-Goetz. A genome-wide study of gene activity reveals developmental signaling pathways in the preimplantation mouse embryo. *Dev. Cell*, 6(1):133–144, Jan 2004.

[124] Z. Wang, E. Oron, B. Nelson, S. Razis, and N. Ivanova. Distinct lineage specification roles for NANOG, OCT4, and SOX2 in human embryonic stem cells. *Cell Stem Cell*, 10(4):440–454, Apr 2012.

[125] K. Watanabe, M. Ueno, D. Kamiya, A. Nishiyama, M. Matsumura, T. Wataya, J. B. Takahashi, S. Nishikawa, S. Nishikawa, K. Muguruma, and Y. Sasai. A ROCK inhibitor permits survival of dissociated human embryonic stem cells. *Nat. Biotechnol.*, 25(6):681–686, Jun 2007.

[126] G. Weitzer. Embryonic stem cell-derived embryoid bodies: an in vitro model of eutherian pregastrulation development and early gastrulation. *Handb Exp Pharmacol*, (174):21–51, 2006.

[127] G. Winnier, M. Blessing, P. A. Labosky, and B. L. Hogan. Bone morphogenetic protein-4 is required for mesoderm formation and patterning in the mouse. *Genes Dev.*, 9(17):2105–2116, Sep 1995.

[128] J. M. Wozney, V. Rosen, A. J. Celeste, L. M. Mitsock, M. J. Whitters, R. W. Kriz, R. M. Hewick, and E. A. Wang. Novel regulators of bone formation: molecular clones and activities. *Science*, 242(4885):1528–1534, Dec 1988.

[129] J. Wray, T. Kalkan, S. Gomez-Lopez, D. Eckardt, A. Cook, R. Kemler, and A. Smith. Inhibition of glycogen synthase kinase-3 alleviates Tcf3 repression of the pluripotency

network and increases embryonic stem cell resistance to differentiation. *Nat. Cell Biol.*, 13(7):838–845, Jul 2011.

[130] K. H. Wrighton, X. Lin, and X. H. Feng. Phospho-control of TGF-beta superfamily signaling. *Cell Res.*, 19(1):8–20, Jan 2009.

[131] C. I. Wu, J. A. Hoffman, B. R. Shy, E. M. Ford, E. Fuchs, H. Nguyen, and B. J. Merrill. Function of Wnt/-catenin in counteracting Tcf3 repression through the Tcf3--catenin interaction. *Development*, 139(12):2118–2129, Jun 2012.

[132] M. Y. Wu and C. S. Hill. Tgf-beta superfamily signaling in embryonic development and homeostasis. *Dev. Cell*, 16(3):329–343, Mar 2009.

[133] R. H. Xu, X. Chen, D. S. Li, R. Li, G. C. Addicks, C. Glennon, T. P. Zwaka, and J. A. Thomson. BMP4 initiates human embryonic stem cell differentiation to trophoblast. *Nat. Biotechnol.*, 20(12):1261–1264, Dec 2002.

[134] R. H. Xu, R. M. Peck, D. S. Li, X. Feng, T. Ludwig, and J. A. Thomson. Basic FGF and suppression of BMP signaling sustain undifferentiated proliferation of human ES cells. *Nat. Methods*, 2(3):185–190, Mar 2005.

[135] R. H. Xu, T. L. Sampsell-Barron, F. Gu, S. Root, R. M. Peck, G. Pan, J. Yu, J. Antosiewicz-Bourget, S. Tian, R. Stewart, and J. A. Thomson. NANOG is a direct target of TGFbeta/activin-mediated SMAD signaling in human ESCs. *Cell Stem Cell*, 3(2):196–206, Aug 2008.

[136] F. Yi, L. Pereira, J. A. Hoffman, B. R. Shy, C. M. Yuen, D. R. Liu, and B. J. Merrill. Opposing effects of Tcf3 and Tcf1 control Wnt stimulation of embryonic stem cell self-renewal. *Nat. Cell Biol.*, 13(7):762–770, Jul 2011.

[137] F. Yi, L. Pereira, and B. J. Merrill. Tcf3 functions as a steady-state limiter of transcriptional programs of mouse embryonic stem cell self-renewal. *Stem Cells*, 26(8):1951–1960, Aug 2008.

[138] X. Zhou, H. Sasaki, L. Lowe, B. L. Hogan, and M. R. Kuehn. Nodal is a novel TGF-beta-like gene expressed in the mouse node during gastrulation. *Nature*, 361(6412):543–547, Feb 1993.

[139] Z. Zhu and D. Huangfu. Human pluripotent stem cells: an emerging model in developmental biology. *Development*, 140(4):705–717, Feb 2013.

