LHX1 IS A MASTER REGULATOR OF INTERMEDIATE MESODERM FORMATION FROM HUMAN EMBRYONIC STEM CELLS

by

BENJAMIN R BOWARD

(Under the Direction of Stephen Dalton)

ABSTRACT

The adult kidney is derived from the intermediate mesoderm, an early mesoderm subtype located between the paraxial and lateral plate mesoderm. Genetic regulation and fate determination of intermediate mesoderm are poorly understood. The homeobox transcription factor LHX1 is expressed in the developing intermediate mesoderm and is required for the specification of several kidney lineages, but its function in determining IM cell fate is unknown. Throughout this research, I show that LHX1 is necessary and sufficient for differentiation of human ESCs to intermediate mesoderm. Bioinformatics analysis has unveiled an LHX1 interaction network that positions LHX1 as a master regulator of IM formation. LHX1 regulates IM gene expression through binding to distal elements and its binding is required for the function of IM enhancers. This research is the first to describe a role for a transcription factor in IM lineage specification. Further understanding of the intermediate mesoderm transcriptional fate map should enable higher efficiency kidney differentiations for use in developmental and disease models. This work broadens the understanding of

early human embryonic development and provides insight into early mesoderm cell fate decisions.

INDEX WORDS: Stem cells, differentiation, intermediate mesoderm, LHX1, kidney, sequencing, development

LHX1 IS A MASTER REGULATOR OF INTERMEDIATE MESODERM FORMATION FROM HUMAN EMBRYONIC STEM CELLS

by

BENJAMIN R BOWARD

BS, JAMES MADISON UNIVERSITY, 2013

A Dissertation Submitted to the Graduate Faculty of The University of Georgia in Partial Fulfillment of the Requirements for the Degree

DOCTOR OF PHILOSOPHY

ATHENS, GEORGIA

2019

LHX1 IS A MASTER REGULATOR OF INTERMEDIATE MESODERM FORMATION FROM HUMAN EMBRYONIC STEM CELLS

by

BENJAMIN R BOWARD

Major Professor: Stephen Dalton
Committee: Jonathan Eggenshwiler
Hang Yin
Shaying Zhao

Electronic Version Approved:

Suzanne Barbour
Dean of the Graduate School
The University of Georgia
May 2019

DEDICATION

This work is dedicated to my parents, Ken and Laura Boward. All of my academic achievements are the result of your continued support and encouragement throughout my life. You have always encouraged me to be the best man that I can be in my professional and personal life. You will always be role models for me, and I still have so much to learn from you both. Without your endless support, none of this would have been possible. Thanks Mom and Dad.

ACKNOWLEDGEMENTS

I would first like to thank my family for their continued support in my graduate career. Even though I have been hundreds of miles away, I have felt that my family has been there for me every step of the way. Thank you Mom, Dad, Emily and Sarah, Nana, and Granddad.

I would like to thank Dr. Stephen Dalton for taking me in as a graduate student in his lab and providing me with guidance during my development as a scientist. He has provided me with a deep understanding of how science is done, and an appreciation of the rigor and attention to detail that high-level science requires.

I feel very fortunate to have worked with the members of the Dalton Lab over the years. One of the things that I will miss most about Athens is the time spent at lunch with my colleagues discussing everything from stem cell biology to current events in sports. They have made going to work every day a lot of fun, and will be life-long friends.

I would also like to thank the Innovation Gateway at UGA, and Georgia Bio. The opportunities that I have had through these organizations have provided me with considerable knowledge beyond the laboratory and have given me a tremendous advantage in my career search, ultimately leading to successfully landing a job at a major biotechnology company in Boston.

Finally, I would like to thank my friends. My time spent in Athens has been something that I will always treasure. The life-long relationships that I have formed in Athens, as well as the ones that I have maintained from my life before graduate school have made my time in Athens truly unforgettable. Thank you for all of the great times, and your continued love and support.

TABLE OF CONTENTS

Page

LIST OF TABLES

Page

LIST OF FIGURES

Page

CHAPTER 1

INTRODUCTION AND LITERATURE REVIEW

Pluripotent stem cells

Pluripotent stem cells (PSCs) are defined by their ability to proliferate indefinitely *in vitro*, or to self-renew, and to give rise to every adult cell lineage in the body, a term called pluripotency(Thomson et al., 1998; Zwaka and Thomson, 2005). Self-renewal is the ability to divide symmetrically into two identical daughter cells indefinitely. These properties make pluripotent stem cells an ideal system to study things like human development and disease, thus providing a system for investigating the mechanisms of early embryonic development(Harrison et al., 2017). PSCs also have enormous potential to replace damaged or diseased tissue, a development known as regenerative medicine. Specifically, this provides a platform for the investigation of signaling pathways, genetic regulation, and protein regulation among others to enhance the knowledge of early human development and provide targets for regenerative medicine. PSC lines are primarily generated from embryonic stem cell isolation or reprogramming, a process of de-differentiation.

Embryonic Stem Cells

Embryonic stem cells (ESCs), first derived from murine origin in 1981(Evans and Kaufman, 1981; Martin, 1981) and human origin in 1998(Thomson et al., 1998), are cells derived from the inner cell mass (ICM) of pre-implantation blastocyst-stage embryos. This resulted in immortalized ESCs, capable of self-renewal, that were a tremendous improvement over existing *in vitro* models for studying early embryonic development that had limited differentiation capacity, such as embryonic carcinoma cells(Martin, 1981). Pluripotency is maintained in murine (mPSCs) and human (hPSCs) by a set of core pluripotency transcription factors, OCT4, SOX2 and NANOG(Boyer et al., 2005) as well as other implicated factors such as MYC and KLF4(Jiang et al., 2008; Kim et al., 2012; Sumi et al., 2007). Furthermore, in order to maintain pluripotency and self-renew indefinitely *in vitro*, hPSCs must be cultured in the presence of defined extrinsic factors such as Activin A and FGF2(D. James et al., 2005; Vallier et al., 2005). mESCs can be cultured to maintain pluripotency with self-renewal capabilities on murine embryonic fibroblasts in the presence of fetal calf serum(Brook and Gardner, 1997), or in defined culture conditions containing leukemia inhibitory factor and inhibitors of signaling factors ERK and GSK3β(Silva et al., 2008; Ying et al., 2008). With the exception of certain difficult experimental conditions, the culture of PSCs on a layer of feeder cells is considered obsolete with the substitution of feeder cells with gelatinized matrices, and the discovery of defined medium conditions that maintain pluripotency.

Induced Pluripotent Stem Cells

In 2006, Shinya Yamanaka and colleagues discovered that cells could be reversed back to the pluripotent state through a process called reprogramming(K. Takahashi and Yamanaka, 2006). This was accomplished by viral introduction of core pluripotency transcription factors OCT4, SOX2, KLF4, and MYC (OSKM) in mouse embryonic fibroblasts. Introduction of these factors into somatic cells is sufficient to induce a genetic reprogramming back to the pluripotent state. These induced pluripotent stem cells (iPSCs) have remarkably similar characteristics as PSCs derived from embryonic origin and maintain pluripotent differentiation potential(Meissner, 2010; Polo et al., 2012). In 2007 Takahashi showed that adult human skin fibroblasts could be reprogrammed to the pluripotent state using the same factors, opening the door for using iPSCs for things like disease modeling and patient specific stem cell therapies(K. Takahashi et al., 2007). Since then, significant resources have been devoted to using iPSCs for disease modeling and will continue to provide unique and valuable insight into the mechanisms of disease and provide novel treatment avenues.

Mesoderm Formation in the Embryo

The body plan of vertebrate embryos arises as a result of a process called gastrulation by which much of the patterning is coordinated by signals coming from a structure in the midline, the primitive streak. This results in the spatial organization of tissues including the somitic mesoderm, splanchnic mesoderm, and intermediate mesoderm, which gives rise to tissues such as skeletal muscle,

heart and kidneys. This arrangement of tissues along the body axis is highly conserved among vertebrate species(Wilson et al., 2009). The primitive streak arises at embryonic day (E) 6.5 where pluripotent epiblast cells undergo an epithelial to mesenchymal transition resulting in the formation of the three germ layers, the ectoderm, endoderm and early mesoderm lineages(Tam and Behringer, 1997). Different mesoderm cell types are formed based on the time and the site of migration through the primitive streak. The earliest, posterior mesoderm population is patterned in response to BMP4 signaling from the extraembryonic ectoderm and give rise to the extraembryonic tissues(Arnold and Robertson, 2009). Lateral plate mesoderm, as well as paraxial mesoderm and cardiac mesoderm follow, arising from the intermediate and anterior regions of the primitive streak. Lastly, epiblast cells that migrate to the anterior give rise to tissues such as the notochord and the definitive endoderm (DE) lineage.

These processes are controlled by opposing gradients of Nodal and BMP signaling(Arnold and Robertson, 2009). Experiments in amphibian embryos showed these interactions between tissues and provided the groundwork for further identification of secreted and intracellular factors that are involved in germ layer formation. The first identified factor, basic fibroblast growth factor (bFGF) was shown to induce mesoderm in amphibians and has been implicated in inducing mesoderm tissue in many different organisms(Alev et al., 2013; Burdsal et al., 1998; Cao et al., 2004). TGFβs were also found to induce mesoderm, for example Activin A(Kimelman and Kirschner, 1987), Nodals(Shen, 2007) and BMPs(Köster et al., 1991). Nodal signals are essential for mesoderm induction

and have also been shown to be mesoderm inducers in gain of function experiments(Jones et al., 1995). BMP signaling has also been implicated in mesoderm induction. BMPs signal through a pathway that involves Smad1/5, whereas the aforementioned signaling pathways act through Smad2/3, when phosphorylated, resulting in activation of different genetic components driving different mesoderm programs(Kiecker et al., 2016). Finally, Wnt/β-Catenin signaling is involved in induction of early mesoderm. Multiple Wnt downstream components are enriched in the dorsal half of the embryo including β-Catenin, Dishevelled and glycogen synthase kinase 3 (GSK3) accumulate in the dorsal side(Miller et al., 1999). These major signaling pathways, among others, drive the mesendoderm and mesoderm genetic programs that give rise to the early mesoderm cell types of the embryo.

The first gene that was discovered to be essential for mesendoderm formation in vivo was a T-box gene, *Brachyury (T)*(Clark, 1934). Mice with homozygous mutations die during embryogenesis because of severe mesoderm formation and morphogenesis defects(Fujimoto and Yanagisawa, 1983). *Brachyury* is expressed in the primitive streak of the early embryo and has been shown to bind DNA as a tissue-specific transcription factor(Kispert et al., 1995) and when overexpressed in *Xenopus*, results in ectopic induction of mesoderm(Cunliffe and Smith, 1992). Another major inducer of mesendoderm is a T-box gene, *Eomesodermin (Eomes)* which precedes the expression of *T*. *Eomes* is induced by TGFβs and its expression in the primitive streak and early

mesendoderm results in induction of a broad range of transcription factors, including *T*(Ryan et al., 1996).

The signaling pathways and genes described thus far are the primary components driving the genetic programs that specify the early mesoderm subtypes that will be described in following sections. The signaling gradients created during this early stage, as well as the essential early transcription factors induce germ layer-specific transcriptional programs that regulate and drive subsequent cell fate decisions.

Lateral Plate and Paraxial Mesoderm Formation

Early embryonic mesoderm, which is formed by the primitive streak as described above, migrates and is organized along the mediolateral axis into three major mesoderm subtypes which lead to further specification of the mesoderm germ layer, the paraxial, intermediate, and lateral plate mesoderm. BMP is a major signaling component driving the further specification of the lateral plate mesoderm, specifically BMP4 whose antagonistic activity controls differentiation of mesoderm and is highly expressed in the lateral plate, promoting its formation(Tonegawa and Y. Takahashi, 1998). The lateral plate mesoderm goes on to form several, more specialized progenitor cells such as the coelom, somatic mesoderm, splanchnic mesoderm, and gut mesenchyme. The splanchnic mesoderm, which is more pertinent to this thesis, gives rise to the heart tissue such as the first and secondary heart field, endocardium and proepicardium, and outflow epicardium among others. Cardiac mesoderm relies

on BMP signaling for maturation of the tissue, specifically BMP2 and BMP4(Schultheiss et al., 1997).

Several important families of transcription factors have been implicated in cardiac cell commitment and differentiation. These include the Nk2 and GATA family factors. *Nkx2-5* is expressed in the cardiac mesoderm of vertebrate embryos and is thought to be one of the earliest markers of cardiac mesoderm differentiation. In mouse studies, disruption of the *Nkx2-5* gene has a severe heart formation defect(Lyons et al., 1995). GATA family members of zinc finger domain transcription factors have also been implicated in cardiac mesoderm development. *GATA4, 5*, and *6* are expressed in different domains of cardiac mesoderm prior to heart tube formation. GATA-binding sites have been discovered in the promoter regions of several important cardiac-specific genes such as *ANP*, *BNP*, and *α-MHC*(Mohun and Sparrow, 1997). In GATA4 knockout mice, it was discovered that the heart tube fails to form(Kuo et al., 1997). Finally, the LIM homeodomain protein Islet1 (Isl1) is important for self-renewal, differentiation and lineage commitment of cardiac precursors. Isl1 knockout mice have sever cardiac defects including the appearance of a common atrium and ventricular chamber, and the absence of the right ventricle and the outflow tract(Cai et al., 2003).

Paraxial mesoderm is located directly next to the neural tube in the early developing embryo and gives rise to tissues such as the somites, cartilage, skeletal muscle and endothelial cells. Paraxial mesoderm and its somite

derivatives are specified via inverse gradients of retinoic acid (RA) and FGF/Wnt signaling(Aulehla and Pourquié, 2010).

A paired-domain family member, the transcription factor Pax3, is an essential factor in the specification of paraxial mesoderm as well as down stream processes such as myogenesis. Pax3 is expressed in the dorsal neural tube, the posterior presomitic mesoderm, and the developing somites during development of the embryo(Magli et al., 2013). Pax3 expression in mouse embryonic stem cells was shown to be sufficient for activation of the myogenic program(Darabi et al., 2008) elucidating its importance in paraxial mesoderm genetic regulation. One of the more well known regulators of paraxial mesoderm, the Forkhead family transcription factor *FoxC2* is a direct downstream target of Pax3(Lagha et al., 2009). In addition, Pax3 was shown to induce *Myf5* expression in mouse ESCs(Magli et al., 2013). Furthermore, the mutation of Pax3 results in the absence of limb muscles in mice(Epstein et al., 1991) as well as other skeletal muscle and vertebral defects(Henderson et al., 1999; Schubert et al., 2001). It was also found that expression of Pax3 during differentiation of EBs resulted repression of key cardiac genes, *Nkx2-5* and cardiac *Troponin I* (cTnI)(Magli et al., 2013).

Intermediate Mesoderm Formation

The final early mesoderm subtype, the intermediate mesoderm, is flanked by the paraxial mesoderm and the lateral plate mesoderm in the developing embryo(Davidson, 2008) (Figure 1.1). Due to its “intermediate” positioning in the

embryo, the intermediate mesoderm is surrounded by a variety of signals coming from both the paraxial and lateral plate mesoderm, as well as the neural tube and notochord. Thus, the intermediate mesoderm is surrounded by multiple gradients of signaling highlighting its difficulty in recreating this cell type in vitro and making in vivo studies more difficult than the other, more segregated mesoderm subtypes. One such signaling pattern is provided by BMP. As mentioned above, high BMP4 signaling leads cells to adopt a lateral plate mesoderm fate at the expense of paraxial mesoderm. In the chick, it has been shown that high levels of BMP induce lateral plate mesoderm however, lower levels are effective at inducing intermediate mesoderm markers(R. G. James and Schultheiss, 2005). Likely due to the complex signaling conditions in vivo, BMP is not sufficient to induce intermediate mesoderm(Davidson et al., 2018). Studies in zebrafish indicate that FGF and Wnt signaling are essential in determining the number of intermediate mesoderm cells that form during development(Naylor et al., 2017; Ueno et al., 2007; Warga et al., 2013). This indicates that BMP, Wnt and FGF signaling gradients are all potentially involved in the formation of intermediate mesoderm in vivo. Finally, retinoic acid (RA) has been implicated in intermediate mesoderm induction. Specifically, RA has been shown to be required for specification of the pronephric cell fate, a downstream kidney progenitor(Cartry et al., 2006; Taira et al., 1994). Furthermore RA induces expression of LHX1, a known marker of several kidney lineages throughout development, in Xenopus as well as in human embryonic stem cells in vitro(Lam et al., 2014; Taira et al., 1994). Finally, *Aldh1a2* knockout mice show a delay in *Pax2* expression in the

intermediate mesoderm(Cartry et al., 2006). Since each of these signaling pathways operate in gradients in the developing embryo, it is likely that the intermediate mesoderm is exposed to “intermediate” levels of signaling from several or all of the discussed pathways.

Some evidence exists in mice that there may be inhibitory genetic pathways in the paraxial and intermediate mesoderms to help form the boundary between the two mesoderm subtypes(Wilm et al., 2004). The genes *Foxc1* and *Foxc2* have been shown to be involved in the delineation of paraxial versus intermediate mesoderm. These genes are expressed in the presomitic mesoderm and when both genes are ablated, mice show an expansion of the intermediate mesoderm markers *Lhx1* and *Osr1* with no observed effect on lateral plate mesoderm(Wilm et al., 2004). In the double knockout, the expansion is specific to the region where the intermediate mesoderm is specified in vivo, agreeing with the view that there is a specific region in the rostro-caudal axis that is conducive to intermediate mesoderm due to the specific signaling conditions in that region. Similarly to the boundary between paraxial and intermediate mesoderm, it was recently discovered in zebrafish that the bHLH transcription factor hand2, expressed in the lateral plate, is responsible for the boundary between the intermediate mesoderm and the lateral plate mesoderm(Perens et al., 2016). Embryos lacking hand2 showed an expansion of intermediate mesoderm while overexpression resulted in a reduction of intermediate mesoderm. Chromatin immunoprecipitation (CHIP) analysis has revealed Hand2-binding peaks on the *Pax2* and *Lhx1* genes indicating that Hand2 functions as a

direct repressor of intermediate mesoderm specification(Osterwalder et al., 2014). Contrariwise, knockdown of *Pax8* in *Xenopus* results in an expansion of angioblast cells and a contraction of the pronephric tubule(Buisson et al., 2015). This would suggest that *Hand2* and *Pax8* function as repressors to define the boundary between the intermediate and lateral plate mesoderm. Although it has been shown that several transcription factors regulate the boundaries between intermediate mesoderm and the other two early mesoderm subtypes, little evidence exists characterizing the transcriptional network that gives rise to and specifies the intermediate mesoderm cell fate. One member of this network, *LHX1* will be a central focus of this thesis.

The LIM Homeobox Transcription Factor, LHX1

During embryogenesis, the LIM homeodomain family of transcription factors is required for many processes including body axis formation and tissue specification(Dawid et al., 1998). The LIM family of transcription factors contain two cysteine-histidine rich motifs (LIM domains), a central homeodomain, and a C-terminal transactivation domain(Dawid et al., 1995) and are thought to function as components of protein interaction modules that serve to regulate parts of a transcriptional complex(Dawid et al., 1998). The LIM homeodomain transcription factor *Lhx1* is expressed in the early stages of gastrulation in *Xenopus* embryos(Taira et al., 1992) and is required in mouse and *Xenopus* for cell movement *in vivo* during gastrulation(Hukriede et al., 2003). This suggests a conserved role for Lhx1 in early embryonic patterning. *Lhx1* is one of the earliest

known genes expressed in the pronephros(Carroll and Vize, 1999; Tsang et al., 2000). Lhx1 is initially observed in the intermediate mesoderm stage of both mouse and Xenopus(Fujii et al., 1994; Taira et al., 1994) and is required for the proper patterning of the kidney field(Shawlot and Behringer, 1995) (Figure 1.3). Coexpression of kidney markers *lhx1* and *pax8* shows an enlarged kidney and ectopic formation of pronephric tubules(Carroll and Vize, 1999). In Xenopus, experiments focusing on the early stages of kidney development following the intermediate mesoderm have shown that Lhx1 plays a role in driving the specification of renal progenitor cells from the intermediate mesoderm(Cirio et al., 2011). In relation to signaling, lhx1 expression has been shown to be regulated by RA signaling, and RA responsive elements (RAREs) have been identified in the promoter of lhx1(Cartry et al., 2006). This suggests that RA signaling is required for further specification of the kidney field following intermediate mesoderm in addition to being essential for intermediate mesoderm formation following gastrulation. Further along during development of the embryonic kidney, Lhx1 has been shown to be required for ureteric bud morphogenesis and nephric vesicle patterning(Kobayashi et al., 2005; Pedersen et al., 2005). Therefore, LHX1 is involved in many different stages of kidney development and is indispensible in the formation of the organ.