Appendix A

TCF7L1 siRNA Knockdown Microarray Gene List

Microarray analysis

Microarray analyses were performed using GeneChip Human Gene 1.0 ST Arrays (Affymetrix). Probe cell intensity files (*.CEL) were analyzed in Affymetrix Expression Console software v1.1.1 using the PLIER algorithm to generate probe level summarization files. Analysis of differential gene expression was performed using the CyberT algorithm (http://cybert.microarray.ics.uci.edu/). Genes were considered significantly affected if they had a p-value <0.05 and fold change cutoff of +/-1.5. Gene ontology analysis was performed using DAVID Bioinformatic Resources 6.7 (http://david.abcc.ncifcrf.gov).

3 Day TCF7L1 siRNA Knockdown Microarray Analysis Gene List

Gene	Fold Change	p-value
DACT1	3.1	0.00
NODAL	2.9	0.00
MIR124-1	2.8	0.03
FAT4	2.6	0.00
SMARCA2	2.4	0.00
IAPP	2.4	0.01
FST	2.4	0.00
LPAR6	2.3	0.00
THBS2	2.0	0.00
ABHD12B	2.0	0.02
MEIS2	1.9	0.00
RPRM	1.9	0.00
FLJ38379	1.9	0.00
CC2D2B	1.8	0.03
HHLA1	1.8	0.00
LMO3	1.8	0.00
OR4K15	1.7	0.05
DUSP10	1.7	0.01
SLCO4C1	1.7	0.00
TMPRSS11E	1.7	0.00
THRB	1.6	0.01
FLJ44838	1.6	0.04
NR5A2	1.6	0.00
SMYD2	1.6	0.00
SNORD45B	1.6	0.00
TNFAIP6	1.6	0.04
SULF1	1.6	0.01
LRRTM1	1.6	0.01
GCNT4	1.6	0.00
STOX2	1.6	0.00
GPR50	1.6	0.00
CDK6	1.5	0.00
BMP4	1.5	0.00
LRRC17	1.5	0.02
SALL3	1.5	0.00
SATB2	1.5	0.00
CSPG4P5	1.5	0.05
HEY2	1.5	0.00
IL1RAPL1	1.5	0.04
SLC39A8	1.5	0.00

ZNF257	1.5	0.04
RASGRF2	1.5	0.00
ANGPT1	1.5	0.04
RGMB	1.5	0.01
MCC	1.5	0.00
MGAT4C	1.5	0.05
PDE1B	1.5	0.01
PDE1C	1.5	0.05
PHLDB2	1.5	0.00
TMEM130	-1.5	0.01
CAMK2N1	-1.5	0.00
HLA-DRA	-1.5	0.03
PTTG3P	-1.5	0.04
AK5	-1.5	0.05
GABRP	-1.5	0.02
HLA-DRA	-1.5	0.03
C3orf57	-1.5	0.01
NTS	-1.5	0.01
ST8SIA5	-1.5	0.01
HLA-DRA	-1.5	0.02
TMEM59L	-1.5	0.00
TARP	-1.5	0.04
DGKB	-1.5	0.04
DGKK	-1.6	0.00
TTC23L	-1.6	0.03
EFHC2	-1.6	0.02
EFCAB4B	-1.6	0.02
PGA3	-1.6	0.00
PGA4	-1.6	0.00
PGA5	-1.6	0.00
CMKLR1	-1.6	0.00
TMEM14A	-1.6	0.00
AIF1	-1.6	0.00
JAKMIP2	-1.7	0.01
BNC1	-1.7	0.01
GRPR	-1.7	0.00
FAM75C1	-1.7	0.05
OR10K2	-1.7	0.04
ST18	-1.8	0.03
CEACAM1	-1.8	0.00
SPRR2B	-1.9	0.04
FAM65B	-2.0	0.00

LOC440944	-2.2	0.02
CLC	-2.2	0.00
EYS	-4.0	0.02
TCF7L1	-4.6	0.00
FRG2C	-5.3	0.00

Appendix B

Materials and Methods

hESC Culture

Human ESC lines H1, H9 and H14 were maintained on mitotically inactivated mouse embryonic fibroblasts (MEFs, Millipore). Human ESCs were grown in medium containing; DMEM/F12 (Invitrogen), 20% KnockOut serum (Invitrogen), 4 ng/ml recombinant human basic fibroblast growth factor (bFGF, Invitrogen), 5 mM GlutaMax (Invitrogen), 0.1 mM non-essential amino acids (Invitrogen), and 0.1 mM beta-mercaptoethanol. For general hESC propagation and maintenance, cells were mechanically passaged every 5-7 days using the StemPro EZPassage tool (Invitrogen). For feeder-free culture, hESCs were grown on Geltrex (Invitrogen) and maintained with mTeSR1 (StemCell Technologies) medium.