Lhx1 has also been implicated in a number of other tissues and systems. One study demonstrated that LHX1 binds to the putative enhancers of Otx2 and Foxa2 to regulate the anterior mesoderm and midline development(Costello et al., 2015). Purkinje cells of the cerebellar cortex have been shown to be

controlled by Lhx1/5 to undergo dendritogenesis and spine morphogenesis in mice(Lui et al., 2017). Lhx1 has been implicated in the terminal differentiation and circadian function of the suprachiasmatic nucleus, the hypothalamic region of the brain that controls vertebrate circadian rhythm(Bedont et al., 2014). Another example of *LHX1* activity in tissues outside of the kidney is its association with congenital absence of the uterus and vagina (CAUV)(Zhang et al., 2017). This study identified a novel missense mutation in *LHX1* in a patient diagnosed with CAUV. This mutation was found to alter the transcriptional activity of *LHX1* and its downstream effect on GSC, which has been shown to be associated with the development of the urogenital system(Zhang et al., 2017). These studies highlight the importance of *LHX1* as a transcriptional activator in a variety of tissue systems and underscore its essentiality for the proper development of several tissue and cell types.

Kidney Development from the Intermediate Mesoderm

In the early embryo, between E8.5-9, cells committed to an intermediate mesoderm fate express the transcription factors Pax2, Pax8, Lhx1, Osr1 and Gata3(Marcotte et al., 2014) (Figure 1.2). Paired-box transcription factors Pax2 and Pax8 are essential early in kidney development as they direct part of the intermediate mesoderm to differentiate into the pronephric anlagen(Bouchard et al., 2002). The mammalian pronephros is non-functional and is likely an intermediate to the nephric duct(Vetter and Gibley, 1966). At E9 the nephric duct extends toward the urogenital sinus guided by FGF8 for maintenance of the

extending nephric duct, as well as guiding the duct toward the cloaca(Atsuta and Y. Takahashi, 2015; Attia et al., 2015). Again, Pax2 and Pax8 are involved in nephric duct formation by acting directly upstream of *Gata3* and *Lhx1*, establishing a transcriptional network(Boualia et al., 2013). Conditional deletion of *Gata3* results in a decrease in proliferation and spontaneous differentiation of the nephric duct(Grote et al., 2008), while similar deletion of *Lhx1* results in a delay in the elongation of the nephric duct again highlighting the role of Lhx1 in specialized kidney structure formation(Pedersen et al., 2005).

Wnt signaling becomes important for further differentiation of the kidney following nephric duct formation. Wnt9b is produced by the nephric duct as it migrates, and induces the formation of mesonephric tubules between E9-11(Carroll et al., 2005). At the rostral end of the embryo, transcription factors Wt1 and Six1 are required for formation of mesonephric tubules. The tubules become segmented and express proximal and distal tubule markers while fusing with the nephric duct(Sainio et al., 1997).

At E10.5 formation of the metanephric mesenchyme begins, providing the source of nephron progenitors and inducing the formation of the ureteric bud(Davidson et al., 2018). Nephron progenitor cells are a very commonly studied cell type for their therapeutic potential as they give rise to adult nephrons, which do not regenerate in humans and are a cause of chronic kidney disease (CKD). Cross-talk between the ureteric bud, the metanephric mesenchyme, and endothelial and stromal cells results in the formation of the metanephros. The completed kidney in the mouse is a collection of a highly branched collecting duct

system formed from the ureteric bud and roughly 13,000 nephrons that are formed from the metanephric mesenchyme(McMahon, 2016).

The transcription factors that play a role in the development of more specialized kidney tissue have also been reasonably defined. The metanephric mesenchyme is marked by expression of *Eya1, Six1, Six2, Osr1, Wt1* and *Pax2* among others(Davidson, 2008). *Eya1* has been shown to form a complex with the Six family of transcription factors, complexing with *Six2* by E10 in the metanephric mesenchyme(J. Xu and P.-X. Xu, 2015). *Osr1* and *Eya1* have been implicated as early metanephric mesenchyme regulators, as mutants of these genes do not form the metanephric mesenchyme(R. G. James et al., 2006; Sajithlal et al., 2005). Wilm's tumor 1 (Wt1), a zinc finger transcription factor essential for kidney development, promotes the survival and proliferation of the metanephric mesenchyme via Fgf signaling and repression of BMP4 signaling(Kreidberg et al., 1993; Motamedi et al., 2014). *Pax2*, another major transcription factor essential for kidney development, is responsible for inducing development of the ureteric bud in a complex with Eya1 and Hox11 paralogs(Gong et al., 2007). *Six2* is also required for nephron progenitor maintenance and inactivation of this gene results in ectopic differentiation of mesenchymal cells and the loss of the progenitor cells in the metanephric mesenchyme resulting in severe renal hypoplasia, thus making Six2 a good marker of nephron progenitor cells(Self et al., 2006). Understanding of development of the kidney has led to in vitro models that aim to recapitulate organogenesis of the kidney in organoid models(Boreström et al., 2018;

Czerniecki et al., 2018; Kumar et al., 2019; Li et al., 2016; Takasato et al., 2016). These models, as well as better characterization of the events governing early kidney generation, will facilitate development of therapies for various kidney diseases using techniques such as disease modeling and drug screening. Furthermore, such advanced kidney organoid models can lead to therapies involving regeneration of damaged or diseased kidney tissue(Morizane et al., 2017) (Figure 1.4).

Human vs. Mouse Kidney Organogenesis

Studies in the mouse have elucidated the mechanism by which the mammalian kidney develops. Genetic and cellular analyses have provided a detailed framework of the cell types and molecular processes by which the mammalian kidney is formed and the developmental programs responsible for initiating the many stages of kidney formation. Although the genetic programs by which the kidney is formed is largely conserved between mouse and human, the scale, organization and molecular identity of many major cell types differ significantly(Lindström et al., 2018). Although mouse kidney development has been thoroughly characterized, the same cannot be said of human kidney development. One important difference between mouse and human kidney is the scale of nephron proliferation. In the moue, approximately 13,000 nephrons make up the mature kidney whereas in humans the number of nephrons exceeds 1,000,000(Lindström et al., 2018). This is an important feature as low nephron number is linked to disease such as Chronic Kidney Disease (CKD). Some major

genes essential for mouse kidney development appear to be non-essential in humans. For example, Spp2 in mouse encodes for a phosphoprotein that binds to members of the TGF-β family and is specific to the proximal tubule(Tian et al., 2015). It is thought that this protein is absent in the human kidney. Additionally, research suggests that the gene *SIX1* is largely responsible for the large number of nephrons in the human kidney, while it plays a fleeting and early role in mouse kidney development(O'Brien et al., 2016). Thus it is important to employ modern molecular and cellular analyses to provide a more robust understanding of human kidney development, as the differences between mouse and human are likely to affect our understanding kidney disease and development, and our ability to treat disease.

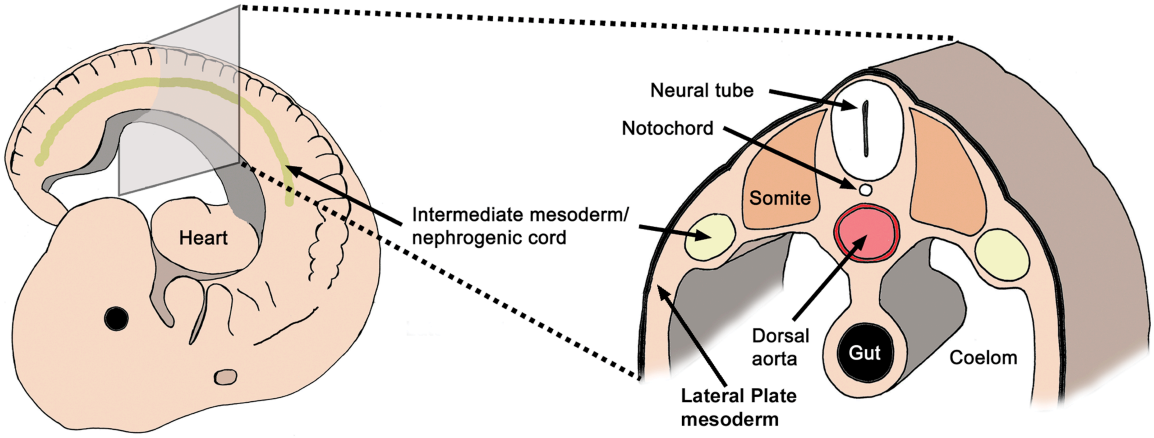


Figure 1.1 – Location of the intermediate mesoderm/nephrogenic chord. Schematic representation of a mouse embryo at E9.5 (left) and a cross-section through the trunk (right) showing the position of the intermediate mesoderm/nephrogenic cord. *(Adapted from Davidson et al., 2008).*

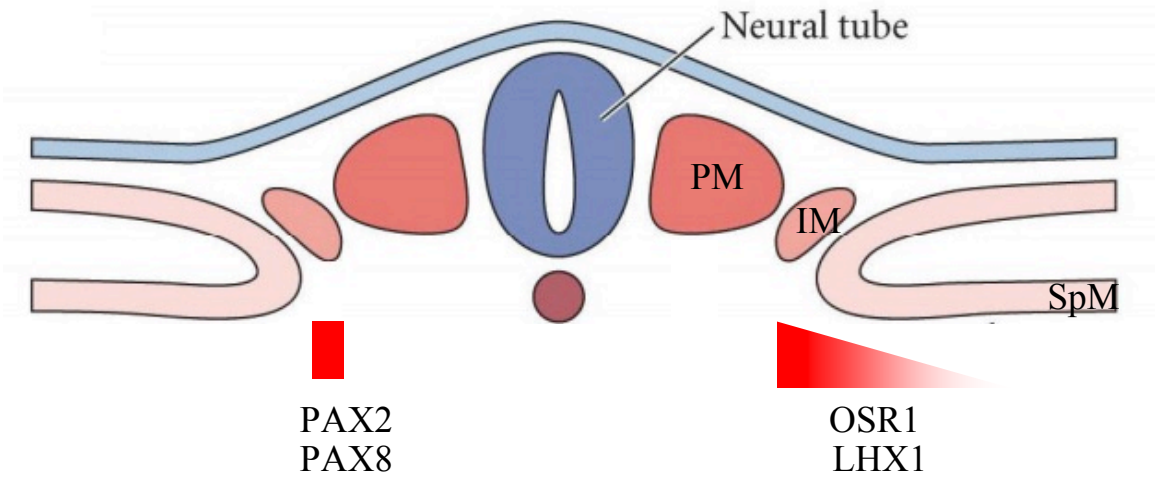


Figure 1.2 – IM marker expression in early mesoderm. Illustration of the early developing embryo directly following gastrulation. Highlighted are several key transcription factors involved in specification of intermediate mesoderm derivatives and markers of intermediate mesoderm. Red gradients represent the specificity of expression in the mesoderm subtypes. *(Adapted from http://www.d.umn.edu/~pschoff/documents/ParaxialandIntermediateMesoderm_002.pdf).*

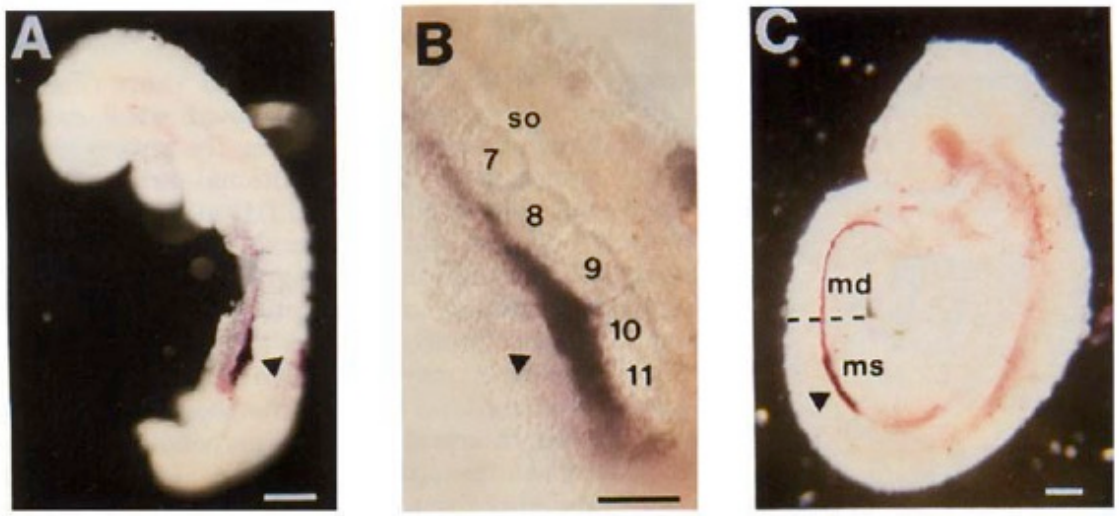


Figure 1.3 - Spatial distribution of Liml mRNA in mouse embryos as visualized by whole mount in situ hybridization. A: Lateral view of an E 8.5 embryo. Staining is detected in the presumptive mesonephric region indicated by a solid triangle (A,B,C). B: Higher magnification of A. Somites are numbered. C: Lateral view of an E 9.5 embryo. Staining is detected in the mesonephric duct (solid triangle). *(Adapted from Fujii et al., 1994).*

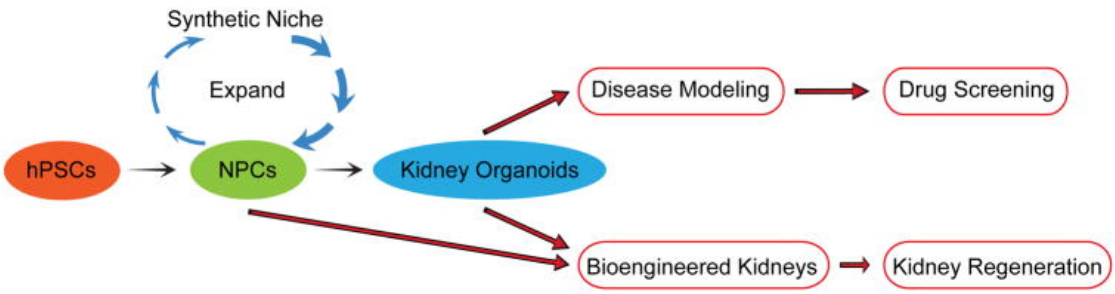


Figure 1.4 – Future applications of NPCs and kidney organoids for drug screening and kidney regeneration. hPSCs: human pluripotent stem cells. NPCs: nephron progenitor cells. *(Adapted from Morizane et al., 2017).*

REFERENCES

Alev, C., Wu, Y., Nakaya, Y., Sheng, G., 2013. Decoupling of amniote gastrulation and streak formation reveals a morphogenetic unity in vertebrate mesoderm induction. Development 140, 2691–2696. doi:10.1242/dev.094318

Arnold, S.J., Robertson, E.J., 2009. Making a commitment: cell lineage allocation and axis patterning in the early mouse embryo. Nat. Rev. Mol. Cell Biol. 10, 91–103. doi:10.1038/nrm2618

Atsuta, Y., Takahashi, Y., 2015. FGF8 coordinates tissue elongation and cell epithelialization during early kidney tubulogenesis. Development 142, 2329–2337. doi:10.1242/dev.122408

Attia, L., Schneider, J., Yelin, R., Schultheiss, T.M., 2015. Collective cell migration of the nephric duct requires FGF signaling. Dev. Dyn. 244, 157–167. doi:10.1002/dvdy.24241

Aulehla, A., Pourquié, O., 2010. Signaling gradients during paraxial mesoderm development. Cold Spring Harb Perspect Biol 2, a000869–a000869. doi:10.1101/cshperspect.a000869

Bedont, J.L., LeGates, T.A., Slat, E.A., Byerly, M.S., Wang, H., Hu, J., Rupp, A.C., Qian, J., Wong, G.W., Herzog, E.D., Hattar, S., Blackshaw, S., 2014.

Lhx1 controls terminal differentiation and circadian function of the suprachiasmatic nucleus. Cell Rep 7, 609–622. doi:10.1016/j.celrep.2014.03.060

Boreström, C., Jonebring, A., Guo, J., Palmgren, H., Cederblad, L., Forslöw, A., Svensson, A., Söderberg, M., Reznichenko, A., Nyström, J., Patrakka, J., Hicks, R., Maresca, M., Valastro, B., Collén, A., 2018. A CRISP(e)R view on kidney organoids allows generation of an induced pluripotent stem cell-derived kidney model for drug discovery. Kidney Int. 94, 1099–1110. doi:10.1016/j.kint.2018.05.003

Boualia, S.K., Gaitan, Y., Tremblay, M., Sharma, R., Cardin, J., Kania, A., Bouchard, M., 2013. A core transcriptional network composed of Pax2/8, Gata3 and Lim1 regulates key players of pro/mesonephros morphogenesis. Developmental biology 382, 555–566. doi:10.1016/j.ydbio.2013.07.028

Bouchard, M., Souabni, A., Mandler, M., Neubüser, A., Busslinger, M., 2002. Nephric lineage specification by Pax2 and Pax8. Genes & development 16, 2958–2970. doi:10.1101/gad.240102

Boyer, L.A., Lee, T.I., Cole, M.F., Johnstone, S.E., Levine, S.S., Zucker, J.P., Guenther, M.G., Kumar, R.M., Murray, H.L., Jenner, R.G., Gifford, D.K., Melton, D.A., Jaenisch, R., Young, R.A., 2005. Core transcriptional regulatory

circuitry in human embryonic stem cells. Cell 122, 947–956. doi:10.1016/j.cell.2005.08.020

Brook, F.A., Gardner, R.L., 1997. The origin and efficient derivation of embryonic stem cells in the mouse. Proceedings of the National Academy of Sciences of the United States of America 94, 5709–5712.

Buisson, I., Le Bouffant, R., Futel, M., Riou, J.-F., Umbhauer, M., 2015. Pax8 and Pax2 are specifically required at different steps of Xenopus pronephros development. Developmental biology 397, 175–190. doi:10.1016/j.ydbio.2014.10.022

Burdsal, C.A., Flannery, M.L., Pedersen, R.A., 1998. FGF-2 alters the fate of mouse epiblast from ectoderm to mesoderm in vitro. Developmental biology 198, 231–244.

Cai, C.-L., Liang, X., Shi, Y., Chu, P.-H., Pfaff, S.L., Chen, J., Evans, S., 2003. Isl1 identifies a cardiac progenitor population that proliferates prior to differentiation and contributes a majority of cells to the heart. Dev. Cell 5, 877–889.

Cao, Y., Zhao, J., Sun, Z., Zhao, Z., Postlethwait, J., Meng, A., 2004. fgf17b, a novel member of Fgf family, helps patterning zebrafish embryos.

Developmental biology 271, 130–143. doi:10.1016/j.ydbio.2004.03.032

Carroll, T.J., Park, J.-S., Hayashi, S., Majumdar, A., McMahon, A.P., 2005. Wnt9b plays a central role in the regulation of mesenchymal to epithelial transitions underlying organogenesis of the mammalian urogenital system. Dev. Cell 9, 283–292. doi:10.1016/j.devcel.2005.05.016

Carroll, T.J., Vize, P.D., 1999. Synergism between Pax-8 and lim-1 in embryonic kidney development. Developmental biology 214, 46–59. doi:10.1006/dbio.1999.9414

Cartry, J., Nichane, M., Ribes, V., Colas, A., Riou, J.-F., Pieler, T., Dollé, P., Bellefroid, E.J., Umbhauer, M., 2006. Retinoic acid signalling is required for specification of pronephric cell fate. Developmental biology 299, 35–51. doi:10.1016/j.ydbio.2006.06.047

Cirio, M.C., Hui, Z., Haldin, C.E., Cosentino, C.C., Stuckenholz, C., Chen, X., Hong, S.-K., Dawid, I.B., Hukriede, N.A., 2011. Lhx1 is required for specification of the renal progenitor cell field. PLoS ONE 6, e18858. doi:10.1371/journal.pone.0018858

Clark, F.H., 1934. Linkage Studies of Brachyury (Short Tail) in the House Mouse. Proceedings of the National Academy of Sciences of the United States of

America 20, 276–279.

Costello, I., Nowotschin, S., Sun, X., Mould, A.W., Hadjantonakis, A.-K., Bikoff, E.K., Robertson, E.J., 2015. Lhx1 functions together with Otx2, Foxa2, and Ldb1 to govern anterior mesendoderm, node, and midline development. Genes & development 29, 2108–2122. doi:10.1101/gad.268979.115

Cunliffe, V., Smith, J.C., 1992. Ectopic mesoderm formation in Xenopus embryos caused by widespread expression of a Brachyury homologue. Nature 358, 427–430. doi:10.1038/358427a0

Czerniecki, S.M., Cruz, N.M., Harder, J.L., Menon, R., Annis, J., Otto, E.A., Gulieva, R.E., Islas, L.V., Kim, Y.K., Tran, L.M., Martins, T.J., Pippin, J.W., Fu, H., Kretzler, M., Shankland, S.J., Himmelfarb, J., Moon, R.T., Paragas, N., Freedman, B.S., 2018. High-Throughput Screening Enhances Kidney Organoid Differentiation from Human Pluripotent Stem Cells and Enables Automated Multidimensional Phenotyping. Cell stem cell 22, 929–940.e4. doi:10.1016/j.stem.2018.04.022

Darabi, R., Gehlbach, K., Bachoo, R.M., Kamath, S., Osawa, M., Kamm, K.E., Kyba, M., Perlingeiro, R.C.R., 2008. Functional skeletal muscle regeneration from differentiating embryonic stem cells. Nat. Med. 14, 134–143. doi:10.1038/nm1705

Davidson, A.J., 2008. Mouse kidney development. StemBook. doi:10.3824/stembook.1.34.1

Davidson, A.J., Lewis, P., Przepiorski, A., Sander, V., 2018. Turning mesoderm into kidney. Semin. Cell Dev. Biol. doi:10.1016/j.semcdb.2018.08.016

Dawid, I.B., Breen, J.J., Toyama, R., 1998. LIM domains: multiple roles as adapters and functional modifiers in protein interactions. Trends Genet. 14, 156–162.

Dawid, I.B., Toyama, R., Taira, M., 1995. LIM domain proteins. C. R. Acad. Sci. III, Sci. Vie 318, 295–306.

Epstein, D.J., Vekemans, M., Gros, P., 1991. Splotch (Sp2H), a mutation affecting development of the mouse neural tube, shows a deletion within the paired homeodomain of Pax-3. Cell 67, 767–774.

Evans, M.J., Kaufman, M.H., 1981. Establishment in culture of pluripotential cells from mouse embryos. Nature 292, 154–156.

Fujii, T., Pichel, J.G., Taira, M., Toyama, R., Dawid, I.B., Westphal, H., 1994. Expression patterns of the murine LIM class homeobox gene lim1 in the

developing brain and excretory system. Dev. Dyn. 199, 73–83. doi:10.1002/aja.1001990108

Fujimoto, H., Yanagisawa, K.O., 1983. Defects in the archenteron of mouse embryos homozygous for the T-mutation. Differentiation 25, 44–47.

Gong, K.-Q., Yallowitz, A.R., Sun, H., Dressler, G.R., Wellik, D.M., 2007. A Hox-Eya-Pax complex regulates early kidney developmental gene expression. Molecular and cellular biology 27, 7661–7668. doi:10.1128/MCB.00465-07

Grote, D., Boualia, S.K., Souabni, A., Merkel, C., Chi, X., Costantini, F., Carroll, T., Bouchard, M., 2008. Gata3 acts downstream of beta-catenin signaling to prevent ectopic metanephric kidney induction. PLoS Genet. 4, e1000316. doi:10.1371/journal.pgen.1000316

Harrison, S.E., Sozen, B., Christodoulou, N., Kyprianou, C., Zernicka-Goetz, M., 2017. Assembly of embryonic and extraembryonic stem cells to mimic embryogenesis in vitro. Science 356, eaal1810. doi:10.1126/science.aal1810

Henderson, D.J., Conway, S.J., Copp, A.J., 1999. Rib truncations and fusions in the Sp2H mouse reveal a role for Pax3 in specification of the ventro-lateral and posterior parts of the somite. Developmental biology 209, 143–158. doi:10.1006/dbio.1999.9215

Hukriede, N.A., Tsang, T.E., Habas, R., Khoo, P.-L., Steiner, K., Weeks, D.L., Tam, P.P.L., Dawid, I.B., 2003. Conserved requirement of Lim1 function for cell movements during gastrulation. Dev. Cell 4, 83–94.

James, D., Levine, A.J., Besser, D., Hemmati-Brivanlou, A., 2005. TGFbeta/activin/nodal signaling is necessary for the maintenance of pluripotency in human embryonic stem cells. Development 132, 1273–1282. doi:10.1242/dev.01706

James, R.G., Kamei, C.N., Wang, Q., Jiang, R., Schultheiss, T.M., 2006. Odd-skipped related 1 is required for development of the metanephric kidney and regulates formation and differentiation of kidney precursor cells. Development 133, 2995–3004. doi:10.1242/dev.02442

James, R.G., Schultheiss, T.M., 2005. Bmp signaling promotes intermediate mesoderm gene expression in a dose-dependent, cell-autonomous and translation-dependent manner. Developmental biology 288, 113–125. doi:10.1016/j.ydbio.2005.09.025

Jiang, J., Chan, Y.-S., Loh, Y.-H., Cai, J., Tong, G.-Q., Lim, C.-A., Robson, P., Zhong, S., Ng, H.-H., 2008. A core Klf circuitry regulates self-renewal of embryonic stem cells. Nat. Cell Biol. 10, 353–360. doi:10.1038/ncb1698

Jones, C.M., Kuehn, M.R., Hogan, B.L., Smith, J.C., Wright, C.V., 1995. Nodal-related signals induce axial mesoderm and dorsalize mesoderm during gastrulation. Development 121, 3651–3662.

Kiecker, C., Bates, T., Bell, E., 2016. Molecular specification of germ layers in vertebrate embryos. Cell. Mol. Life Sci. 73, 923–947. doi:10.1007/s00018-015-2092-y

Kim, M.O., Kim, S.-H., Cho, Y.-Y., Nadas, J., Jeong, C.-H., Yao, K., Kim, D.J., Yu, D.-H., Keum, Y.-S., Lee, K.-Y., Huang, Z., Bode, A.M., Dong, Z., 2012. ERK1 and ERK2 regulate embryonic stem cell self-renewal through phosphorylation of Klf4. Nat. Struct. Mol. Biol. 19, 283–290. doi:10.1038/nsmb.2217

Kimelman, D., Kirschner, M., 1987. Synergistic induction of mesoderm by FGF and TGF-beta and the identification of an mRNA coding for FGF in the early Xenopus embryo. Cell 51, 869–877.

Kispert, A., Koschorz, B., Herrmann, B.G., 1995. The T protein encoded by Brachyury is a tissue-specific transcription factor. EMBO J. 14, 4763–4772.

Kobayashi, A., Kwan, K.M., Carroll, T.J., McMahon, A.P., Mendelsohn, C.L.,

Behringer, R.R., 2005. Distinct and sequential tissue-specific activities of the LIM-class homeobox gene Lim1 for tubular morphogenesis during kidney development. Development 132, 2809–2823. doi:10.1242/dev.01858

Köster, M., Plessow, S., Clement, J.H., Lorenz, A., Tiedemann, H., Knöchel, W., 1991. Bone morphogenetic protein 4 (BMP-4), a member of the TGF-beta family, in early embryos of Xenopus laevis: analysis of mesoderm inducing activity. Mech. Dev. 33, 191–199.

Kreidberg, J.A., Sariola, H., Loring, J.M., Maeda, M., Pelletier, J., Housman, D., Jaenisch, R., 1993. WT-1 is required for early kidney development. Cell 74, 679–691.

Kumar, S.V., Er, P.X., Lawlor, K.T., Motazedian, A., Scurr, M., Ghobrial, I., Combes, A.N., Zappia, L., Oshlack, A., Stanley, E.G., Little, M.H., 2019. Kidney micro-organoids in suspension culture as a scalable source of human pluripotent stem cell-derived kidney cells. Development 146, dev172361. doi:10.1242/dev.172361

Kuo, C.T., Morrisey, E.E., Anandappa, R., Sigrist, K., Lu, M.M., Parmacek, M.S., Soudais, C., Leiden, J.M., 1997. GATA4 transcription factor is required for ventral morphogenesis and heart tube formation. Genes & development 11, 1048–1060.

Lagha, M., Brunelli, S., Messina, G., Cumano, A., Kume, T., Relaix, F., Buckingham, M.E., 2009. Pax3:Foxc2 reciprocal repression in the somite modulates muscular versus vascular cell fate choice in multipotent progenitors. Dev. Cell 17, 892–899. doi:10.1016/j.devcel.2009.10.021

Lam, A.Q., Freedman, B.S., Morizane, R., Lerou, P.H., Valerius, M.T., Bonventre, J.V., 2014. Rapid and efficient differentiation of human pluripotent stem cells into intermediate mesoderm that forms tubules expressing kidney proximal tubular markers. J. Am. Soc. Nephrol. 25, 1211–1225. doi:10.1681/ASN.2013080831

Li, Z., Araoka, T., Wu, J., Liao, H.-K., Li, M., Lazo, M., Zhou, B., Sui, Y., Wu, M.-Z., Tamura, I., Xia, Y., Beyret, E., Matsusaka, T., Pastan, I., Rodriguez Esteban, C., Guillen, I., Guillen, P., Campistol, J.M., Izpisua Belmonte, J.C., 2016. 3D Culture Supports Long-Term Expansion of Mouse and Human Nephrogenic Progenitors. Cell stem cell 19, 516–529. doi:10.1016/j.stem.2016.07.016

Lindström, N.O., McMahon, J.A., Guo, J., Tran, T., Guo, Q., Rutledge, E., Parvez, R.K., Saribekyan, G., Schuler, R.E., Liao, C., Kim, A.D., Abdelhalim, A., Ruffins, S.W., Thornton, M.E., Baskin, L., Grubbs, B., Kesselman, C., McMahon, A.P., 2018. Conserved and Divergent Features of Human and

Mouse Kidney Organogenesis. J. Am. Soc. Nephrol. 29, 785–805. doi:10.1681/ASN.2017080887

Lui, N.C., Tam, W.Y., Gao, C., Huang, J.-D., Wang, C.C., Jiang, L., Yung, W.H., Kwan, K.M., 2017. Lhx1/5 control dendritogenesis and spine morphogenesis of Purkinje cells via regulation of Espin. Nat Commun 8, 15079. doi:10.1038/ncomms15079

Lyons, I., Parsons, L.M., Hartley, L., Li, R., Andrews, J.E., Robb, L., Harvey, R.P., 1995. Myogenic and morphogenetic defects in the heart tubes of murine embryos lacking the homeo box gene Nkx2-5. Genes & development 9, 1654–1666.

Magli, A., Schnettler, E., Rinaldi, F., Bremer, P., Perlingeiro, R.C.R., 2013. Functional dissection of Pax3 in paraxial mesoderm development and myogenesis. Stem Cells 31, 59–70. doi:10.1002/stem.1254

Marcotte, M., Sharma, R., Bouchard, M., 2014. Gene regulatory network of renal primordium development. Pediatr. Nephrol. 29, 637–644. doi:10.1007/s00467-013-2635-0

Martin, G.R., 1981. Isolation of a pluripotent cell line from early mouse embryos cultured in medium conditioned by teratocarcinoma stem cells. Proceedings

of the National Academy of Sciences of the United States of America 78, 7634–7638.

McMahon, A.P., 2016. Development of the Mammalian Kidney. Curr. Top. Dev. Biol. 117, 31–64. doi:10.1016/bs.ctdb.2015.10.010

Meissner, A., 2010. Epigenetic modifications in pluripotent and differentiated cells. Nat. Biotechnol. 28, 1079–1088. doi:10.1038/nbt.1684

Miller, J.R., Rowning, B.A., Larabell, C.A., Yang-Snyder, J.A., Bates, R.L., Moon, R.T., 1999. Establishment of the dorsal-ventral axis in Xenopus embryos coincides with the dorsal enrichment of dishevelled that is dependent on cortical rotation. The Journal of cell biology 146, 427–437.

Mohun, T., Sparrow, D., 1997. Early steps in vertebrate cardiogenesis. Curr. Opin. Genet. Dev. 7, 628–633.

Morizane, R., Miyoshi, T., Bonventre, J.V., 2017. Concise Review: Kidney Generation with Human Pluripotent Stem Cells. Stem Cells 35, 2209–2217. doi:10.1002/stem.2699

Motamedi, F.J., Badro, D.A., Clarkson, M., Lecca, M.R., Bradford, S.T., Buske, F.A., Saar, K., Hübner, N., Brändli, A.W., Schedl, A., 2014. WT1 controls

antagonistic FGF and BMP-pSMAD pathways in early renal progenitors. Nat Commun 5, 4444. doi:10.1038/ncomms5444

Naylor, R.W., Han, H.I., Hukriede, N.A., Davidson, A.J., 2017. Wnt8a expands the pool of embryonic kidney progenitors in zebrafish. Developmental biology 425, 130–141. doi:10.1016/j.ydbio.2017.03.027

O'Brien, L.L., Guo, Q., Lee, Y., Tran, T., Benazet, J.-D., Whitney, P.H., Valouev, A., McMahon, A.P., 2016. Differential regulation of mouse and human nephron progenitors by the Six family of transcriptional regulators. Development 143, 595–608. doi:10.1242/dev.127175

Osterwalder, M., Speziale, D., Shoukry, M., Mohan, R., Ivanek, R., Kohler, M., Beisel, C., Wen, X., Scales, S.J., Christoffels, V.M., Visel, A., Lopez-Rios, J., Zeller, R., 2014. HAND2 targets define a network of transcriptional regulators that compartmentalize the early limb bud mesenchyme. Dev. Cell 31, 345–357. doi:10.1016/j.devcel.2014.09.018

Pedersen, A., Skjong, C., Shawlot, W., 2005. Lim 1 is required for nephric duct extension and ureteric bud morphogenesis. Developmental biology 288, 571–581. doi:10.1016/j.ydbio.2005.09.027

Perens, E.A., Garavito-Aguilar, Z.V., Guio-Vega, G.P., Peña, K.T., Schindler,

Y.L., Yelon, D., 2016. Hand2 inhibits kidney specification while promoting vein formation within the posterior mesoderm. Elife 5, 231. doi:10.7554/eLife.19941

Polo, J.M., Anderssen, E., Walsh, R.M., Schwarz, B.A., Nefzger, C.M., Lim, S.M., Borkent, M., Apostolou, E., Alaei, S., Cloutier, J., Bar-Nur, O., Cheloufi, S., Stadtfeld, M., Figueroa, M.E., Robinton, D., Natesan, S., Melnick, A., Zhu, J., Ramaswamy, S., Hochedlinger, K., 2012. A molecular roadmap of reprogramming somatic cells into iPS cells. Cell 151, 1617–1632. doi:10.1016/j.cell.2012.11.039

Ryan, K., Garrett, N., Mitchell, A., GURDON, J.B., 1996. Eomesodermin, a key early gene in Xenopus mesoderm differentiation. Cell 87, 989–1000.

Sainio, K., Hellstedt, P., Kreidberg, J.A., Saxén, L., Sariola, H., 1997. Differential regulation of two sets of mesonephric tubules by WT-1. Development 124, 1293–1299.

Sajithlal, G., Zou, D., Silvius, D., Xu, P.-X., 2005. Eya 1 acts as a critical regulator for specifying the metanephric mesenchyme. Developmental biology 284, 323–336. doi:10.1016/j.ydbio.2005.05.029

Schubert, F.R., Tremblay, P., Mansouri, A., Faisst, A.M., Kammandel, B., Lumsden, A., Gruss, P., Dietrich, S., 2001. Early mesodermal phenotypes in

splotch suggest a role for Pax3 in the formation of epithelial somites. Dev. Dyn. 222, 506–521. doi:10.1002/dvdy.1211

Schultheiss, T.M., Burch, J.B., Lassar, A.B., 1997. A role for bone morphogenetic proteins in the induction of cardiac myogenesis. Genes & development 11, 451–462.

Self, M., Lagutin, O.V., Bowling, B., Hendrix, J., Cai, Y., Dressler, G.R., Oliver, G., 2006. Six2 is required for suppression of nephrogenesis and progenitor renewal in the developing kidney. EMBO J. 25, 5214–5228. doi:10.1038/sj.emboj.7601381

Shawlot, W., Behringer, R.R., 1995. Requirement for Lim1 in head-organizer function. Nature 374, 425–430. doi:10.1038/374425a0

Shen, M.M., 2007. Nodal signaling: developmental roles and regulation. Development 134, 1023–1034. doi:10.1242/dev.000166

Silva, J., Barrandon, O., Nichols, J., Kawaguchi, J., Theunissen, T.W., Smith, A., 2008. Promotion of reprogramming to ground state pluripotency by signal inhibition. PLoS biology 6, e253. doi:10.1371/journal.pbio.0060253

Sumi, T., Tsuneyoshi, N., Nakatsuji, N., Suemori, H., 2007. Apoptosis and

differentiation of human embryonic stem cells induced by sustained activation of c-Myc. Oncogene 26, 5564–5576. doi:10.1038/sj.onc.1210353

Taira, M., Jamrich, M., Good, P.J., Dawid, I.B., 1992. The LIM domain-containing homeo box gene Xlim-1 is expressed specifically in the organizer region of Xenopus gastrula embryos. Genes & development 6, 356–366.

Taira, M., Otani, H., Jamrich, M., Dawid, I.B., 1994. Expression of the LIM class homeobox gene Xlim-1 in pronephros and CNS cell lineages of Xenopus embryos is affected by retinoic acid and exogastrulation. Development 120, 1525–1536.

Takahashi, K., Tanabe, K., Ohnuki, M., Narita, M., Ichisaka, T., Tomoda, K., Yamanaka, S., 2007. Induction of pluripotent stem cells from adult human fibroblasts by defined factors. Cell 131, 861–872. doi:10.1016/j.cell.2007.11.019

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

Takasato, M., Er, P.X., Chiu, H.S., Little, M.H., 2016. Generation of kidney organoids from human pluripotent stem cells. Nat Protoc 11, 1681–1692.

doi:10.1038/nprot.2016.098

Tam, P.P., Behringer, R.R., 1997. Mouse gastrulation: the formation of a mammalian body plan. Mech. Dev. 68, 3–25.

Thomson, J.A., Itskovitz-Eldor, J., Shapiro, S.S., Waknitz, M.A., Swiergiel, J.J., Marshall, V.S., Jones, J.M., 1998. Embryonic stem cell lines derived from human blastocysts. Science 282, 1145–1147.