Western blot

Whole-cell protein extracts were prepared as follows: trypsinized (0.05% trypsin-EDTA) single cell suspensions were counted and 500,000 to 1 million cells (depending on the number cells required for a given experiment) were lysed in 95°C SDS Sample Buffer, boiled for 5 minutes at 95-100°C and then briefly sonicated to shear DNA. Ten percent acrylamide

gels were used for SDS-PAGE, then proteins were transferred to a PVDF membrane (GE Healthcare) for immunoblotting and membranes were imaged using a VersaDoc MP 5000 (Bio-Rad). All antibodies used in this study are listed in **Table B.1**.

Immunofluorescence staining

For immunofluorescence staining analysis, hESCs were grown on 35 mm glass-bottom imaging dishes (Matek). Cells were fixed in 4% paraformaldehyde in PBS(+) for 20 minutes at room temperature. After fixation, the cells were gently washed with PBS(+) and permeabilized with 0.5% Triton X-100 in PBS(+) for 10 minutes at room temperature. Cells were rinsed again and then blocked with a 10% serum (this was dependent upon the species of the secondary antibody) solution for 1 hour at room temperature. Primary antibodies were then diluted in 1% of the appropriate serum blocking buffer and incubated with the cells overnight at 4°C. The following day the cells were rinsed 3 times with PBS(+) and then incubated with the secondary antibodies for 2 hours, in the dark, at room temperature. Once this incubation was complete, the cells were rinsed with PBS(+), incubated with Hoechst (1:5000) for 7 minutes, rinsed again and then left in 1.5 - 2.0 ml of PBS(+). Stained cells were imaged on either a Zeiss LSM 510 confocal microscope or an Olympus FSX100. All antibodies used for IF are listed in **Table B.1**

Reverse transcription polymerase chain reaction (PCR)

Total RNA was isolated from hESCs using the RNeasy Mini Kit (Qiagen) according to manufacturers protocol. In addition, total RNA samples were treated with RNase-free DNase I (Qiagen) to remove genomic DNA. Total RNA was quantitated using a NanoDrop 2000 (Thermo Scientific) and 0.25 μg of total RNA was used for PCR analysis. PCR conditions followed the SuperScript One-Step (Invitrogen) protocol with an annealing temperature of 55°C for 35 cycles of amplification. β-ACTIN, the loading control, was amplified over 25 cycles with an annealing temperature 55°C. All PCR primer sets used in this study are listed

in **Table B.2**.

Quantitative PCR (qPCR)

Total RNA was isolated from hESCs and genomic DNA removed as described above. Total RNA was quantitated using a NanoDrop 2000 (Thermo Scientific) and 1.0 μg of total RNA was used for cDNA synthesis (High Capacity cDNA Reverse Transcriptase Kit; Invitrogen). Taqman and/or SYBR Green assays (Invitrogen) were used to measure RNA expression levels using a ViiA 7 RT-PCR System (Invitrogen). Data was analyzed by the $\Delta\Delta$CT method and normalized to either 18S or GAPDH [62]. GraphPad Prism version 5 and 6 were used to create graphs and perform two-tailed student's t-test statistical analyses. SYBR Green primer sets and Taqman probes are listed in **Table B.3** and **Table B.5**.

siRNA knockdown

siRNA mediated knockdowns in hESCs were performed under feeder-free conditions on Geltrex (Invitrogen) in mTeSR1 (STEMCELL Technologies) defined medium. Human ESCs were dissociated into single cells using 0.05% trypsin-EDTA (Invitrogen), filtered through a 40 μm filter to remove clumps and then 180,000 - 200,000 cells were plated per well of a 6-well plate in the presence of 10 μM ROCK inhibitor (Calbiochem) to increase single cell survival [125]. After 18-20 hours, siRNAs were transfected into the hESCs using the RNAiMAX (Invitrogen) transfection reagent according to the manufacturers protocol. Briefly, transfection of a single well of a 6-well plate with an siRNA at a concentration of 50 nM was performed as follows: an eppendorf tube containing 6 μl of RNAiMAX + 94 μl of Opti-MEM (Invitrogen) (total volume 100 μl) and a second tube containing 5 μl siRNA + 95 μl Opti-MEM (total volume 100 μl) are mixed together, incubated at room temperature for 15 minutes and then added to the well of hESCs. Twenty-four hours later this transfection procedure was repeated for a second time. *TCF7L1* and the non-targeting siRNAs were purchased from Dharmacon and used at a concentration of 50 nM. RNA was isolated at

appropriate time points and analyzed using qPCR as previously described. All siRNAs used in this study are listed in **Table B.6**.

Primitive streak-like differentiation assay

Human ESCs were trypsinized and 200,000 single cells were plated per well of a six-well plate as described above under feeder-free conditions. The following day, the cells were given mTeSR1 medium to allow them to recuperate from the single cell passaging and give them time to form small colonies. Twenty-four hours later, the cells were given primitive streak-like (PS) differentiation medium which contains: RPMI 1640 (Invitrogen), 2% B-27 supplement (Invitrogen), 1X non-essential amino acids (Invitrogen), 50 μg BMP4 (R&D Systems) and 50 μg ACTIVIN A (R&D Systems). After 24 hours in PS differentiation medium, hESCs were given fresh PS differentiation medium for another 24 hours. After 48 hours of PS differentiation, the cells were harvested for experiment specific analyses.

Chromatin immunoprecipitation (ChIP)

Each ChIP experiment was performed using approximately 60 million cells grown under feeder-free conditions in mTeSR1 defined medium as described above. Twenty million cells were used for each IP condition.

Once the hESCs reached confluency they were cross-linked in **Fixing Buffer** (PBS + 1% methanol-free formaldehyde) for 10 minutes at room temperature. Cross-linking was inhibited by adding 2.5 M glycine to a final concentration of 0.125 M glycine and incubated for 10 minutes at room temperature with gentle rocking. Afterwards the wells were aspirated and rinsed with cold (4°C) PBS + protease inhibitors (PIs) twice. Next, the cells were scraped from the wells of the tissue culture plate with a cell scraper (Corning) and added to a 50 ml conical tube on ice. The hESC suspension was centrifuged at 1,350 x g for 10 minutes at 4°C. The supernatant was carefully removed and the cell pellet was resuspended with 10 ml cold **Lysis Buffer** (50 mM HEPES pH 7.9; 140 mM NaCl; 1 mM EDTA;

10% Glycerol; 0.5% NP-40; 0.25% Triton X-100; fresh PIs) to release the nuclei from the cells.The cell suspension was then transferred to a new 15 ml conical tube and incubated at 4°C on a nutator for 10 minutes. Cells were pelleted at 1,350 x g for 5 minutes at 4°C. The supernatant was carefully removed and the pellet was resuspended with 5 ml **Wash Buffer** (10 mM Tris-HCl pH 8.0; 200 mM NaCl; 1 mM EDTA pH 8.0; 0.5 mM EGTA pH 8.0; fresh PIs) and incubated again at 4°C for 10 minutes on a nutator. Finally, the cell nuclei were pelleted by centrifugation at 1,350 x g for 5 minutes at 4°C.

The nuclear pellet was resuspended in 1 ml **Buffer NUC** (15 mM HEPES pH 7.5; 60 mM KCl; 15 mM NaCl; 0.32 mM Sucrose; fresh 0.5 mM PMSF) and transferred to a new microfuge tube bringing the total volume to ∼1.2 ml. Next, 12 μl of 100X protease inhibitor (Thermo Scientific), 3.3 μl of 1M CaCl2 and 10 μl micrococcal nuclease (2,000 units/μl) (New England Biolabs) were added to the chromatin suspension and incubated at 37°C for exactly 25 minutes with shaking. Immediately after 25 minutes, 1.2 ml of **2X Sonication Buffer** (90 mM HEPES pH 7.9; 220 mM NaCl; 10 mM EDTA pH 8.0; 1% NP-40; 0.2% Na-deoxycholate; 0.2% SDS) was added to the sample in a 15 ml conical tube and placed on ice. The volume was then equally split between two new 1.5 ml microfuge tubes for gentle sonication (to release chromatin from the nuclei). The QSonica Q700 with micro-tip settings were: Amplitude 1; Time ON 5 seconds; Time OFF 60 seconds. This cycle was four times for each microfuge tube. Sheared and sonicated samples were centrifuged for 15 minutes at 18,000 rpm at 4°C. Chromatin from both microfuge tubes was pooled together in a new 15 ml conical tube and the volume was brought to exactly 3 ml by adding **1X Sonication Buffer** (1 volume **Buffer NUC** to 1 volume **2X Sonication Buffer** with protease inhibitors) as needed. Lastly, 30 μl was removed for **1% Input** and DNA digestion analysis.

For a single ChIP, 5 μg of antibody (or IgG for the control) was added to 1 ml of the chromatin suspension (1 ml is roughly equal to the chromatin from 20 million cells) and incubated overnight at 4°C on a nutator. The next day, 50 μl of Dynabeads (Invitrogen) were