Tian, H., Li, C.-S., Zhao, K.-W., Wang, J.C., Duarte, M.E.L., David, C.L., Phan, K., Atti, E., Brochmann, E.J., Murray, S.S., 2015. A carboxy terminal BMP/TGF-β binding site in secreted phosphoprotein 24 kD independently affects BMP-2 activity. J. Cell. Biochem. 116, 667–676. doi:10.1002/jcb.25023

Tonegawa, A., Takahashi, Y., 1998. Somitogenesis controlled by Noggin. Developmental biology 202, 172–182. doi:10.1006/dbio.1998.8895

Tsang, T.E., Shawlot, W., Kinder, S.J., Kobayashi, A., Kwan, K.M., Schughart, K., Kania, A., Jessell, T.M., Behringer, R.R., Tam, P.P., 2000. Lim1 activity is required for intermediate mesoderm differentiation in the mouse embryo. Developmental biology 223, 77–90. doi:10.1006/dbio.2000.9733

Ueno, S., Weidinger, G., Osugi, T., Kohn, A.D., Golob, J.L., Pabon, L., Reinecke, H., Moon, R.T., Murry, C.E., 2007. Biphasic role for Wnt/beta-catenin signaling in cardiac specification in zebrafish and embryonic stem cells. Proceedings of the National Academy of Sciences of the United States of America 104, 9685–9690. doi:10.1073/pnas.0702859104

Vallier, L., Alexander, M., Pedersen, R.A., 2005. Activin/Nodal and FGF pathways cooperate to maintain pluripotency of human embryonic stem cells. J. Cell. Sci. 118, 4495–4509. doi:10.1242/jcs.02553

Vetter, M.R., Gibley, C.W., 1966. Morphogenesis and histochemistry of the developing mouse kidney. J. Morphol. 120, 135–155. doi:10.1002/jmor.1051200203

Warga, R.M., Mueller, R.L., Ho, R.K., Kane, D.A., 2013. Zebrafish Tbx16 regulates intermediate mesoderm cell fate by attenuating Fgf activity. Developmental biology 383, 75–89. doi:10.1016/j.ydbio.2013.08.018

Wilm, B., James, R.G., Schultheiss, T.M., Hogan, B.L.M., 2004. The forkhead genes, Foxc1 and Foxc2, regulate paraxial versus intermediate mesoderm cell fate. Developmental biology 271, 176–189. doi:10.1016/j.ydbio.2004.03.034

Wilson, V., Olivera-Martinez, I., Storey, K.G., 2009. Stem cells, signals and vertebrate body axis extension. Development 136, 1591–1604. doi:10.1242/dev.021246

Xu, J., Xu, P.-X., 2015. Eya-six are necessary for survival of nephrogenic cord progenitors and inducing nephric duct development before ureteric bud formation. Dev. Dyn. 244, 866–873. doi:10.1002/dvdy.24282

Ying, Q.-L., Wray, J., Nichols, J., Batlle-Morera, L., Doble, B., Woodgett, J., Cohen, P., Smith, A., 2008. The ground state of embryonic stem cell self-renewal. Nature 453, 519–523. doi:10.1038/nature06968

Zhang, W., Zhou, X., Liu, L., Zhu, Y., Liu, C., Pan, H., Xing, Q., Wang, J., Wang, X., Zhang, X., Cao, Y., Wang, B., 2017. Identification and functional analysis of a novel LHX1 mutation associated with congenital absence of the uterus and vagina. Oncotarget 8, 8785–8790. doi:10.18632/oncotarget.14455

Zwaka, T.P., Thomson, J.A., 2005. A germ cell origin of embryonic stem cells? Development 132, 227–233. doi:10.1242/dev.01586

CHAPTER 2

LHX1 CONTROLS INTERMEDIATE MESODERM FORMATION FROM HUMAN EMBRYONIC STEM CELLS

Boward BR, Zhang L, Wu T, Tam P, Li Z, Dalton S. To be submitted to *Genes and Development*.

ABSTRACT

The adult kidney is derived from the intermediate mesoderm, an early mesoderm subtype located between the paraxial and lateral plate mesoderm. Genetic regulation and fate determination of intermediate mesoderm are poorly understood. The homeobox transcription factor LHX1 is expressed in the developing intermediate mesoderm and is required for specification of several kidney lineages, but its function in determining IM cell fate is unknown. Here we show that LHX1 is both necessary and sufficient for differentiation of human ESCs to intermediate mesoderm. RNA-seq and ATAC-seq have unveiled an LHX1 interaction network that positions LHX1 as a master regulator of IM formation. LHX1 regulates IM gene expression through binding to distal elements and its binding is required for the function of IM enhancers. This is the first report describing a role for a transcription factor in IM lineage specification. Further understanding of the intermediate mesoderm transcriptional fate map should enable higher efficiency kidney differentiations for use in developmental and disease models.

INTRODUCTION

The body plan of vertebrate embryos results from a process called gastrulation, or formation of the three germ layers. This results in the spatial organization of tissues including the somitic, or paraxial, mesoderm, lateral plate mesoderm, and intermediate mesoderm, whose arrangement along the body axis is highly conserved among vertebrates (Wilson et al., 2009). These early

mesoderm subtypes gives rise to adults tissues such as skeletal muscle, heart and kidneys.

The intermediate mesoderm (IM) is flanked by the paraxial mesoderm and the lateral plate mesoderm in the developing embryo. Due to its “intermediate” positioning in the embryo, the intermediate mesoderm is surrounded by a variety of signals coming from both the paraxial and lateral plate mesoderm, as well as the neural tube and notochord(Davidson et al., 2018; Naylor et al., 2017; Ueno et al., 2007; Warga et al., 2013), highlighting the difficulty in recreating this cell type in vitro.

One essential signal, retinoic acid (RA), has been implicated in intermediate mesoderm induction. Specifically, RA has been shown to be required for specification of the pronephric cell fate, a downstream kidney progenitor(Cartry et al., 2006; Taira et al., 1994). Furthermore, RA induces expression of LHX1, a known marker of several kidney lineages throughout development, in Xenopus as well as in human(Cartry et al., 2006; Lam et al., 2014; Taira et al., 1994).

LHX1 is a member of the LIM homeodomain family of transcription factors, which have been shown to be required for processes including body axis formation and tissue specification(Dawid et al., 1998). Lhx1 is initially observed in the intermediate mesoderm stage of both mouse and Xenopus(Fujii et al., 1994; Taira et al., 1994) and is one of the earliest known genes expressed in the pronephros(Carroll and Vize, 1999; Tsang et al., 2000). Coexpression of kidney markers *lhx1* and *pax8* shows an enlarged kidney and ectopic formation of

pronephric tubules(Carroll and Vize, 1999). Further along during development of the embryonic kidney, Lhx1 has been shown to be required for ureteric bud morphogenesis and nephric vesicle patterning(Kobayashi et al., 2005; Pedersen et al., 2005). Importantly, Lhx1 has been shown to be required for the proper patterning of the kidney, as Lhx1 null mice show a loss of the kidney field(Shawlot and Behringer, 1995). The role of LHX1 in the development of specialized kidney cells has been extensively studied, however, its role in specification of the intermediate mesoderm is completely unknown.

Pluripotent stem cells (PSCs) are a valuable model system to study human development as it can provide detailed insights into the molecular mechanisms of tissue specification, and holds tremendous potential for genome-wide studies and regenerative medicine due to the ease of generating large cell numbers *in vitro*. To date, several labs have developed protocols for differentiation of intermediate mesoderm from human pluripotent stem cells (hPSCs) that express key IM lineage markers, and have potential for differentiation to more specialized kidney cell types downstream of the IM(Araoka et al., 2014; Kumar et al., 2015; Lam et al., 2014; Mae et al., 2013). These studies however, do not address the transcriptional program or the mechanisms by which intermediate mesoderm is formed from PSCs. Therefore, the transcriptional network responsible for specification of IM, especially in humans, is almost completely uncharacterized.

To achieve efficient directed differentiation to renal lineages that have therapeutic potential, a step-wise differentiation that goes through the

intermediate mesoderm is required. Thus, understanding the developmental program that leads to IM formation is essential for the future development of cell therapies using kidney progenitor cells. In this report we show that LHX1, a transcription factor known to be involved in formation of specialized kidney tissues during development, is actually a master regulator that drives intermediate mesoderm formation from hESCs through binding to distal elements around key IM lineage genes. This is the first report describing LHX1 as a necessary factor for IM differentiation, an earlier role than previously described, elucidating its central role in kidney development and serving as the earliest known regulator of the kidney transcriptional program.

MATERIALS AND METHODS

Cell culture and differentiations

Human embryonic stem cell (hESC) lines WA09 (WiCell) were cultured as previously detailed(Singh et al., 2015). Specifically, hESCs were seeded at 50,000 cells/cm^2 on Geltrex LDEV-Free qualified reduced growth factor basement membrane matrix (Thermo Scientific) or Cultrex coated polystyrene culture plates (Thermo Scientific). Geltrex or Cultrex were used at a 1:200 dilution in DMEM/F-12 w/o glutamine (Corning). To culture hESCs that would used for differentiation to DE, SpM, and PM, a chemically defined base medium (DM) was prepared with DMEM/F-12 w/o glutamine supplemented with 2% Probumin (EMD Milipore), 1x Antibiotic-Antimyotic (Corning), 1x MEM non-essential amino acids (Corning), 1x Trace Elements A (Corning), 1x Trace

Elements B (Corning), 1x Trace Elements C (Corning) 50 ug/mL Ascorbic Acid (Sigma), 10 ug/mL Transferrin (Athens Research and Technology), 0.1 mM 2-mercaptoethanol (Gibco) and 1x glutagro (Corning). Several key growth factors were added for the maintenance of hESCs, specifically, 8 ng/mL human basic-FGF (R&D Systems), 200 ng/mL LONG® R3 human IGF-I (Sigma), 10 ng/mL Activin A (R&D Systems) and 10 ng/mL human Heregulinβ-1 (Peprotech) was added to DM, resulting in a complete defied media (CDM). For cell culture for maintenance of pluripotent cells, hESCs were cultured in a 37 °C incubator containing 5% CO_2, in CDM with medium changes every 24 hours until cells reached 90% confluency after 3 to 4 days. To detatch hESCs from culture plates after reaching confluency, cells were treated with Accutase (ICT) at room temperature (RT) for 5-10 minutes. hESCs were pelleted by centrifugation at 200xG for 4 minutes at RT. Accutase was aspirated and the cell pellet was resuspended in CDM. Cell number was obtained using a hemocytometer and cells were reseeded at a density of 50,000 cells/cm^2 onto geltrex-coated plates as described above. Cells were maintained between passages 30 to 80 and were replaced with cryo-preserved lower passage cells when passage 80 was reached. Mycoplasma detection kit was used to check for mycoplasma contamination regularly. To confirm maintenance of pluripotency and ESC identity, quantitative polymerase chain reaction (qPCR), immunofluorescence, or flow cytometry was used to verify positive expression of OCT4, SOX2, and NANOG.

Definitive endoderm (DE) cells were generated by seeding hESCs at 50,000 cells/cm^2 on geltrex or cultrex-coated plates, as described above in differentiation medium, specifically, DM supplemented with 100 ng/mL Activin A and 8 ng/mL human basic-FGF. 25 ng/mL of human Wnt-3a (R&D Systems) was added to the medium for the first 24 hours only. Medium was changed daily for 4 days. Cells were collected at day 4 and DE cell identity was confirmed using qPCR, immunofluorescence or flow cytometry to verify positive expression of SOX17, FOXA2, GATA6 and GSC and negative expression of OCT3/4, SOX2, and NANOG, as previously described(Singh et al., 2015).

Splanchnic mesoderm (SpM) cells were generated by seeding hESCs at 50,000 cells/cm^2 on geltrex or cultrex-coated plates, as described above in differentiation medium, specifically, CDM supplemented with 25 ng/mL human Wnt-3a and 100ng/mL human BMP-4 (R&D Systems). Medium was changed daily for 4 days. Cells were collected at day 4 and cell identity was confirmed using qPCR, immunofluorescence or flow cytometry to verify positive expression of ISL1, GATA4, NKX2.5 and negative expression of OCT3/4, SOX2, and NANOG, as previously described(Berger et al., 2016).

Intermediate mesoderm (IM) cells were generated by seeding hESCs at 50,000 cells/cm^2 on geltrex or cultrex-coated plates, as described above, in mTeSR1 (STEMCELL Technologies) supplemented with mTeSR1 5x supplement (STEMCELL Technologies) and 1x Antibiotic-Antimyotic. After 24 hours, medium was changed to differentiation medium, specifically, Advanced RPMI (A-RPMI; Invitrogen) supplemented with 1x glutagro, 1x Antibiotic-

Antimyotic, and 5 μM CHIR (R&D Systems) for 24 hours. Medium was then changed to A-RPMI supplemented with 1x glutagro, 1x Antibiotic-Antimyotic, 100ng/mL FGF2, and 1 μM retinoic acid (RA; Thermo Fisher) and was changed daily for 3 days as described in (Lam et al., 2014). Cells were collected at day 4 and cell identity was confirmed using qPCR, immunofluorescence or flow cytometry to verify positive expression of PAX2, PAX8, LHX1, OSR1 and negative expression of OCT3/4, SOX2, and NANOG.

Immunostaining

To stain for protein, cells were cultured on Lab-Tek 4-well chamber slides (Fisher Scientific). After differentiation, medium on slides was removed and cells were washed with DPBS w/o calcium and magnesium and fixed with 4% paraformaldehyde (Electron Microscopy Sciences) in DPBS solution for 15 minutes. Cells were washed 3 times with DPBS and subsequently blocked with blocking buffer containing 10% donkey serum (Equitech-Bio), 0.2 M Triton X-100 (Fisher Scientific) and 0.3 M glycine (Sigma) in DPBS solution for 1 hour at room temperature. Primary antibodies, for antibody information see Supplemental Table 1, were diluted in blocking buffer and incubated overnight at 4 °C. The next day cells were washed 3 times with DPBS and secondary antibodies were added. For secondary antibody information, see Supplemental Table 2. Secondary antibodies were diluted in blocking buffer and incubated for 1 hour at room temperature protected from light. After removing the secondary antibody solution, 1 μg/mL 4′,6-Diamidino-2-phenylindole dihydrochloride (Sigma) diluted

in DPBS was added to each well for 5 minutes. The cells were then washed 3 times in DPBS and coverslips were mounted to the slides with ProLong Gold Antifade (Invitrogen). To take images of the slides, a Leica DM6000B upright microscope and a Zeiss LSM 710 confocal microscope were used at 10X, 20X, and 40X. Images were further processed with Slidebook 6 (Intelligent Imaging Solutions) and videos were made with the same software.

qRT-PCR

To collect cells for qRT-PCR, cells were scraped from the plate using a cell scraper in E.Z. RNA Isolation Kit Lysis Buffer (Omega) and frozen at -80 °C. The E.Z. RNA Isolation Kit (Omega) was used to isolate RNA from cells following the manufacturer's protocol and the RNA was measured and quantified using a Biotek Synergy 2 plate reader. To make cDNA, 1 µg of RNA was used with the Iscript cDNA synthesis kit (Bio-Rad) with the manufacturer's protocol and synthesized with a ViiA 7 Real Time PCR system (Life Technologies). Synthesized cDNA was diluted in molecular grade water to a final volume of 500 µL. To perform the qPCR, a ViiA7 Real-Tie PCR System (Life Technologies) was used for ΔΔCt qRT-PCR analysis in a 384 well plate containing 5 µL TaqMan Universal PCR Master Mix No AmpErase UNG (appliedbiosystems), 0.5 µL TaqMan primer (Life Technologies), 0.5 µL molecular grade water (Fisher Scientific), and 4 µL synthesized cDNA. See Supplemental Table 3 for a list of primers used. Normalization was performed using expression of 18S ribosomal

RNA. Each reaction was done in triplicate and plotted as the mean ± standard deviation.

Western blotting

To prepare cells for western blotting, cells were scrapped from the plate using a cell scrapper in cold DPBS and pelleted by centrifugation at 1,100 rpm for 4 minutes at room temperature and subsequently flash-frozen in liquid nitrogen and stored at -80 °C as pellets. Cell pellets were resuspended and then lysed for 30 minutes in 50uL of lysis buffer made up of RIPA lysis buffer (Sigma), 1X protease inhibitor (Roche), 1X phosphatase inhibitor (Calbiochem) and 100 mM dithithreitol. Cells were then incubated on ice for 30 minutes in lysis buffer and centrifuged at 20,000 x g. Supernatant containing protein lysate was transferred to a new microcentrifuge tube. A Bradford assay was used to quantify isolated protein by measuring absorbance at 595nm in a Biotek Synergy 2 plate reader. Sample buffer, containing 950 µL of Laemmli buffer (Bio-Rad) supplemented with 50 µL of 2-mercaptoethanol, was added to the samples in a volume equal to the protein lysate.

40 µg of protein from each sample was then loaded in Bolt bis-Tris precast gels (Life Technologies) and separated by electrophoresis at 165V for 35 minutes in a Bolt Mini Gel Tank (Life Technologies). Protein was then transferred on to a nitrocellulose membrane using a Bolt transfer system (Life Technologies) at 10V for 60 minutes. Membranes were blocked with 0.5% TBST containing 2% nonfat milk (Bio-Rad) for 60 minutes at room temperature, followed by washes

with TBST while rocking. Primary antibodies, found in Supplemental Table 1, were diluted in blocking buffer and incubated with membranes overnight at 4 °C while rotating. The next day, membranes were washed in TBST while rocking, 3 times for 5 minutes each. HRP conjugated secondary antibodies, Supplemental Table 1, were diluted in blocking buffer and incubated with membranes for 60 minutes at room temperature while rocking, followed by 3 5-minute washes with TBST. Protein levels were detected following incubation with Amersham ECL detection reagent (GE) for 1 minute, using Amersham hyperfilm (GE).

CRISPR-Cas9 Genome Editing

The summary of the CRISPR-Cas9 genome editing strategy is to design and clone guide RNAs (gRNAs) to target a specific genomic locus, transfect Cas9 and gRNAs into cells, drug selection for a polyclonal population of edited cells, clonal isolation and expansion of edited cells, and validation of properly edited cells resulting in a knockout cell line. Each step will be explained in detail as follows.

Guide RNA design and build

To design a strategy for CRISPR-Cas9 knockout, gene loci from the NCBI database (www.ncbi.nlm.nig.gov) to the SnapGene software to determine sites for gRNA binding. The strategy for knockout of the human LHX1 gene was to delete a ~250 base pair segment of the second exon in part of the LIM domain, using a gRNA on each side of the knockout region. To design gRNAs, the

genomic region of the second exon of the human LHX1 gene was uploaded to the Optimized CRISPR Design web-based portal (crispr.mit.edu/), which analyzed the uploaded genomic regions for access to protospacer adjacent motifs (PAMs) and assigned a score based on the frequency of off target binding throughout the genome(Ran et al., 2013).

20 base pair gRNA sequences with off target scores above 90 were selected and purchased as DNA oligos and were cloned into a pSpCas9(BB)-2A-GFP plasmid (Addgene plasmid #48138), which will be referred to as pCas9-GFP, using BbsI overhangs. The composition of the sgRNA oligos contains the sequence CACCGGNNNNNNNNNNNNNNNNNNN, the N's being the gRNA sequence, and CCNNNNNNNNNNNNNNNNNNNCAAA, the N's being the gRNA reverse compliment. The oligos were annealed via the T4 PNK (New England Biolabs) and ligated into the pCas9-GFP vector following BbsI digestion at 16 °C for 60 minutes using ligation mix (Takara). Ligated plasmids were transformed into Mix & Go Competent Cells (Zymo Research) and plated on carbenicillin resistant bacterial plates to isolate clones. Clones were sequenced (Eurofins Scientific) to validate insertion of gRNAs.

Transfection of gRNA-Cas9 plasmids

Once gRNA-pCas9-GFP plasmids were generated and validated they were transfected into WA09 hESCs using Lipofectamine Stem Reagent (Invitrogen) following the protocol for transfection of cells in mTeSR1 medium. To transfect cells in a 100mm dish, 10 µg of DNA of each gRNA-Cas9 plasmid was

diluted in Opti-MEM I Reduced Serum Medium (Thermo Fisher) and mixed with 64 µL of Lipofectamine Stem Reagent diluted in Opti-MEM I and incubated for 15 minutes at room temperature. The transfection mix was then added drop wise to a 100mm dish of hESCs cultured in mTeSR1 medium at ~ 50% confluency. After 24 hours, fresh medium was added on top of the medium containing the transfection mix.

Selection for polyclonal edited cells

Selection started 72 hours after transfection using mTeSR1 medium supplemented with puromycin to select for cells that stably integrated the gRNA constructs into the genome. Under selection conditions, cells were passaged after 4-6 days to maintain proper cell confluency. This process was repeated into progressively smaller cell culture plates until cell death stopped and cells began to amplify. Cells were amplified and cells from a 60mm plate were frozen in liquid nitrogen as a stock of polyclonal edited cells labeled “LHX1-KO”. Genomic DNA was extracted from the LHX1-KO cells and PCR was performed using primers inside and outside of the deleted region of the second exon, as well as primers both outside of the edited region. The polyclonal population showed both bands indicating that some successfully edited clones were present.

Clonal cell line generation

Isolation of clones was accomplished via single cell sorting using a Beckman Coulter MoFlo XDP into each well of a 96-well plate containing Mouse

Embryonic Fibroblasts (MEFS, Gibco) in 200 µL per well of CDM supplemented with 10 µM Y-27632 (ROCK inhibitor, Stem Cell Technologies) and 10% KSR. Medium was changed every other day using a volume of 100 µL per well. Wells were examined under a microscope two weeks after sorting and wells containing hESC colonies were passed via Accutase to a 96-well plate coated with Geltrex. Once confluent, cells were passed to a 24-well plate and cultured in mTeSR1 medium.

Validation of edited cell lines

Clones were collected from 24-well plates and genomic DNA was isolated. PCR was performed using a Phusion High Fidelity PCR kit (New England Biolabs) with primers inside and outside of the edited region and run on a 1% agarose gel. Clones that were successfully edited lacked a band and were further analyzed by Sanger sequencing to confirm successful editing of the LHX1 second exon. qRT-PCR for LHX1 using Taqman primers inside of the second exon also confirmed loss of the genomic region as indicated by a loss of LHX1 trascript. Three validated clones were frozen as stocks and one clone was used for experimentation in this manuscript.