washed once with 1 ml of **1X Sonication Buffer** and then transferred to the chromatin and antibody suspension. Beads were incubated with chromatin for 2 hours at 4°C on a rotator. Afterwards, beads were collected using a magnet and then washed as follows: 5 washes with 1 ml cold **RIPA Buffer** (50 mM HEPES-KOH pH 7.9; 500 mM LiCl; 1 mM EDTA; 1% NP-40; 0.7% Na-deoxycholate) and twice with 1 ml **TE Buffer** (10 mM Tris-HCl pH 7.5 + 1 mM EDTA). After the final wash, the beads were centrifuged at room temperature for 1 minute at 1,000 rpm and the supernatant was discarded. The beads were resuspended in 210 μl **Direct Elution Buffer** (10 mM Tris-HCl pH 8.0; 300 mM NaCl; 5 mM EDTA pH 8.0; 1% SDS; 0.2 μg/ml RNase A) and incubated at 65°C overnight with shaking. The following day, the beads were centrifuged for 1 minute at 18,000 rpm and 200 μl of the supernatant was transferred to a new microfuge tube. 2 μl of Proteinase K (20 mg/ml stock) (Thermo Scientific) was added to the sample, incubated at 55°C for 2 hours and then purified using a Mini Elute PCR Purification Kit (Qiagen) according to the manufacturers protocol. DNA was eluted with 30 μl warmed (50°C) **EB Buffer** and the DNA concentration was measured with Qubit HS Kit (Invitrogen) following the manufacturers instructions.

ChIP-seq Peak calling and Irreproducibility Discovery Rate (IDR)

TCF7L1 ChIP-seq and Input samples from H9 hESCs were sequenced at the UCI High-Throughput Facility on the Illumina HiSeq 2500 platform. Reads were mapped to the human genome (hg19) with Bowtie version 0.12.8 using -v 2, -S, -k 1, -m 1 parameters to report only uniquely mapped reads with at most 2 mismatches. The resulting sam files were converted to bam files using Samtools version 0.1.19. To ensure greater than or equal read depth in Input samples versus ChIP samples (as suggested by ENCODE), TCF7L1 ChIP bam files were random down sampled (using Samtools -view, -b, -s parameters) so that they contained fewer reads than Input, as suggested by ENCODE. Samtools was also used to remove PCR duplicates (using the rmdup parameter). The resulting number of mapped reads for each sample was:

Input: 55,127,188

TCF7L1 Replicate #1: 53,797,735

TCF7L1 Replicate #2: 54,273,005

IDR analysis was performed as suggested by ENCODE to determine the consistency of the ChIP-seq biological replicates. IDR was performed using the Macs2 peak caller as in:

https://sites.google.com/site/anshulkundaje/projects/idr

Briefly, peaks were called on individual TCF7L1 biological replicates as well as on the pooled replicates with Macs2 using a p-value threshold of 1E-2. The top 50,000 peaks from TCF7L1 ChIP-seq replicate #1 and replicate #2 were entered into IDR analysis. To get there final like os ChIP-seq peaks an IDR of 0.05 was used, which is within the range suggested by:

https://sites.google.com/site/anshulkundaje/projects/idr

Genomic Regions Enrichment of Annotations Tool (GREAT) analysis

GREAT analysis was performed on the top 1,000 TCF7L1 peaks +/- 5 kb from annotated transcription start sites. The GREAT analysis application can be accessed here:

http://bejerano.stanford.edu/great/public/html/index.php

The Human: GRCn37 hg19 build was selected as the "Species Assembly" and the whole genome was chosen for the "Background regions" field. The "Basal plus extension" analysis method was used along with its default settings: Proximal 5.0 kb upstream, 1.0 kb downstream, plus distal up to 1000.0 kb.

Cloning and Site-Directed Mutagenesis

Cloning was performed with the GeneArt Seamless Cloning Kit (Invitrogen) in accordance with the manufacturers instructions. This kit was also used for site-directed mutagenesis

applications. PCR amplification of cDNA fragments was performed using Phusion High-Fidelity DNA Polymerase (Thermo Scientific). All constructs generated and used in this study are listed in **Table A.3**.

Table B.1: **Antibodies**

Antibody	**Company**	**Cat. #**	**[WB]**	**[IF]**
TCF7L1	Cell Signaling	2883	1:1000	1:200
TCF7L1	Santa Cruz	sc-8635	1:200	1:50
OCT4	R&D	AF1759	1:1000	1:100
β-catenin	Cell Signaling	2677		1:200
β-catenin	Cell Signaling	8814	1:1000	
FLAG (M2)	Sigma	F3165	1:2500	1:375
SMARCA2	Abcam	ab15597	1:1000	
CASP3 (Cleaved)	Cell Signaling	9661	1:1000	
β-ACTIN	Santa Cruz	sc-1616	1:1000	
anti-goat HRP	Santa Cruz	sc-2350	1:5000	
anti-mouse HRP	Promega	G770B	1:5000	
Donk. anti-rabbit 488	Invitrogen	A21206		1:400
Donk. anti-goat 568	Invitrogen	A11057		1:400