RNA-Seq

For RNA-Seq, total RNA samples were isolated using E.Z.N.A Total RNA kit and DNase treated to remove potential genomic DNA contamination (Omega, E1091). RNA samples with a RIN > 9 were processed for RNA-Seq library

preparation. An average of 30 million paired-end reads with a length of 75 bp were generated per library on a NextSeq platform by Georgia Genomic and Bioinformatics Core. Raw fastq files were mapped to the human genome (hg38) by STAR v2.5.3a using default setting and read counts were obtained in STAR quant-mode. Gene expression analysis was preformed using limma, Glimma and edgeR in R Studio. TMM normalization (edgeR) was performed when conducting across sample comparison and Z score transformed values were used for plotting heatmaps in R package 'pheatmap'. Lineage specific genes were defined by 2-fold difference between selected lineage and all the other lineages. Gene Ontology analysis was performed in Metascape (http://metascape.org/gp/index.html). For principal component analysis, raw read count data matrix is loaded and analyzed by R package 'DESeq2' with its plotPCA function. R package 'ggplot2' is employed to better visualize the result plot. The wild type to LHX1 KO RNA-Seq comparison scatter plot was drawn using R package 'ggplot2' and the cutoff line was set to 2 folds change.

ATAC-Seq preparation

For ATAC-Seq, 50,000 cells in single-cell suspension were prepared as described in a previously published ATAC-Seq protocol (Buenrostro, Jason D., et al. 2015) and treated in 50 μl of Tn5 transposase containing transposition reaction mixture (Nextera DNA Sample Prep Kit, Illumina) and 0.01% Digitonin (Thermo Fisher) at 37 °C for 30 minutes. Genomic DNA was extracted by Zymo

DNA clean&concentrator-5 kit (Zymo Reseaarch) and subjected to preliminary PCR amplification using indexed primers and NEB Next High-Fiedelity 2X PCR master mix. Aliquots of preliminary PCR production were then amplified using quantitative PCR to determine the number of additional PCR cycles needed for one-third of maximal amplification. The additional cycles were applied to the remaining preliminary PCR amplicon accordingly and then purified by Mag-Bind RXNPure plus beads from Omega Bio-tek. The purified final amplicons were sent to Georgia Genomic and Bioinformatics Core for fragment-size distribution analysis using Bioanalyzer to confirm similar fragment-size distribution of each samples and then sequenced in Illumina NextSeq platform.

ATAC-seq bioinformatic analysis

For analysis of ATAC-Seq data, raw fastq files were aligned to UCSC hg38 human genome with Bowtie2 (bowtie2 -p 16 --local -N 1 --phred33 -x /hg38_Bowtie2_Index *.fastq -S ADSC_Rep2_ATAC.sam). PCR duplicates, mitochondria DNA, unmapped reads and Y chromosome DNA were removed by samtools. For visualization of ATAC-Seq data in Integrated Genomics Viewer, mapped bam files were converted to Bigwig format in Deeptools using its bamcoverage function. Peak calling were conducted in MACS2 with parameter '-g hs -B --keep-dup 1 --nomodel --shift -75 --extsize 150 -q 0.01'. The resultant narrow peak files from MACS2 output were merged in HOMER with its mergePeaks function with parameter '-d 100'. Tag density of each merged peak was calculated in HOMER by using makeTagDirectory and annotatePeaks.pl.

The obtained tag counts were normalized to RPKM (read count per kilobase of peak width) in R. For later clustering analysis, distal peaks of merged peak dataset were selected by picking peaks with distance larger than 3Kb to the closest transcription start site. The filtered peaks were then used in K-mean clustering analysis by using scale (package 'scales') and kmean function in R. Peaks of each individual cluster were then used for motif analysis in HOMER using its findMotifsGenome.pl function.

To test the relevance between lineage specific ATAC-Seq peaks and gene expression, each ATAC-Seq peak was assigned to a gene basing on HOMER peak annotation result. The expression of corresponding genes in each lineage was extracted from TMM normalized RNA-Seq log2 transformed CPM (read count per million) matrix of 4 cell types (H9ESC, H9PM, H9IM and H9ISL1). The statistical analysis was performed using one-way ANOVA and visualized in Graphpad Prism 8.

Rescue of LHX1

To rescue IM differentiation in LHX1-KO cells, LHX1 was exogenously and transiently expressed in differentiating LHX1-KO cells. To express LHX1 in cells, LHX1 cDNA (Dharmacon) was cloned into a Cag-GFP vector resulting in a Cag-LHX1-GFP plasmid (LHX1-GFP). LHX1-GFP was transfected into LHX1-KO cells at day 2 of intermediate mesoderm differentiation using Lipofectamine Stem Reagent as described above. At day 3 of differentiation, fresh medium was added on top of medium containing transfection mix. Cells were collected at day

4 of differentiation, pelleted, washed and sorted into GFP- and GFP+ populations on a Becton Dickinson FACSMelody. Both populations were analyzed by qPCR for IM markers.

EdU labeling

For cell cycle analysis of intermediate mesoderm cells, both wildtype and LHX1-KO cells were differentiated to intermediate mesoderm and analyzed using the Click-iT EdU Alexa Fluor 647 Flow Cytometry Assay Kit (Thermo Fisher). Cells were labeled with EdU for 60 minutes prior to collection. After labeling, cells were harvested from plates using Accutase, pelleted and fixed using the fixative included in the kit according to protocol. Cells were then labeled with Click-iT reaction cocktail according to protocol. Cells were stained for DNA content using 1 µL of FxCycle Violet (Thermo Fisher) per 1 million cells and analyzed on a CytoFLEX S Flow Cytometer (Beckman Coulter).

Luciferase assay for *PAX8* enhancer activity

To test the functionality of a putative *PAX8* enhancer element, a genomic region downstream of the *PAX8* gene was chosen based on a unique ATAC peak in IM cells that contains two LHX1 motifs. Two DNA sequences were synthesized, one with the wildtype sequence, and one with the LHX1 motifs mutated at the putative DNA binding motif where 3 base pairs were changed to interrupt binding. These genomic regions were cloned into pGL4.23[***luc2***/minP] vectors as described above. The LHX1-luciferase plasmids were transiently

transfected into differentiating wildtype IM cells at day two as described in "Rescue of LHX1". Cells were also transfected with a Renilla expression plasmid (Promega) at day 2 to normalize for transfection efficiency differences between replicates and samples. Cells were collected at day 4 and assayed for luciferase activity using a Dual-Luciferase Reporter Assay System kit (Promega) and analyzed on a BioTek Synergy 2 plate reader. Luficerase activity was a function of raw luciferase activity divided by Renilla activity.

SIX2-EGFP cell line generation

To generate SIX2-EGFP knock-in reporter ESCs, a TALEN-based homologous recombination strategy was used to insert a 2A-EGFP-PGK-Neo cassette downstream of the stop codon of the endogenous *SIX2* gene according to Li et al 2016(Li et al., 2016). TALEN and donor plasmids were provided by the Zhongwei Li lab at University of Southern California. 10 µg of both TALENs and donor plasmids were transfected into both WA09 ESCs and LHX1-KO ESCs via Lipofectamine Stem Reagent. 4 days following transfection, neomycin selection was used to select for cells that have incorporated the 2A-EGFP-PGK-Neo cassette. Following selection, polyclonal cells were single-cell sorted into 96-well plates using a Beckman Coulter MoFLo XDP. Colonies were passed to a 24 well plate and collected for genomic DNA extraction and genotyping. Clones were genotyped by PCR using one primer in the endogenous *SIX2* locus and another primer inside of the 2A-EGFP-PGK-Neo cassette insert. The presence of a band indicated successful recombination and successful clones were further validated

via Sanger sequencing. Three lines for both wildtype and LHX1-KO cells were frozen and one line for each was used for further studies.

RESULTS

Efficient generation of intermediate mesoderm cells from human PSCs

To differentiate hESCs to intermediate mesoderm (IM) in vitro, we utilized a protocol from Lam et al.(Lam et al., 2014) consisting of 24 hour exposure to Wnt antagonist CHIR99201 followed by 3 additional days of exposure to high concentration FGF2 and low dose retinoic acid (Fig. 2.1A). As cells differentiate toward an intermediate mesoderm cell fate, pluripotency marker expression decreases while characteristic intermediate mesoderm gene expression increased as early as day 3 (Fig. 2.1B). Cells were collected for analysis at day 4, where *PAX2, PAX8*, and *OSR1* expression are highest (Fig. 2.1B). Intermediate mesoderm cells begin to specify in the embryo following the primitive streak and during gastrulation. At this developmental stage, two other major mesoderm cell types exist, specifically the paraxial mesoderm (PM) and the lateral plate or splanchnic mesoderm (SpM). To further characterize the IM generated in culture, we compared the cells to splanchnic mesoderm and paraxial mesoderm differentiated from hESCs in vitro via a panel of genes by qRT-PCR. Signature IM genes *PAX2, PAX8, LHX1* and *OSR1* are highly expressed in the IM while lower expression levels are observed in the SpM and PM (Fig. 2.1C). Pan-mesoderm marker GATA6 expression levels were consistent across mesoderm subtypes while PM marker FOXC1 was highest in

PM, and SpM markers *PITX1, FOXF1* and *NKX2-5* were highest in the SpM (Fig. 2.1C). These results were confirmed via western blotting, where PAX2 and LHX1 protein expression were highest in IM, while the SpM marker ISL1 protein expression was highest in SpM (Fig. 2.1D). The lowers levels of PAX2 and ISL1 protein observed in PM and IM, respectively, likely come from small populations of contaminating cells in culture. OCT4 protein expression was specific to ESCs (Fig. 2.1D). To visualize expression of these key lineage markers, ESCs and IM were analyzed via immunofluorescence. PAX2 and LHX1 expression are completely absent in ESCs while in the IM, ~70-80% of cells express high levels of each (Fig. 2.1F, Supp. Fig. 2.6). OCT4 expression is expressed in nearly 100% of ESCs and 0% of IM (Fig. 2.1F). Finally, principal component analysis of global transcript shows that each cell type is distinct and notably, IM shows significant separation from ESCs, PM, and SpM(Fig. 2.1E).

Discovery of LHX1 as a potential regulator of IM specification

After efficiency and specificity of intermediate mesoderm differentiation was confirmed, RNA-seq and ATAC-seq analyses were performed to investigate the transcriptional program responsible for intermediate mesoderm specification. Cells were differentiated to IM and at day 4, collected for RNA-seq and ATAC-seq (Fig. 2.2A). RNA-seq and ATAC-seq results for IM were compared to our previously generated data from hESCs, PM, and SpM to reveal lineage specific ATAC peaks and transcript (Fig. 2.2A, B). RNA-seq analysis of hESCs and the three mesoderm subtypes reveals global lineage specific transcription of the IM,

which includes key genes *LHX1, PAX2,* and *PAX8* (Fig. 2.2B). GO analysis for each mesoderm-specific gene cluster shows expected, developmentally relevant GO terms. For the IM, these include kidney development, urogenital system development, and metanephros development ranking among the highest by p-value (Fig. 2.2C). GO terms from the PM and SpM cluster include expected terms such as somite development, and heart development for the PM and SpM, respectively (Fig. 2.2C). To discover transcription factors that may be responsible for IM lineage commitment, ATAC-seq peaks were clustered by Z-score which revealed peak clusters specific to each lineage (Fig. 2.2D). Motif discovery analysis was performed for each highlighted peak cluster (Fig. 2.2D). ATAC-seq peaks were then integrated with RNA-seq gene expression data, which showed that gene expression is significantly higher for genes around lineage-specific up-peaks in each respective lineage (Fig. 2.2D). In the IM-specific cluster, homeobox factors are highly ranked by p-value. Specifically, the LIM homeobox transcription factor family LHX was among the highest ranked factors in the IM-specific cluster (Fig. 2.2D). Eight LHX family members have been characterized in the human genome however, only LHX1 is expressed in our IM (Fig. 2.2E). LHX1 is known to be expressed as early as the intermediate mesoderm stage in mice(Fujii et al., 1994) and is required for proper patterning of the kidney field(Shawlot and Behringer, 1995), however its role in determining intermediate mesoderm cell fate is unknown.

LHX1 is required for efficient differentiation to intermediate mesoderm

To investigate the role of LHX1 in intermediate mesoderm formation, we generated an *LHX1* knockout cell line in hESCs using a two-gRNA CRISPR-Cas9 system as described in Vidigal and Ventura et. al.(Vidigal and Ventura, 2015). When LHX1-KO cells were differentiated to intermediate mesoderm, key lineage markers *PAX2, PAX8, LHX1*, and *OSR1* expression are lost while pluripotency gene expression is lost in both wildtype and knockout cells (Fig. 2.3A,C). Furthermore, pan-mesoderm markers *GATA6, GATA5* as well as splanchnic mesoderm, paraxial mesoderm, mesendoderm, endoderm and ectoderm markers are all weakly expressed and relatively unchanged suggesting that the loss of gene expression is specific to IM genes (Fig. 2.3C). At the protein level, PAX2 and LHX1 expression is lost as shown by both immunofluorescence and western blotting (Fig. 2.3B,D). Global transcript analysis of LHX1-KO cells confirms enrichment of key lineage genes in wildtype IM (Supp. Fig. 2.1A). GO analysis of wildtype-enriched genes reveals a global loss of kidney-related gene expression in LHX1-KO IM (Supp. Fig. 2.1A).

To determine if the differentiation defect observed in the LHX1-KO cells is specific to IM, both wildtype and LHX1-KO cells were differentiated to SpM and definitive endoderm (DE). Differentiation efficiency in the knockout cells was comparable to the wildtype cells as nearly every cell in both cell types expresses lineage markers NKX2-5 and ISL1, and SOX17 and FOXA2 in SpM and DE, respectively (Supp. Fig. 2.1A,B). It was further confirmed that the differentiation defect in the LHX1-KO cells is not due to differential kinetics as knockout cells fail

to upregulate key lineage markers by day 5 and day6 of differentiation (Supp. Fig. 2.1C). *OSR1* expression is observed the later stages of differentiation in the LHX1-KO cells, which is likely due to increased proliferation of lateral plate cells in the absence of LHX1. Finally, LHX1-KO cells do not have a proliferation defect as confirmed by cell cycle analysis (Supp. Fig. 2.2). These data indicate that LHX1 is required for proper formation of intermediate mesoderm from human ESCs.

LHX1 is a master regulator of human intermediate mesoderm

To investigate whether LHX1 is sufficient for the intermediate mesoderm transcriptional program, *LHX1* with a GFP tag was exogenously expressed in differentiating LHX1-KO IM cells to see if LHX1 can rescue IM gene expression. qRT-PCR for key lineage markers was performed on GFP- and GFP+ cells. Upon exogenous expression of LHX1, we saw induction of *PAX2, PAX8,* and *WT1* indicating that LHX1 drives expression of the IM transcriptional program (Fig. 2.3E). Exogenous *LHX1* expression does not induce expression of key SpM or PM markers, suggesting that LHX1 specifically rescues transcription of IM genes (Supp. Fig. 2.4). To further show the sufficiency of LHX1 in driving the IM transcriptional program, LHX1-GFP was exogenously expressed in hESCs under self-renewal conditions. qRT-PCR analysis shows that, upon LHX1 expression in hESCs, expression of key lineage genes *PAX2, PAX8,* and *OSR1* is induced while pluripotency genes *OCT4* and *NANOG* are unchanged (Fig. 2.3F). This indicates that LHX1 is a master regulator of IM specification and is sufficient to

drive the IM transcriptional program, although some upstream factor or signaling pathway is likely required for the exit from pluripotency.

It has been shown previously that retinoic acid is required for LHX1 expression, and that there are retinoic acid receptors in the promoter region of the LHX1 gene(Lam et al., 2014; Taira et al., 1994) however, the mechanism by which LHX1 drives expression of the IM transcriptional program is unknown. Analysis of LHX1 motifs in our ATAC-seq data suggests that LHX1 is found at distal elements around key IM lineage genes. This led us to hypothesize that LHX1 binds enhancer elements to drive expression of IM genes. To characterize the binding of LHX1 protein in IM, we exogenously expressed FLAG-tagged LHX1 in differentiating IM cells. RESULTS FOR CHIP EXPERIMENTS HERE.

To test the functionality of LHX1-bound enhancer activity, we designed a luciferase assay to assess the function of LHX1 at a PAX8 distal element. To do this, a PAX8 distal element, selected from a unique ATAC peak containing two LHX1 motifs, was cloned upstream of a luciferase gene with a minimal promoter (Fig. 2.4A). To assess the function of LHX1, a mutant version of this element was also designed by mutating both LHX1 motifs. The wildtype PAX8 element strongly induced luciferase activity while in the LHX1 mutant, luciferase activity was lost. This confirmed that LHX1 binding is required for the function of a putative PAX8 enhancer and serves as a model for how LHX1 drives expression of IM lineage genes.

In this report we have shown that LHX1 is necessary and sufficient for differentiation of hPSCs to intermediate mesoderm. Without the ability to form

intermediate mesoderm, LHX1-KO cells should not be able to form derivatives of the IM such as nephron progenitor cells (NPCs), a progenitor cell population originating from the metanephric mesenchyme in the developing kidney that gives rise to functional cells of the adult nephrons(Davidson et al., 2018). To address this, wildtype and LHX1-KO were differentiated to nephron progenitor cells. Early results indicate that LHX1-KO cells are able to differentiate to NPCs at lower efficiency than wildtype cells, but are unable to contribute to nephron organoids (results not shown).

DISCUSSION

In this report, we show that intermediate mesoderm cells can be generated *in vitro* from hESCs that express key lineage markers as observed in vivo(Lam et al., 2014; Marcotte et al., 2014), and are transcriptionally distinct from the other developmentally relevant mesoderm subtypes, the paraxial and splanchnic mesoderm. GO analysis of specifically expressed genes in each cell type reveals terms expected of its derivatives, providing another layer of evidence to the specificity of each cell type. To our knowledge, this is the first report that describes all three early mesoderm cell types on the molecular level, which adds confidence to the specificity of results and opens the door for studies of cell fate decisions that ESCs make to give rise to one mesoderm subtype over another.

Open chromatin analysis, paired with RNA-seq, and combined with motif discovery analysis, has been used previously for discovery of regulatory factors,

master regulators, and gene regulatory networks controlling complex processes such as tissue specification, tumor development, tissue development, etc. (Baek et al., 2017; Davie et al., 2015; Hosoya et al., 2018; Tripodi et al., 2018). When genome-wide ATAC- and RNA-seq analyses were performed in IM cells, we were able to further characterize the IM by the most highly representative transcription factor motifs. The pluripotency factors OCT4 and NANOG were among the highest ranked in the ESC cluster, while CTCF, an insulator protein with over 13,000 binding sites and is thought to be essential for transcription in many cell types(Kim et al., 2007), ranks among the highest motifs in the common peak cluster. In the mesoderm clusters, FOX motifs rank highly in paraxial mesoderm while GATA4 is highly ranked in splanchnic mesoderm, consistent with previous studies implicating these factors as principal regulators of each respective cell type(Kuo et al., 1997; Loh et al., 2016; Wilm et al., 2004). This provides a unique level of specificity to dissect the early mesoderm gene network. Although this report focuses on the intermediate mesoderm, it should set up for future studies to determine factors necessary for cell fate decisions leading to demarcation of the mesoderm subtypes. Finally, in the intermediate mesoderm the LHX motif was highly ranked, and supplemented by the RNA-seq analysis where LHX1 is the only LHX family member that is expressed. LHX1 was also found in 5 of the top 7 GO terms from the IM-specific gene cluster. This placed LHX1 as the most likely LHX family member regulating IM, and implicates LHX1 at an earlier regulatory point in kidney development than had been previously described.

Knockout of LHX1 in hESCs revealed a deeper role for LHX1 in IM formation. Loss of LHX1 resulted in an inability to make IM, which could be rescued by re-introduction of LHX1 in knockout cells. Furthermore, we found that exogenous expression of LHX1 in hESCs, under self-renewal conditions, drives IM gene expression independent of IM signaling conditions, positioning LHX1 as a master regulator of IM formation. RNA-seq analysis revealed that the loss of LHX1 lead to a loss of kidney-related gene expression providing further evidence that LHX1 controls a global intermediate mesoderm developmental program. Further studies could reveal a larger LHX1 interaction network and point to unknown genes and signaling components involved in early kidney development. Several studies have placed LHX1 downstream of kidney transcription factors such as PAX2/8 and HNF1B(Boualia et al., 2013; Drews et al., 2011). Here, we report that LHX1 acts upstream of these factors and drives the transcriptional program that specifies intermediate mesoderm, providing evidence that LHX1 jumpstarts the kidney developmental program but then acts as a secondary or tertiary factor in specification of downstream kidney tissues.

Several reports describe the binding of LHX1 to enhancer elements to regulate gene expression in various developmental programs(Costello et al., 2015; Lui et al., 2017). Our results are consistent with LHX1 regulating gene expression through binding to distal elements, indicating that LHX1 uses this mode of regulation regardless of developmental program. Early CHIP results indicate that LHX1 in indeed binding to distal elements around key IM genes. Furthermore, LHX1 seems to be regulated by RA signaling. Further experiments

will need to be done to distinguish between RA directly regulating LHX1, and RA regulating a co-factor that in turn regulates LHX1 expression.

Based on preliminary data, LHX1-KO cells are able to differentiate to nephron progenitor cells, however they are not able to participate in tissue formation in nephron organoid formation assays. This may mean that LHX1 is not required for NPC differentiation but is required for differentiation of NPCs to downstream lineages. It could also be reasoned that the SIX2+ cells in the LHX1-KO are not NPCs. Further transcriptional and functional analysis will need to be done to distinguish between these two possibilities.

Generation of downstream kidney cell types *in vitro*, such as nephron progenitor cells, is currently very inefficient(Li et al., 2016). One future goal of this work is to improve differentiation efficiency of the intermediate mesoderm. If *in vitro* differentiations for kidney cells such as NPCs begin with a 100% pure IM progenitor population, it is likely that differentiation efficiency will be improved which opens the door for development of cell therapies using NPCs and other kidney precursor cells. Similar to Lam et al.(Lam et al., 2014), we observe 70-80% PAX2, LHX1 positive cell upon intermediate mesoderm differentiation. Further bioinformatic analysis of LHX1-KO cells should lead to clues as to signaling components and other factors that may improve differentiation efficiency.

Understanding of the regulators of intermediate mesoderm formation has been very limited. This work has presented the first genetic evidence of the requirement of a transcription factor for specification and differentiation of the

intermediate mesoderm from embryonic stem cells, and describes LHX1 as a central regulator of early kidney development.

ACKNOWLEDGEMENTS

We thank Dr. Zhongwei Li (University of Southern California) for providing the TALEN and SIX2-GFP donor plasmids for creating the SIX2-GFP cell lines. We also thank him for conducting the NPC experiments. This work was supported by NIH grants.

Figure 2.1: Efficient generation of intermediate mesoderm from human PSCs. (A) Differentiation protocol for generation of intermediate mesoderm cells from human PSCs. (B) qRT-PCR time course for pluripotency and key lineage markers during IM differentiation. d3 through d5 of differentiation are shown. Expression data plotted as transcript relative to ESCs. (C) Heatmap of qRT-PCR analysis comparing each relevant mesoderm subtype. Gene list represents IM, SpM and PM lineage genes as well as pan-mesoderm genes. Scale is $\log_2$ transformed transcript relative to ESCs. (D) Western blot for key proteins expressed in ESCs and mesoderm subtypes. Cofilin added as a loading control. (E) Immunofluorescence images for ESCs and IM. Merge compiles DAPI, PAX2, and LHX1. DAPI is also shown for OCT4 staining. Scale bars represent 100 μm. Images were taken at 60x on a confocal microscope. (F) PCA plot comparing RNA-seq replicates of ESCs and the mesoderm subtypes. Percent variance for each component is shown on the axes. Yellow - ESC, embryonic stem cells; blue - PM, paraxial mesoderm; red - SpM, splanchnic mesoderm; green - IM, intermediate mesoderm.

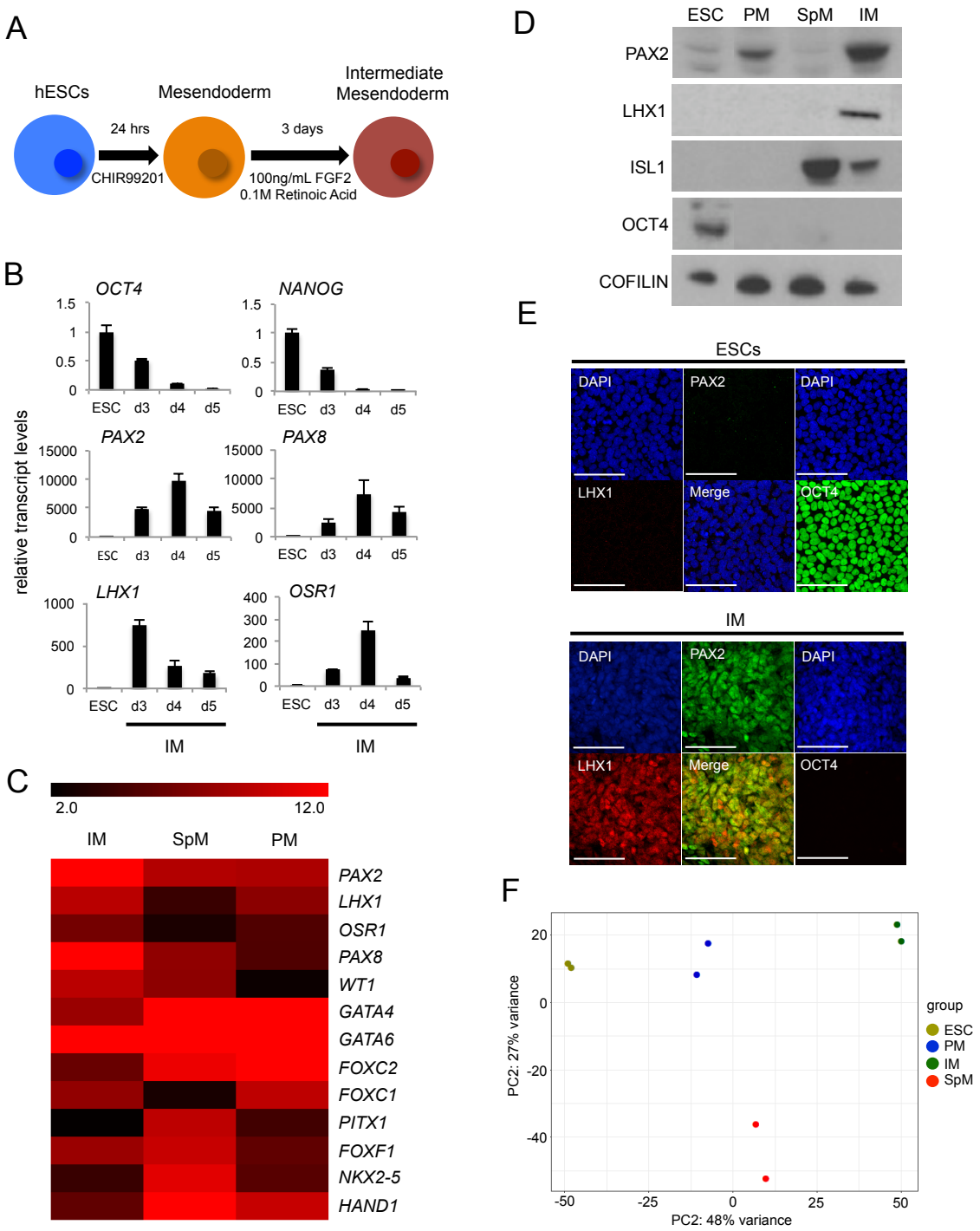

A
hESCs
Mesendoderm
Intermediate Mesendoderm
24 hrs
CHIR99201
3 days
100ng/mL FGF2
0.1M Retinoic Acid
B
relative transcript levels
OCT4
NANOG
PAX2
PAX8
LHX1
OSR1
ESC d3 d4 d5
IM
C
2.0
12.0
IM SpM PM
PAX2
LHX1
OSR1
PAX8
WT1
GATA4
GATA6
FOXC2
FOXC1
PITX1
FOXF1
NKX2-5
HAND1
D
ESC PM SpM IM
PAX2
LHX1
ISL1
OCT4
COFILIN
E
ESCs
IM
DAPI
PAX2
LHX1
Merge
OCT4
F
PC2: 27% variance
PC2: 48% variance
group
ESC
PM
IM
SpM

Figure 2.2: Discovery of LHX1 as a regulator of IM specification. (A) Strategy for RNA- and ATAC-seq of intermediate mesoderm cells. 25M PE and 100M SE represent 25 million paired-end reads and 100 million single-end reads, respectively. (B) Heatmap of RNA-seq data clustered for genes specific to each represented cell type. Colors refer to associated GO analysis plots. Gene expression values were transformed on a Z-score scale. (C) GO analysis of each gene cluster represented by associated colored boxes. Values are plotted as -log of p-value in inverse order of p-value. (D) ATAC-seq heatmap plotted as cluster analysis with each cell-type associated cluster highlighted by associated colored boxes. Two highly ranked motifs are highlighted per cell type cluster via motif discovery analysis. The putative binding motif, transcription factor identity, and associated p-value are shown. Last cluster represents peaks common to all cell types. Values are plotted by Z-score. Far right plots – box plots for ATAC peaks merged with RNA-seq gene expression values showing gene expression of genes with associated unique ATAC up-peaks. Chart titles represent each cluster of cell type-specific peaks. Values plotted as $\log_2$ transformed CPM. (E) Gene expression of LHX family members in IM plotted as FPKM values from RNA-seq data.

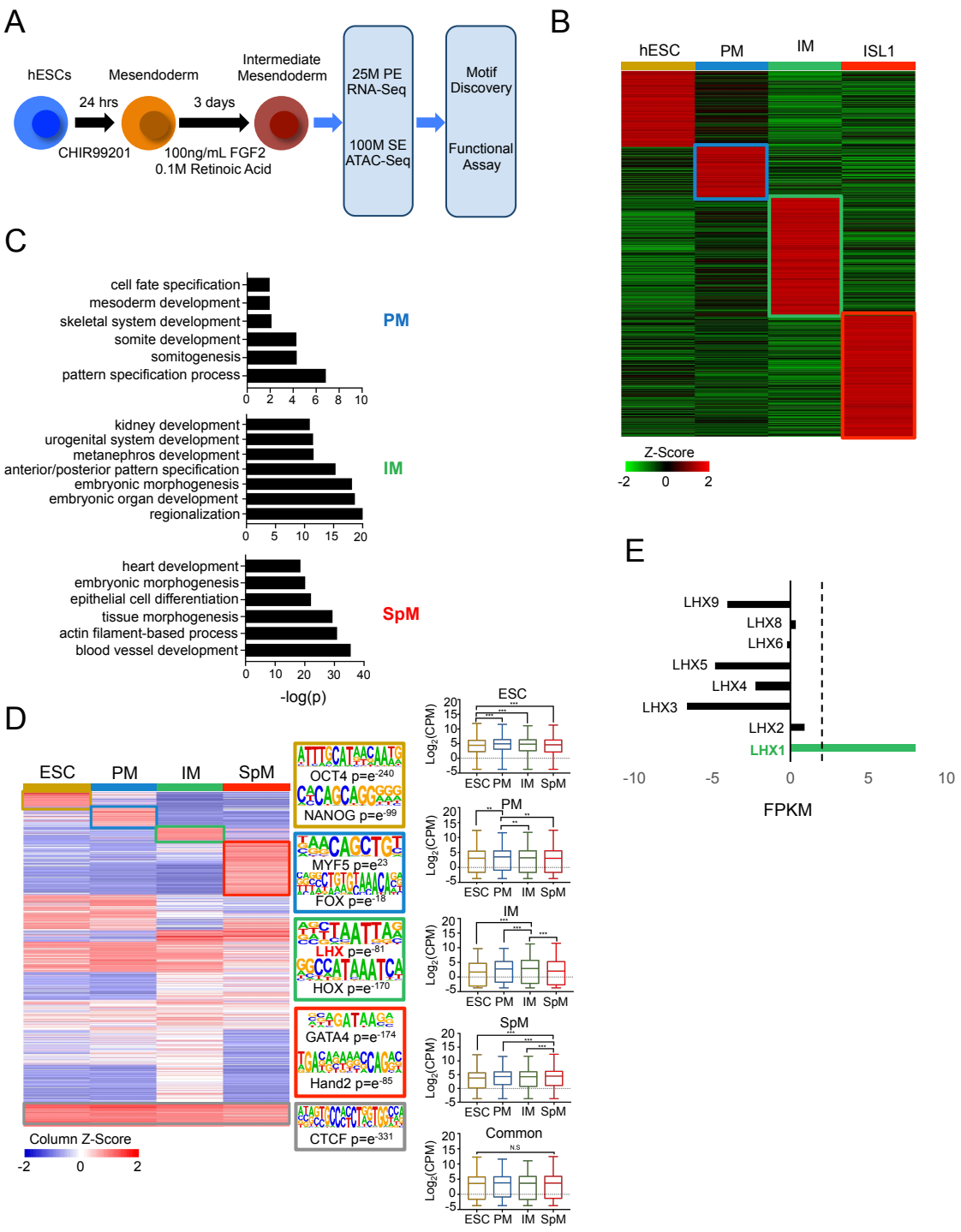

A
hESCs
Mesendoderm
Intermediate Mesendoderm
24 hrs
3 days
CHIR99201
100ng/mL FGF2
0.1M Retinoic Acid
25M PE RNA-Seq
100M SE ATAC-Seq
Motif Discovery
Functional Assay
B
hESC
PM
IM
ISL1
Z-Score
-2 0 2
C
cell fate specification
mesoderm development
skeletal system development
somite development
somitogenesis
pattern specification process
PM
kidney development
urogenital system development
metanephros development
anterior/posterior pattern specification
embryonic morphogenesis
embryonic organ development
regionalization
IM
heart development
embryonic morphogenesis
epithelial cell differentiation
tissue morphogenesis
actin filament-based process
blood vessel development
SpM
-log(p)
D
ESC
PM
IM
SpM
OCT4 p=e-240
NANOG p=e-99
MYF5 p=e23
FOX p=e-18
LHX p=e-81
HOX p=e-170
GATA4 p=e-174
Hand2 p=e-85
CTCF p=e-331
Column Z-Score
-2 0 2
Log2(CPM)
ESC
PM
IM
SpM
Common
N.S
E
LHX9
LHX8
LHX6
LHX5
LHX4
LHX3
LHX2
LHX1
-10 -5 0 5 10
FPKM

Figure 2.3: LHX1 is required for efficient differentiation to intermediate mesoderm. (A) qRT-PCR analysis for pluripotency gene and key IM lineage genes for ESCs, and wildtype IM and LHX1-KO differentiated to IM. Gene expression is relative to ESCs. (B) Heatmap of qRT-PCR data comparing wildtype IM to LHX1-KO IM. Gene list represents pluripotency genes, IM, pan-mesoderm, SpM, PM, mesendoderm, definitive endoderm, and ectoderm genes in that respective order. Scale is $\log_2$ transformed transcript relative to ESCs. (C) Immunofluorescence images for wildtype IM and LHX1-KO IM. Merge compiles PAX2 and LHX1 staining. Scale bars represent 100 μm. Images were taken at 60x on a confocal microscope. (D) Western blot for IM and pluripotency proteins in ESCs, and WT and LHX1-KO IM. CDK2 was used as a loading control. (E) qRT-PCR for ESCs, and FACS sorted GFP- and GFP+ LHX1-KO IM following exogenous LHX1-GFP expression. ESCs, wildtype ESCs; GFP- and GFP+, FACS sorted IM with LHX1-expressing GFP+ cells. Gene expression is plotted relative to ESCs. (F) qRT-PCR analysis for exogenous LHX1 expression in ESCs for pluripotency genes and key IM lineage genes. ESCs, wildtype ESCs; GFP- and GFP+, FACS sorted ESCs with LHX1-expressing GFP+ cells. Values are relative to wildtype ESCs.

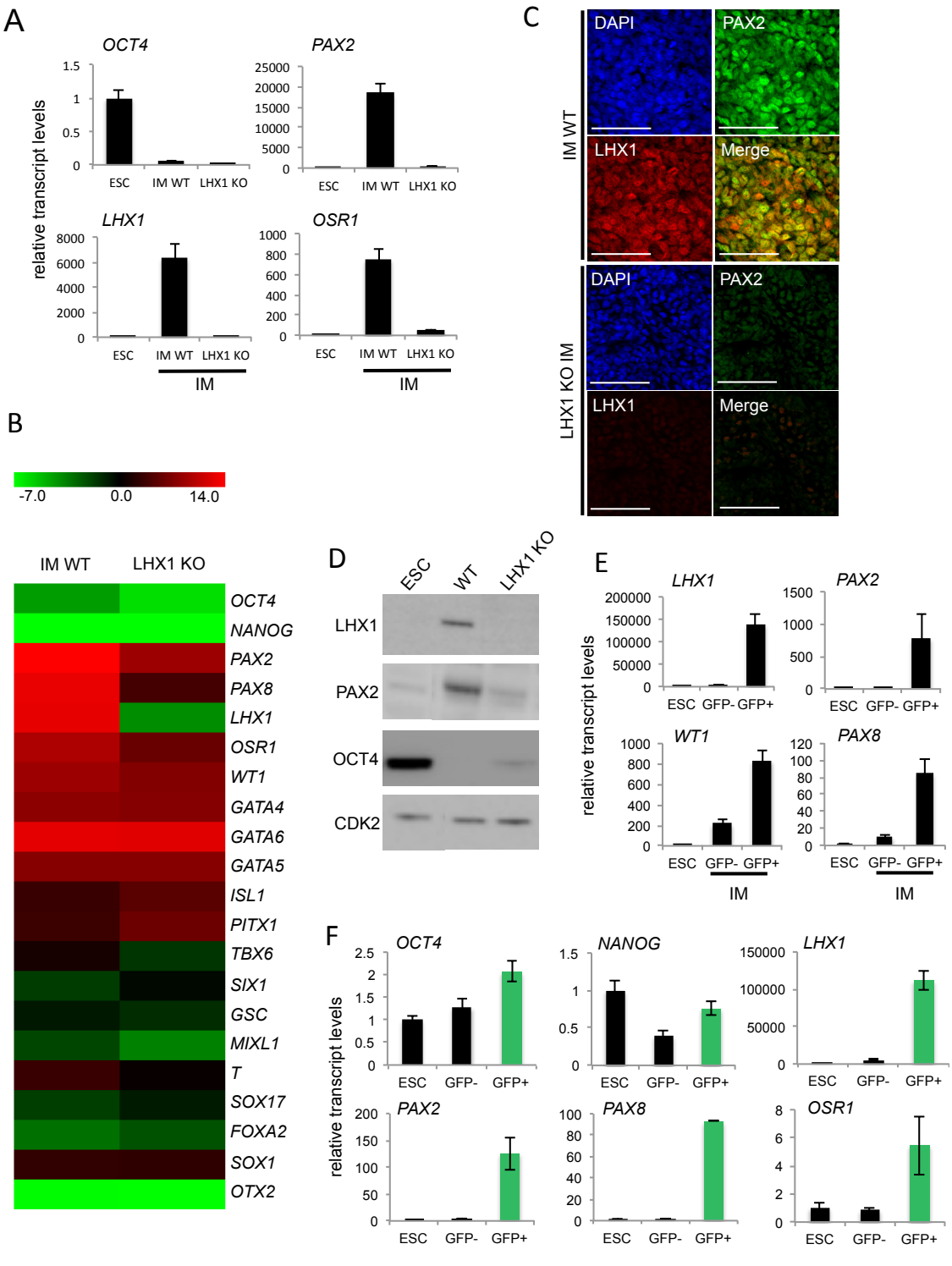
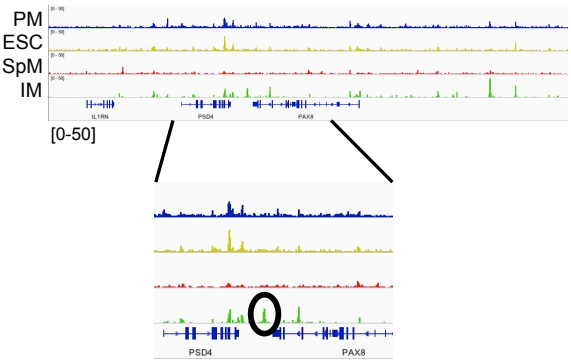
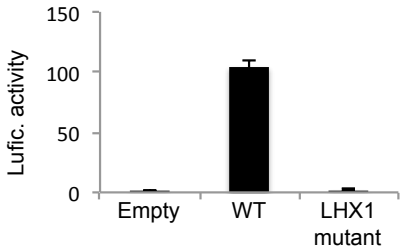

A
OCT4
PAX2
LHX1
OSR1
relative transcript levels
ESC
IM WT
LHX1 KO
IM
B
-7.0
0.0
14.0
IM WT
LHX1 KO
OCT4
NANOG
PAX2
PAX8
LHX1
OSR1
WT1
GATA4
GATA6
GATA5
ISL1
PITX1
TBX6
SIX1
GSC
MIXL1
T
SOX17
FOXA2
SOX1
OTX2
C
DAPI
PAX2
LHX1
Merge
IM WT
LHX1 KO IM
D
ESC
WT
LHX1 KO
LHX1
PAX2
OCT4
CDK2
E
LHX1
PAX2
WT1
PAX8
relative transcript levels
ESC
GFP-
GFP+
IM
F
OCT4
NANOG
LHX1
PAX2
PAX8
OSR1
relative transcript levels
ESC
GFP-
GFP+

Figure 2.4: LHX1 binds functional distal elements around IM genes to control IM gene expression. (A) ATAC-seq peak track around the *PAX8* gene. ATAC peaks are plotted on a scale from 0-50. Bottom panel is an enlargement of the highlighted region. Circled peak highlights the IM-specific *PAX8* peak that contains the DNA used for luciferase experiments. From top to bottom: PM (blue), ESC (yellow), SpM (red), IM (green). (B) Luciferase assay for IM cells transfected with empty luciferase vector, WT *PAX8* peak-containing luciferase vector, and LHX1 mutant *PAX8* peak-containing luciferase vector. Luciferase activity is a measure of the ratio of Luciferase activity to renilla activity.