Table B.2: **PCR primer sets**

Gene	**Primer Sequence (5' - 3')**
Fwd. *TCF7L1*	TCTCCCTCACCACCAAACCAGAAA
Rev. *TCF7L1*	TGTCTTCTCACATGGTGATGGCCT
Fwd. *TCF7L2*	ACTGTCCAGAGAAGAGCAAGCGAA
Rev. *TCF7L2*	TCGGAGGAAGTGAAAGGCAAGGAT
Fwd. *TCF7*	CCATCAAGAAGCCCCTCAATGC
Rev. *TCF7*	GCTCAGTTCAGCCCATCTCTGACC
Fwd. *LEF1*	TATGATTCCCGGTCCTCCTGGTC
Rev. *LEF1*	TGGCTCCTGCTCCTTTCTCTGTTC
Fwd. *β-ACTIN*	TGACGGGGTCACCCACACTGTGCCCATCTA
Rev. *β-ACTIN*	CTAGAAGCATTTGCGGTGGACGATGGAGGG

Table B.3: **SYBR Green qPCR primer sets**

Gene	**Primer Sequence (5' - 3')**
Fwd. *GAPDH*	GTGGACCTGACCTGCCGTCT
Rev. *GAPDH*	GGAGGAGTGGGTGTCGCTGT
Fwd. *NODAL*	TACAGGCAGAAGATGTGGCAGTGGAT
Rev. *NODAL*	ATCCTCTTGTTGGCTCAGGAAGGA
Fwd. *DACT1*	TGGAGGAGAAGTTCTTGGAGGAGA
Rev. *DACT1*	TCCAGTCTCAGGTCACTTATCTGC
Fwd. *BMP4*	ATGATTCCTGGTAACCGAATGC
Rev. *BMP4*	CCCCGTCTCAGGTATCAAACT
Fwd. *WNT3*	AGGGCACCTCCACCATTTG
Rev. *WNT3*	GACACTAACACGCCGAAGTCA

Table B.4: **SYBR Green ChIP-qPCR primer sets**

Gene	**Primer Sequence (5' - 3')**
Fwd. *NODAL*	CCCAGTGATTTCAGGAGGAAAG
Rev. *NODAL*	GCCCAGAGATCAAAGTGAGTG
Fwd. *WNT3*	GTGAATGTGTGGGACCTTAGAC
Rev. *WNT3*	GTGGAGCCTCACTGAATACAC
Fwd. *TDGF1*	ACGTCCGCCTGGAATTTG
Rev. *TDGF1*	GACTAGGAAGCTTGAAACTGAGATAG
Fwd. *BMP4*	GCCTGTGACCAGCTTCTT
Rev. *BMP4*	CCTATGGTGAGCAAGGCTAC
Fwd. *EOMES*	GGAGTCAGTCAGAACCAAAGAG
Rev. *EOMES*	CGGGAGGGCACTTGATTT
Fwd. *GSC*	GTTGTCGATGCTGAACATGC
Rev. *GSC*	CGCTCTCTTTCGGTTTGGT
Fwd. *OTX2*	TGCAAAGTCGGCCCAAAT
Rev. *OTX2*	CCTTAGTTCCACTGCTCCAAAC
Fwd. *MIXL1*	CAGGCTGTAAAGCTGCAAATC
Rev. *MIXL1*	AGCAACTGTCTGGTTCACTATC
Fwd. *FOXD3*	GTGCGCTGCTCTTACTCTTTA
Rev. *FOXD3*	CGAGGTTCCCATATCGTGTTT
Fwd. *BMPR2*	GCACTACACAAATCCTTGGAAAC
Rev. *BMPR2*	GAGTTAGAGTTGTGTCGGGATAG
Fwd. *TCF7L1*	TCATTTAAAGCGAGCGCTGCGACA
Rev. *TCF7L1*	AAGGAATCCGCCCTCATTTGCAC
Fwd. Neg. Control	CCGAATTTGGGCCTCTACAA
Rev. Neg. Control	CATGGTGGCTACGGTGAATAA

Table B.5: **Taqman qPCR probe sets**

Probe	**Company**	**Cat. #**
TCF7L1	Invitrogen	Hs01064111_m1
TCF7L2	Invitrogen	Hs01009038_m1
TCF7	Invitrogen	Hs01556515_m1
LEF1	Invitrogen	Hs01547250_m1
OCT4	Invitrogen	Hs00999632_g1
SOX2	Invitrogen	Hs01053049_s1
NANOG	Invitrogen	Hs02387400_g1
T	Invitrogen	Hs00610080_m1
MIXL1	Invitrogen	Hs00430824_g1
GSC	Invitrogen	Hs00418279_m1
SMARCA2	Invitrogen	Hs00542638_m1
SNAI1	Invitrogen	Hs00195591_m1
MEIS2	Invitrogen	Hs00230534_m1
FST	Invitrogen	Hs00246256_m1
FOXA2	Invitrogen	Hs00232764_m1
WNT8A	Invitrogen	Hs00230534_m1
WNT5A	Invitrogen	Hs00998537_m1
GAPDH	Invitrogen	Hs99999905_m1
18S	Invitrogen	Hs99999901_s1

Table B.6: **siRNAs**

siRNA	**Company**	**Cat. #**	**[nM]**
TCF7L1 SMARTpool	Dharmacon	L-014703-00-0005	50
TCF7L1 #5	Dharmacon	J-014703-05	50
TCF7L1 #6	Dharmacon	J-014703-06	50
Non-targeting pool	Dharmacon	D-001810-10-05	50
Non-targeting #2	Dharmacon	D-001810-02-05	50
NANOG SMARTpool	Dharmacon	L-014489-00-0005	100
SOX2 SMARTpool	Dharmacon	L-011778-00-0005	100
OCT4	Qiagen	AGCAGCTTGGGCTCGAGAA	100

TCF7L1 Construct List

TCF7L1 Constructs	Mutation/Notes	TAG
pCMV-SPORT6-TCF7L1	Contains 5' and 3' UTRs	
pcDNA6.2-EmGFP-TCF7L1	No C-terminal V5 tag	EmGFP
pcDNA6.2-EmGFP-TCF7L1-V5	Contains N-terminal V5 tag	EmGFP
pGEMT-TCF7L1 (+Stop Codon)	Contains Stop Codon	
pGEMT-TCF7L1 (-Stop Codon)	Does not contain a stop codon	
pGEMT-HTBH-TCF7L1	N-terminal HTBH tag	
p3xFLAG-TCF7L1	N-terminal 3xFLAG tag	HTBH
p3xFLAG-EcoRV-TCF7L1-SwaI	Its full length TCF7L1 w/o stop codon, but SwaI contains a stop codon	3xFLAG
pEF1alpha-TCF7L1-ZsGreen1	Full length TCF7L1 with no tags	3xFLAG
pEF1alpha-HTBH-TCF7L1-IRES-ZsGreen1	N-terminal HTBH tag	
pCDH-HTBH-TCF7L1-IRES-Neo	N-terminal HTBH tag	HTBH
pEF1alpha-3xFLAG-EcoRV-TCF7L1-ZraI-HTBH-IRES2-ZsGreen1	C-terminal HTBH Tag	HTBH
pEF1alpha-3xFLAG-EcoRV-TCF7L1(K315R)-ZraI-HTBH-IRES2-ZsGreen1	Putative SUMO site at lysine 315 mutation	HTBH
		HTBH
TCF7L1 Mutant Constructs		
pcDNA6.2-EmGFP-TCF7L1(D80N)-V5	Putative caspase cleavage mutant	EmGFP
pcDNA6.2-EmGFP-TCF7L1(D465N/D476N)-V5	Double putative caspase cleavage mutant 1 & 2	EmGFP
pcDNA6.2-EmGFP-TCF7L1(D80N/D476N)-V5	Double putative caspase cleavage mutant	EmGFP
pcDNA6.2-EmGFP-TCF7L1(D465N)-V5 Casp1mut	D465N	EmGFP
pcDNA6.2-EmGFP-TCF7L1(D476N)-V5 Casp2mut	D476N	EmGFP
pcDNA6.2-EmGFP-TCF7L1(D416N)-V5 Casp3mut	D416N	EmGFP
pcDNA6.2-EmGFP-HMGmut_TCF7L1-V5	K346E & K347E	EmGFP
pcDNA6.2-EmGFP-TCF7L1(ΔBcat)-V5	Deleted first 216 aa (not including start codon met)	EmGFP
p3xFLAG-TCF7L1(D80N/D476N)	Double putative caspase cleavage mutant	3xFLAG
p3xFLAG-TCF7L1(D465N) Casp1mut	D465N	3xFLAG
p3xFLAG-TCF7L1(D476N) Casp2mut	D476N	3xFLAG
p3xFLAG-TCF7L1_HMGmut	K346E & K347E	3xFLAG
p3xFLAG-dnTCF7L1	HMG-NLS portion of TCF7L1	3xFLAG
p3xFLAG-TCF7L1(D36A)	D36A mutation in TCF7L1 B-catenin binding domain to reduce B-catenin binding	3xFLAG
p3xFLAG-dnTCF7L1-EmGFP	dominant negative TCF7L1 fused to EmGFP	3xFLAG
p3xFLAG-EcoRV-TCF7L1-(S140A)-SwaI	Serine 140 phosphorylation site is mutated to alanine	3xFLAG
p3xFLAG-EcoRV-TCF7L1-(T411A)-SwaI	Threonine 411 phosphorylation site is mutated to alanine	3xFLAG
Transgenic Mouse Constructs		
pT-TetOn-FLAG-TCF7L1	T(promoter)-rtTA-P2A-H2B:GFP-insulator-TRE-FLAG-TCF7L1	3xFLAG
pROSA26-loxP:STOP:loxP-FLAG-TCF7L1	ROSA26 targeting construct. CAGGS promoter driving FLAG-TCF7L1-P2A-H2B:DsRed	3xFLAG