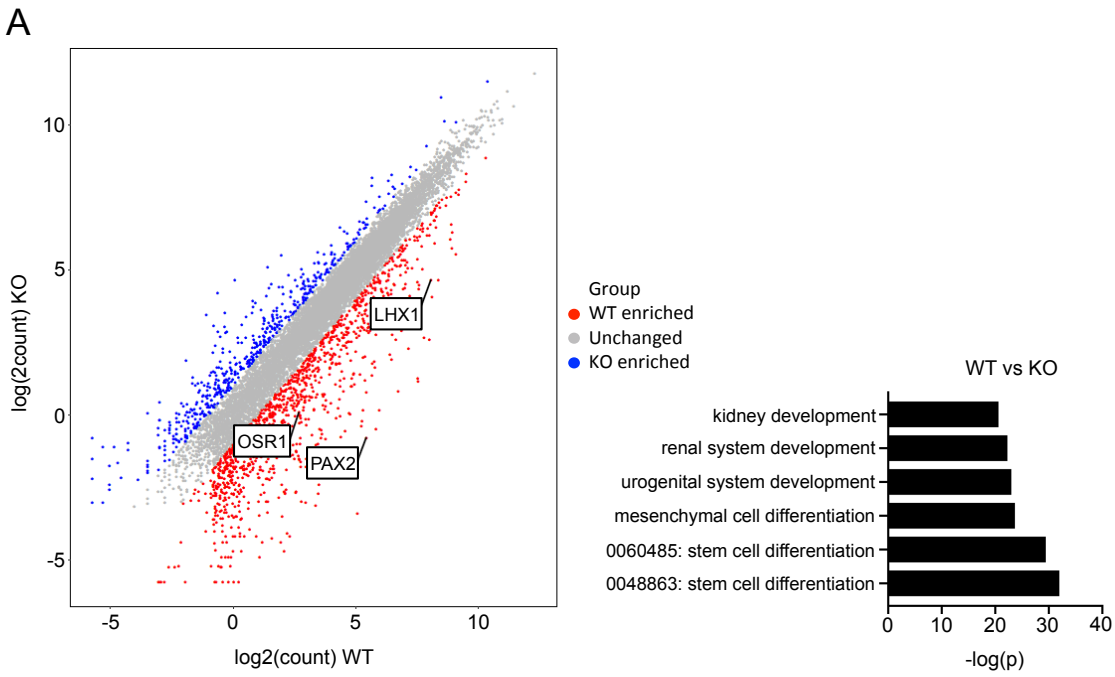
A


PM
ESC
SpM
IM
[0-50]
IL1RN
PSD4
PAX8
PSD4
PAX8

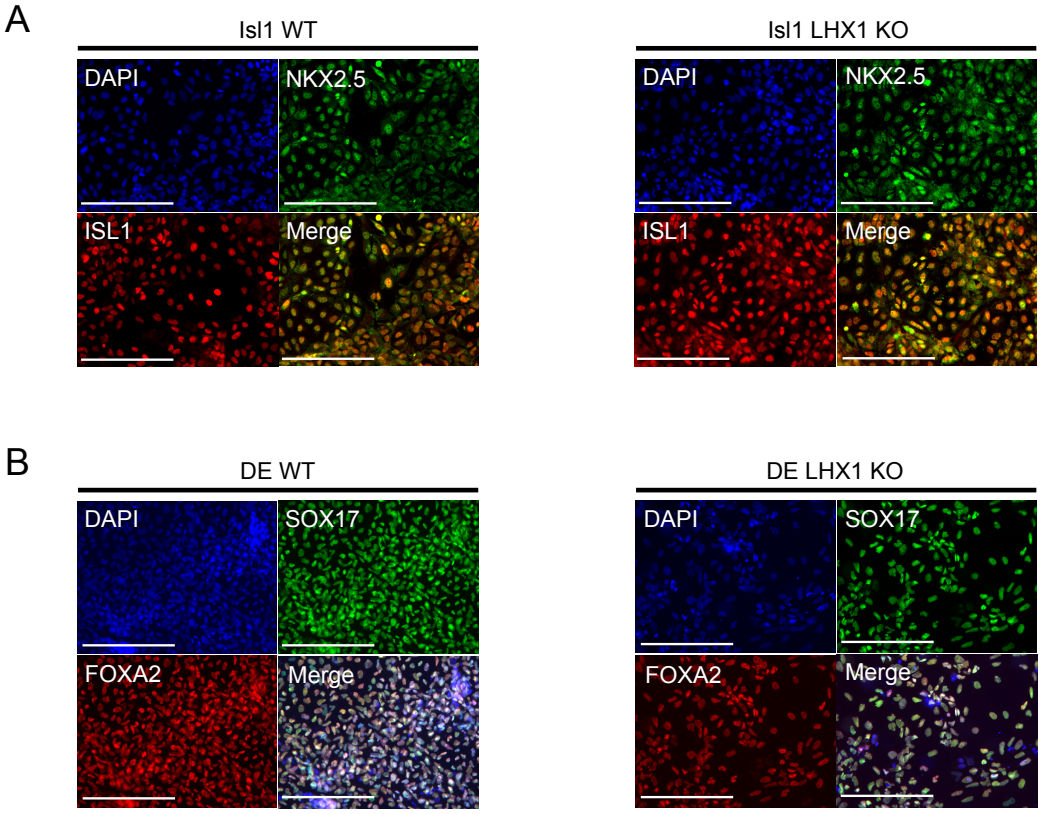
B

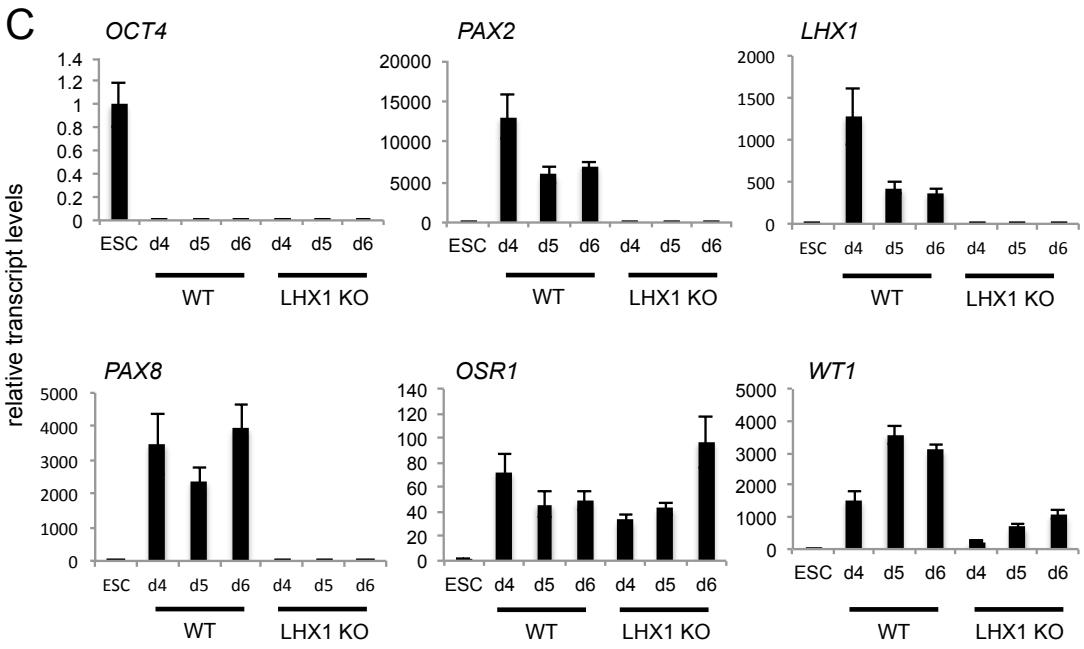

Lufic. activity
150
100
50
0
Empty
WT
LHX1 mutant

Supplementary Figure 2.1: RNA-seq analysis of WT vs. LHX1-KO IM reveals loss of kidney-related gene expression in KO cells. (A) RNA-seq gene expression values transformed on a $\log_2$ scale. Highlighted are WT-enriched genes in red, and LHX1-KO-enriched genes in blue. Cutoff for enriched genes is $\log_2$ value of 2. Grey dots represent genes below the cutoff value. Three key IM genes are highlighted. Right panel – GO analysis performed on WT-enriched genes. GO terms are plotted by -log of p-value.


A
log(2count) KO
log2(count) WT
LHX1
OSR1
PAX2
Group
WT enriched
Unchanged
KO enriched
WT vs KO
kidney development
renal system development
urogenital system development
mesenchymal cell differentiation
0060485: stem cell differentiation
0048863: stem cell differentiation
-log(p)

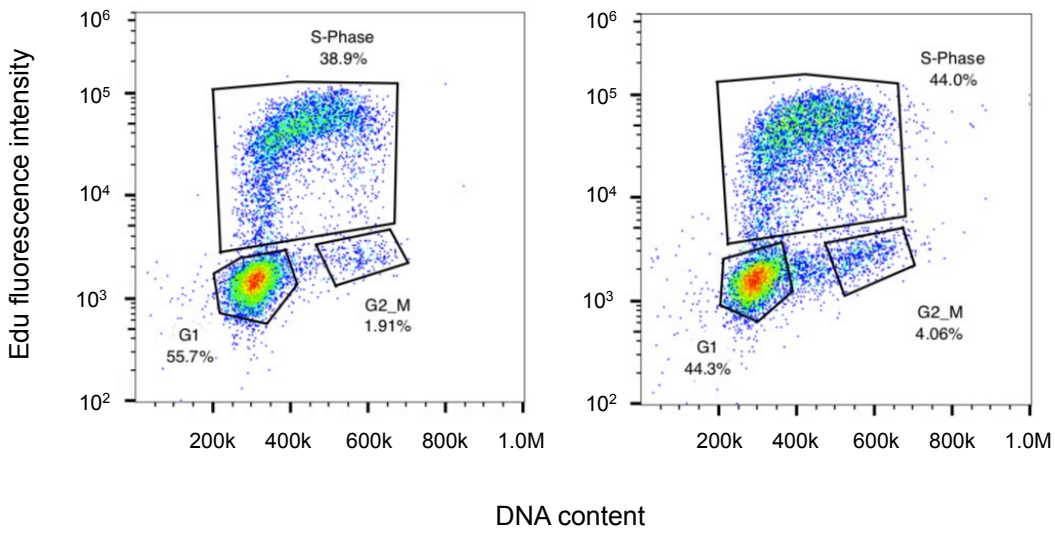
Supplementary Figure 2.2: Differentiation defect in LHX1-KO cells is specific to the intermediate mesoderm. (A) and (B) Immunofluorescence images of both WT and LHX1-KO cells differentiated to SpM and definitive endoderm (DE). Merge compiles NKX2-5 and ISL1, and SOX17 and FOXA2 and DAPI, respectively. Scale bars represent 100 μm. Images were taken at 60x on a confocal microscope. (C) qRT-PCR for WT and LHX1-KO IM at d4-d6 of differentiation. Pluripotency and key IM genes are shown. Gene expression is plotted relative to ESCs.


A
Isl1 WT
DAPI
NKX2.5
ISL1
Merge
Isl1 LHX1 KO
DAPI
NKX2.5
ISL1
Merge
B
DE WT
DAPI
SOX17
FOXA2
Merge
DE LHX1 KO
DAPI
SOX17
FOXA2
Merge

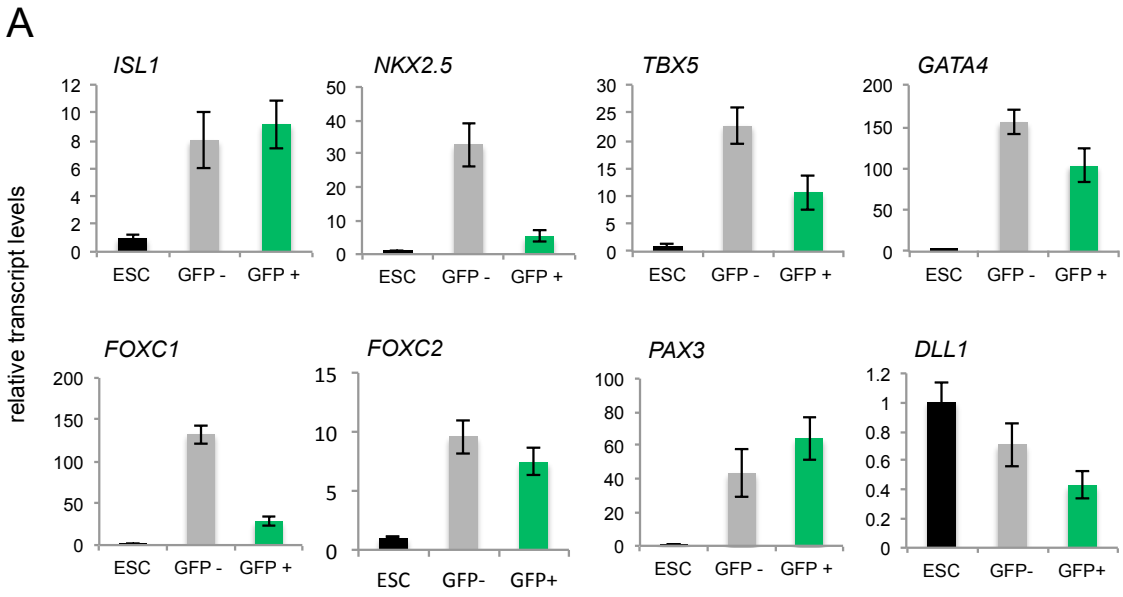

C
relative transcript levels
OCT4
PAX2
LHX1
PAX8
OSR1
WT1
ESC d4 d5 d6 d4 d5 d6
WT
LHX1 KO

Supplementary Figure 2.3: LHX1-KO cells are actively dividing. (A) Edu staining for WT and LHX1-KO IM. Edu fluorescence intensity is plotted on a log scale; DNA content is plotted on a linear scale. Each gate represents its associated cell cycle with percentages representing the percent of total cells in each respective cell cycle.

A

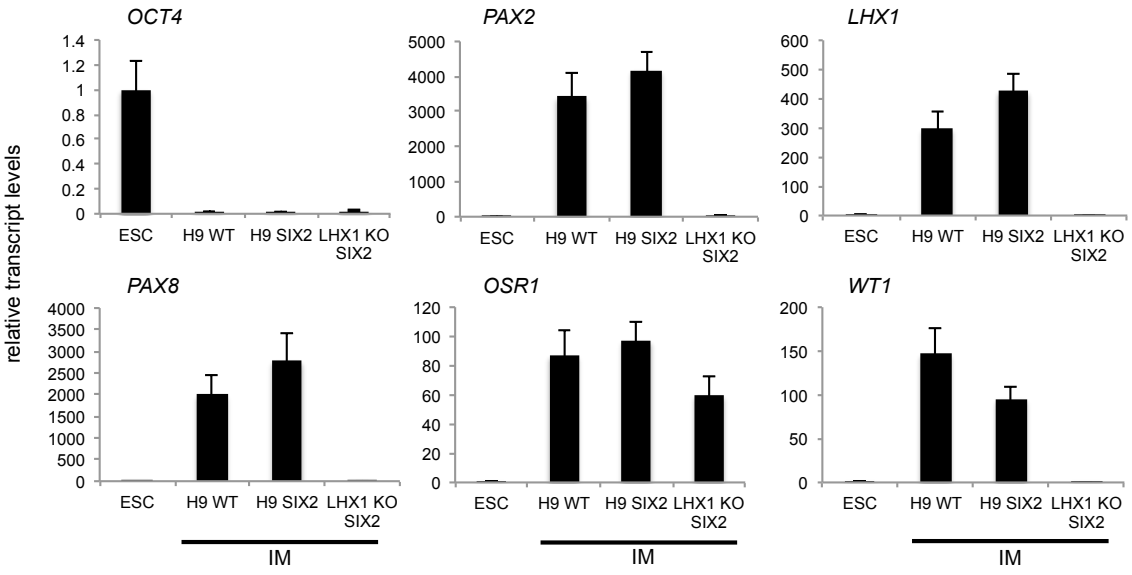

Edu fluorescence intensity
10^6
10^5
10^4
10^3
10^2
S-Phase
38.9%
G1
55.7%
G2_M
1.91%
200k
400k
600k
800k
1.0M
S-Phase
44.0%
G1
44.3%
G2_M
4.06%
DNA content

Supplementary Figure 2.4: LHX1 rescue is specific to intermediate mesoderm genes. (A) qRT-PCR for ESCs, and FACS sorted GFP- and GFP+ LHX1-KO IM cells. Four SpM and four PM genes are shown. Gene expression is plotted relative to ESCs.

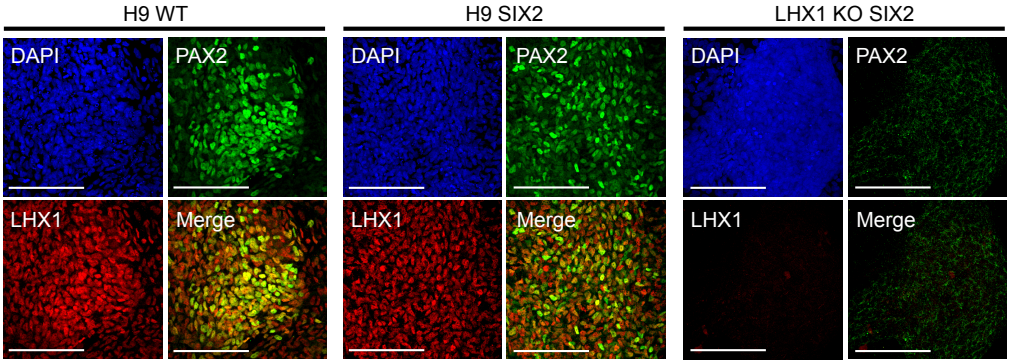

A
relative transcript levels
ISL1
NKX2.5
TBX5
GATA4
FOXC1
FOXC2
PAX3
DLL1
ESC
GFP -
GFP +
GFP-
GFP+

Supplementary Figure 2.5: H9 SIX2-GFP cells differentiate normally to intermediate mesoderm. (A) qRT-PCR for ESCs, and WT H9, WT SIX2, and LHX1-KO SIX2 IM. Pluripotency and key IM genes are shown. Gene expression is plotted relative to ESCs. (B) Immunofluorescence for WT ESC (H9 WT), WT SIX2 (H9 SIX2), and LHX1-KO SIX2 IM. Merge compiles PAX2 and LHX1 staining. Scale bars represent 100 μm. Images were taken at 60x on a confocal microscope.