Lentiviral Vectors		
pLenti-CMVtight-3xFLAG-dnTCF7L1-EmGFP-Puro	Tet inducible dominant negative TCF7L1 fused to EmGFP	3xFLAG
pLenti-CMVtight-3xFLAG-TCF7L1-HTBH-Puro	Tet inducible TCF7L1 fused to HTBH	3xFLAG
pLKO-Tet-On-WPRE	Contains the WPRE element	
pLenti-EF1alpha-rtTA3-Blasticidin	EF1alpha promoter driving rtTA3	
ERT2 Fusions		
pEF1alpha-3xFLAG-dnTCF7L1-ERT2-IRES2-ZsGreen1	dominant negative TCF7L1 fused to estrogen receptor w/ ZsGreen1	3xFLAG
pEF1alpha-3xFLAG-TCF7L1-ERT2-IRES2-ZsGreen1	TCF7L1 fused to estrogen receptor w/ ZsGreen1	3xFLAG
pCDH-EF1a-3xFLAG-dnTCF7L1-ERT2-IRES-Neo	dominant negative TCF7L1 fused to estrogen receptor	3xFLAG
ERtm Fusions		
p3xFLAG-EcoRV-TCF7L1-EcoRV-Ertm	TCF7L1 fused to C-terminal ERtm domain & flanked by EcoRV sites for easy removal and insertion of cDNA of choice	3xFLAG
p3xFLAG-dnTCF7L1-Ertm	dominant negative TCF7L1 w/ C-terminal ERtm fusion	3xFLAG
pEF1alpha-3xFLAG-TCF7L1(1-476)-Ertm-IRES2-ZsGreen1	TCF7L1 clipped isoform w/ C-terminal ERtm fusion & ZsGreen1	3xFLAG
pEF1alpha-3xFLAG-dnTCF7L1-Ertm-IRES2-ZsGreen1	C-terminal ERtm fusion w/ ZsGreen1	3xFLAG
pEF1alpha-3xFLAG-TCF7L1(D476N)-Ertm-IRES2-ZsGreen1	ER inducible caspase cleavage mutant w/ ZsGreen1	3xFLAG
pCDH-EF1a-3xFLAG-dnTCF7L1-Ertm-IRES-Neo	C-terminal ERtm fusion w/ neomycin	3xFLAG
pCDH-EF1a-3xFLAG-TCF7L1(1-476)-Ertm-IRES-Neo	C-temrinal deletion mutant of TCF-3 w/ neomycin	3xFLAG
pCDH-EF1a-3xFLAG-EcoRV-TCF7L1-EcoRV-ERtm-IRES-Neo	C-terminal ERtm fusion w/ neomycin	3xFLAG
pEF1alpha-3xFLAG-EcoRV-TCF7L1-ZraI-Ertm-IRES2-ZsGreen1	C-terminal ERtm fusion w/ neomycin	3xFLAG
pEF1alpha-3xFLAG-EcoRV-TCF7L1-ZraI-HTBH-Ertm-IRES2-ZsGreen1	Frankenstein!!!!!	3xFLAG, HTBH
Vectors without TCF-3		
pEF1alpha-3xFLAG-EcoRV-ERT2-SwaI-IRES2-ZsGreen1	Contains new ERT2 fusion domain w/ ZsGreen1	3xFLAG
pEF1alpha-3xFLAG-EcoRV-Ertm-IRES2-ZsGreen1	Contains C-terminal Ertm fusion domain w/ZsGreen1	3xFLAG
pCDH-EF1a-EmGFP-IRES-Neo	EmGFP for visualizing transfected hESCs	
pCDH-EF1a-3xFLAG-EcoRV-Ertm-IRES-Neo	Control empty vector	3xFLAG
pLenti-EF1alpha-rtTA3-Blasticidin	Cut out CMV and cloned in the EF1alhpa promoter. This is a lenti Tet-On vector from addgene	
Ac5-EF1alpha-TetR-dT2A-GFP-T2A-Neo	EF1alpha promoter. TetR contains the SV40 NLS and everything is separated by T2A	
DSCO shRNA Lenti Constructs		
pDSCO	EF1alpha promoter driving TetR-SV40_PolyA-dT2A-EGFP-T2A-Neomycin. Construct contains responsive H1/TO promotre for shRNA	

"It is not birth, marriage, or death,
but gastrulation which is truly the most important time in your life."
Lewis Wolpert, Ph.D., Developmental Biologist.