A


relative transcript levels
OCT4
PAX2
LHX1
PAX8
OSR1
WT1
ESC
H9 WT
H9 SIX2
LHX1 KO SIX2
IM

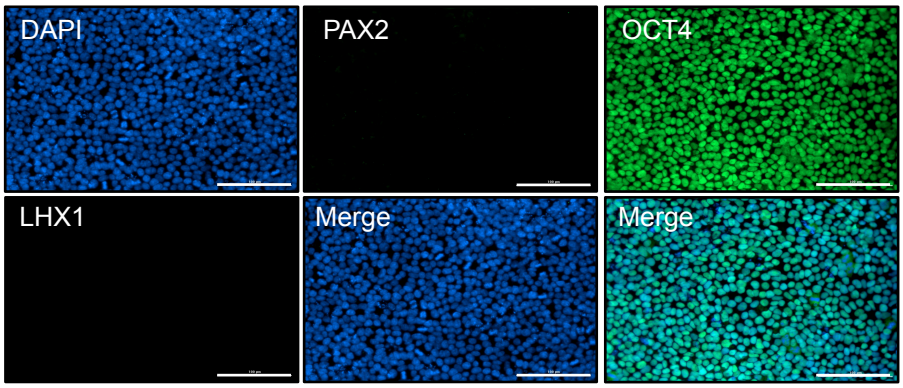
B


H9 WT
H9 SIX2
LHX1 KO SIX2
DAPI
PAX2
LHX1
Merge

Supplementary Figure 2.6: 20x images of WT and LHX1-KO IM. (A) Immunofluorescence for ESCs. (B) Immunofluorescence for WT IM. (C) Immunofluorescence for LHX1-KO IM. Middle panel merge compiles PAX2 and LHX1 staining. Right panel merge compiles DAPI and OCT4 staining. Scale bars represent 100 μm. Images were taken at 20x on a confocal microscope.

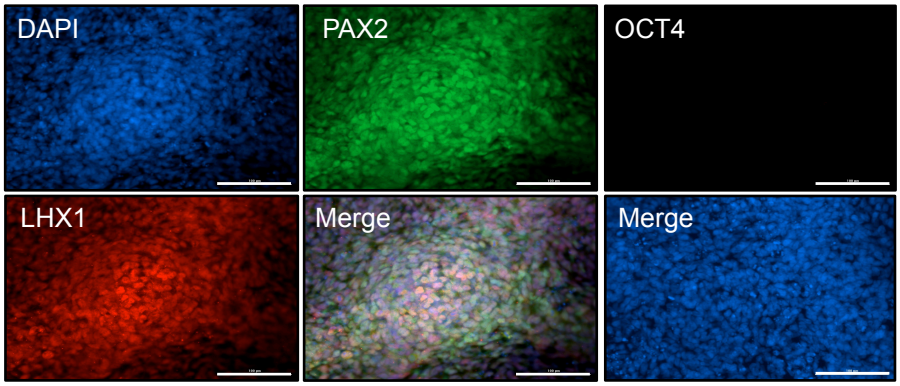
A

ESC



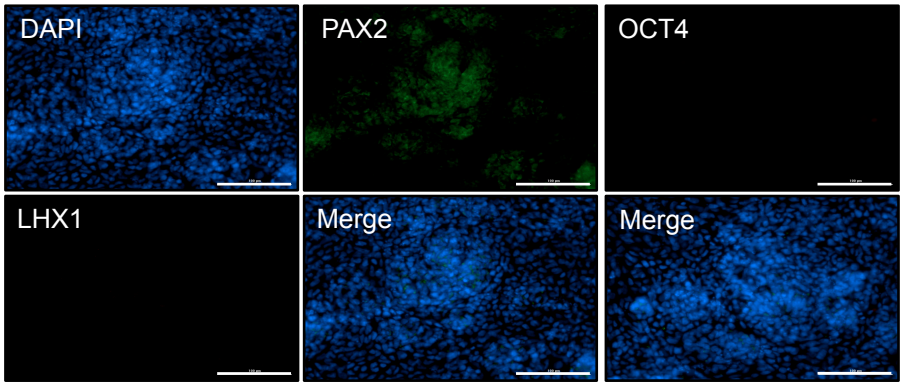
B

IM WT



C

LHX1 KO IM



REFERENCES

Araoka, T., Mae, S.-I., Kurose, Y., Uesugi, M., Ohta, A., Yamanaka, S., Osafune, K., 2014. Efficient and rapid induction of human iPSCs/ESCs into nephrogenic intermediate mesoderm using small molecule-based differentiation methods. PLoS ONE 9, e84881. doi:10.1371/journal.pone.0084881

Baek, S., Goldstein, I., Hager, G.L., 2017. Bivariate Genomic Footprinting Detects Changes in Transcription Factor Activity. Cell Rep 19, 1710–1722. doi:10.1016/j.celrep.2017.05.003

Berger, R.P., Sun, Y.H., Kulik, M., Lee, J.K., Nairn, A.V., Moremen, K.W., Pierce, M., Dalton, S., 2016. ST8SIA4-Dependent Polysialylation is Part of a Developmental Program Required for Germ Layer Formation from Human Pluripotent Stem Cells. Stem Cells 34, 1742–1752. doi:10.1002/stem.2379

Boualia, S.K., Gaitan, Y., Tremblay, M., Sharma, R., Cardin, J., Kania, A., Bouchard, M., 2013. A core transcriptional network composed of Pax2/8, Gata3 and Lim1 regulates key players of pro/mesonephros morphogenesis. Developmental biology 382, 555–566. doi:10.1016/j.ydbio.2013.07.028

Carroll, T.J., Vize, P.D., 1999. Synergism between Pax-8 and lim-1 in embryonic

kidney development. Developmental biology 214, 46–59. doi:10.1006/dbio.1999.9414

Cartry, J., Nichane, M., Ribes, V., Colas, A., Riou, J.-F., Pieler, T., Dollé, P., Bellefroid, E.J., Umbhauer, M., 2006. Retinoic acid signalling is required for specification of pronephric cell fate. Developmental biology 299, 35–51. doi:10.1016/j.ydbio.2006.06.047

Costello, I., Nowotschin, S., Sun, X., Mould, A.W., Hadjantonakis, A.-K., Bikoff, E.K., Robertson, E.J., 2015. Lhx1 functions together with Otx2, Foxa2, and Ldb1 to govern anterior mesendoderm, node, and midline development. Genes & development 29, 2108–2122. doi:10.1101/gad.268979.115

Davidson, A.J., Lewis, P., Przepiorski, A., Sander, V., 2018. Turning mesoderm into kidney. Semin. Cell Dev. Biol. doi:10.1016/j.semcdb.2018.08.016

Davie, K., Jacobs, J., Atkins, M., Potier, D., Christiaens, V., Halder, G., Aerts, S., 2015. Discovery of transcription factors and regulatory regions driving in vivo tumor development by ATAC-seq and FAIRE-seq open chromatin profiling. PLoS Genet. 11, e1004994. doi:10.1371/journal.pgen.1004994

Dawid, I.B., Breen, J.J., Toyama, R., 1998. LIM domains: multiple roles as adapters and functional modifiers in protein interactions. Trends Genet. 14,

156–162.

Drews, C., Senkel, S., Ryffel, G.U., 2011. The nephrogenic potential of the transcription factors osr1, osr2, hnf1b, lhx1 and pax8 assessed in Xenopus animal caps. BMC Dev. Biol. 11, 5. doi:10.1186/1471-213X-11-5

Fujii, T., Pichel, J.G., Taira, M., Toyama, R., Dawid, I.B., Westphal, H., 1994. Expression patterns of the murine LIM class homeobox gene lim1 in the developing brain and excretory system. Dev. Dyn. 199, 73–83. doi:10.1002/aja.1001990108

Hosoya, T., D'Oliveira Albanus, R., Hensley, J., Myers, G., Kyono, Y., Kitzman, J., Parker, S.C.J., Engel, J.D., 2018. Global dynamics of stage-specific transcription factor binding during thymocyte development. Sci Rep 8, 5605. doi:10.1038/s41598-018-23774-9

Kim, T.H., Abdullaev, Z.K., Smith, A.D., Ching, K.A., Loukinov, D.I., Green, R.D., Zhang, M.Q., Lobanenkov, V.V., Ren, B., 2007. Analysis of the vertebrate insulator protein CTCF-binding sites in the human genome. Cell 128, 1231–1245. doi:10.1016/j.cell.2006.12.048

Kobayashi, A., Kwan, K.M., Carroll, T.J., McMahon, A.P., Mendelsohn, C.L., Behringer, R.R., 2005. Distinct and sequential tissue-specific activities of the

LIM-class homeobox gene Lim1 for tubular morphogenesis during kidney development. Development 132, 2809–2823. doi:10.1242/dev.01858

Kumar, N., Richter, J., Cutts, J., Bush, K.T., Trujillo, C., Nigam, S.K., Gaasterland, T., Brafman, D., Willert, K., 2015. Generation of an expandable intermediate mesoderm restricted progenitor cell line from human pluripotent stem cells. Elife 4, a000869. doi:10.7554/eLife.08413

Kuo, C.T., Morrisey, E.E., Anandappa, R., Sigrist, K., Lu, M.M., Parmacek, M.S., Soudais, C., Leiden, J.M., 1997. GATA4 transcription factor is required for ventral morphogenesis and heart tube formation. Genes & development 11, 1048–1060.

Lam, A.Q., Freedman, B.S., Morizane, R., Lerou, P.H., Valerius, M.T., Bonventre, J.V., 2014. Rapid and efficient differentiation of human pluripotent stem cells into intermediate mesoderm that forms tubules expressing kidney proximal tubular markers. J. Am. Soc. Nephrol. 25, 1211–1225. doi:10.1681/ASN.2013080831

Li, Z., Araoka, T., Wu, J., Liao, H.-K., Li, M., Lazo, M., Zhou, B., Sui, Y., Wu, M.-Z., Tamura, I., Xia, Y., Beyret, E., Matsusaka, T., Pastan, I., Rodriguez Esteban, C., Guillen, I., Guillen, P., Campistol, J.M., Izpisua Belmonte, J.C., 2016. 3D Culture Supports Long-Term Expansion of Mouse and Human

Nephrogenic Progenitors. Cell stem cell 19, 516–529. doi:10.1016/j.stem.2016.07.016

Loh, K.M., Chen, A., Koh, P.W., Deng, T.Z., Sinha, R., Tsai, J.M., Barkal, A.A., Shen, K.Y., Jain, R., Morganti, R.M., Shyh-Chang, N., Fernhoff, N.B., George, B.M., Wernig, G., Salomon, R.E.A., Chen, Z., Vogel, H., Epstein, J.A., Kundaje, A., Talbot, W.S., Beachy, P.A., Ang, L.T., Weissman, I.L., 2016. Mapping the Pairwise Choices Leading from Pluripotency to Human Bone, Heart, and Other Mesoderm Cell Types. Cell 166, 451–467. doi:10.1016/j.cell.2016.06.011

Lui, N.C., Tam, W.Y., Gao, C., Huang, J.-D., Wang, C.C., Jiang, L., Yung, W.H., Kwan, K.M., 2017. Lhx1/5 control dendritogenesis and spine morphogenesis of Purkinje cells via regulation of Espin. Nat Commun 8, 15079. doi:10.1038/ncomms15079

Mae, S.-I., Shono, A., Shiota, F., Yasuno, T., Kajiwara, M., Gotoda-Nishimura, N., Arai, S., Sato-Otubo, A., Toyoda, T., Takahashi, K., Nakayama, N., Cowan, C.A., Aoi, T., Ogawa, S., McMahon, A.P., Yamanaka, S., Osafune, K., 2013. Monitoring and robust induction of nephrogenic intermediate mesoderm from human pluripotent stem cells. Nat Commun 4, 1367. doi:10.1038/ncomms2378

Marcotte, M., Sharma, R., Bouchard, M., 2014. Gene regulatory network of renal primordium development. Pediatr. Nephrol. 29, 637–644. doi:10.1007/s00467-013-2635-0

Naylor, R.W., Han, H.I., Hukriede, N.A., Davidson, A.J., 2017. Wnt8a expands the pool of embryonic kidney progenitors in zebrafish. Developmental biology 425, 130–141. doi:10.1016/j.ydbio.2017.03.027

Pedersen, A., Skjong, C., Shawlot, W., 2005. Lim 1 is required for nephric duct extension and ureteric bud morphogenesis. Developmental biology 288, 571–581. doi:10.1016/j.ydbio.2005.09.027

Ran, F.A., Hsu, P.D., Wright, J., Agarwala, V., Scott, D.A., Zhang, F., 2013. Genome engineering using the CRISPR-Cas9 system. Nat Protoc 8, 2281–2308. doi:10.1038/nprot.2013.143

Shawlot, W., Behringer, R.R., 1995. Requirement for Lim1 in head-organizer function. Nature 374, 425–430. doi:10.1038/374425a0

Singh, A.M., Sun, Y., Li, L., Zhang, W., Wu, T., Zhao, S., Qin, Z., Dalton, S., 2015. Cell-Cycle Control of Bivalent Epigenetic Domains Regulates the Exit from Pluripotency. Stem cell reports. doi:10.1016/j.stemcr.2015.07.005

Taira, M., Otani, H., Jamrich, M., Dawid, I.B., 1994. Expression of the LIM class homeobox gene Xlim-1 in pronephros and CNS cell lineages of Xenopus embryos is affected by retinoic acid and exogastrulation. Development 120, 1525–1536.

Tripodi, I.J., Allen, M.A., Dowell, R.D., 2018. Detecting Differential Transcription Factor Activity from ATAC-Seq Data. Molecules 23, 1136. doi:10.3390/molecules23051136

Tsang, T.E., Shawlot, W., Kinder, S.J., Kobayashi, A., Kwan, K.M., Schughart, K., Kania, A., Jessell, T.M., Behringer, R.R., Tam, P.P., 2000. Lim1 activity is required for intermediate mesoderm differentiation in the mouse embryo. Developmental biology 223, 77–90. doi:10.1006/dbio.2000.9733

Ueno, S., Weidinger, G., Osugi, T., Kohn, A.D., Golob, J.L., Pabon, L., Reinecke, H., Moon, R.T., Murry, C.E., 2007. Biphasic role for Wnt/beta-catenin signaling in cardiac specification in zebrafish and embryonic stem cells. Proceedings of the National Academy of Sciences of the United States of America 104, 9685–9690. doi:10.1073/pnas.0702859104

Vidigal, J.A., Ventura, A., 2015. Rapid and efficient one-step generation of paired gRNA CRISPR-Cas9 libraries. Nat Commun 6, 8083. doi:10.1038/ncomms9083

Warga, R.M., Mueller, R.L., Ho, R.K., Kane, D.A., 2013. Zebrafish Tbx16 regulates intermediate mesoderm cell fate by attenuating Fgf activity. Developmental biology 383, 75–89. doi:10.1016/j.ydbio.2013.08.018

Wilm, B., James, R.G., Schultheiss, T.M., Hogan, B.L.M., 2004. The forkhead genes, Foxc1 and Foxc2, regulate paraxial versus intermediate mesoderm cell fate. Developmental biology 271, 176–189. doi:10.1016/j.ydbio.2004.03.034

Wilson, V., Olivera-Martinez, I., Storey, K.G., 2009. Stem cells, signals and vertebrate body axis extension. Development 136, 1591–1604. doi:10.1242/dev.021246

CHAPTER 3

DETAILED EXPERIMENTAL PROCEDURES

MATERIALS AND METHODS

Cell culture and differentiations

Human embryonic stem cell (hESC) lines WA09 (WiCell) were cultured as previously detailed(Singh et al., 2015). Specifically, hESCs were seeded at 50,000 cells/cm^2 on Geltrex LDEV-Free qualified reduced growth factor basement membrane matrix (Thermo Scientific) or Cultrex coated polystyrene culture plates (Thermo Scientific). Geltrex or Cultrex were used at a 1:200 dilution in DMEM/F-12 w/o glutamine (Corning). For culture of hESCs used for differentiation to DE, SpM, and PM, a chemically defined base medium (DM) was prepared with DMEM/F-12 w/o glutamine supplemented with 2% Probumin (EMD Milipore), 1x Antibiotic-Antimyotic (Corning), 1x MEM non-essential amino acids (Corning), 1x Trace Elements A (Corning), 1x Trace Elements B (Corning), 1x Trace Elements C (Corning) 50 ug/mL Ascorbic Acid (Sigma), 10 ug/mL Transferrin (Athens Research and Technology), 0.1 mM 2-mercaptoethanol (Gibco) and 1x glutagro (Corning). To provide factors for the maintenance of hESCs, 8 ng/mL human basic-FGF (R&D Systems), 200 ng/mL LONG® R3 human IGF-I (Sigma), 10 ng/mL Activin A (R&D Systems) and 10 ng/mL human Heregulinβ-1 (Peprotech) was added to DM, resulting in a complete defied media

(CDM). Prior to differentiation, hESCs were cultured ,in a 37 °C incubator containing 5% CO_2, in CDM with medium changes every 24 hours until cells reached 90% confluency after 3 to 4 days. To remove hESCs from culture plates upon reaching confluency, cells were treated with Accutase (ICT) at room temperature (RT) for 5-10 minutes. To pellet hESCs, cells were centrifuged at 200xG for 4 minutes at RT. Accutase was then aspirated and the cell pellet was resuspended in CDM. Cell number was obtained using a hemocytometer and cells were reseeded at a density of 50,000 cells/cm^2 onto geltrex-coated plates as described above. Cells were maintained between passages 30 to 80 and were replaced with lower passage cells when passage 80 was reached. Mycoplasma detection kit was used to check for mycoplasma contamination regularly. To confirm ESC identity, quantitative polymerase chain reaction (qPCR), immunofluorescence, or flow cytometry was used to verify positive expression of OCT3/4, SOX2, and NANOG.

Definitive endoderm (DE) cells were generated by seeding hESCs at 50,000 cells/cm^2 on geltrex or cultrex-coated plates, as described above, and allowing the cells to differentiate in DM supplemented with 100 ng/mL Activin A and 8 ng/mL human basic-FGF, with 25 ng/mL human Wnt-3a (R&D Systems) added to the media for the first 24 hours only. Medium was changed daily for 4 days. Cells were collected at day 4 and cell identity was confirmed using qPCR, immunofluorescence or flow cytometry to verify positive expression of SOX17, FOXA2, GATA6 and GSC and negative expression of OCT3/4, SOX2, and NANOG, as previously described(Singh et al., 2015).

Splanchnic mesoderm (SpM) cells were generated by seeding hESCs at 50,000 cells/cm^2 on geltrex or cultrex-coated plates, as described above, and allowing the cells to differentiate in CDM supplemented with 25 ng/mL human Wnt-3a and 100ng/mL human BMP-4 (R&D Systems). Medium was changed daily for 4 days. Cells were collected at day 4 and cell identity was confirmed using qPCR, immunofluorescence or flow cytometry to verify positive expression of ISL1, GATA4, NKX2.5 and negative expression of OCT3/4, SOX2, and NANOG, as previously described(Berger et al., 2016).

Intermediate mesoderm (IM) cells were generated by seeding hESCs at 50,000 cells/cm^2 on geltrex or cultrex-coated plates, as described above, in mTeSR1 (STEMCELL Technologies) supplemented with mTeSR1 5x supplement (STEMCELL Technologies) and 1x Antibiotic-Antimyotic. After 24 hours, medium was changed to Advanced RPMI (A-RPMI; Invitrogen) supplemented with 1x glutagro, 1x Antibiotic-Antimyotic, and 5 μM CHIR (R&D Systems) for 30 hours. Medium is then changed to A-RPMI supplemented with 1x glutagro, 1x Antibiotic-Antimyotic, 100ng/mL FGF2, and 1 μM retinoic acid (RA; Thermo Fisher) changed daily for 3 days as described in (Lam et al., 2014). Cells were collected at day 4 and cell identity was confirmed using qPCR, immunofluorescence or flow cytometry to verify positive expression of PAX2, PAX8, LHX1, OSR1 and negative expression of OCT3/4, SOX2, and NANOG.

Immunostaining

To stain for protein, medium on slides was removed and cells were washing with DPBS w/o calcium and magnesium and fixed with 4% paraformaldehyde (Electron Microscopy Sciences) in DPBS solution for 15 minutes. Cells were washed 3 times with DPBS and subsequently blocked with blocking buffer containing 10% donkey serum (Equitech-Bio), 0.2 M Triton X-100 (Fisher Scientific) and 0.3 M glycine (Sigma) in DPBS solution for 1 hour at room temperature. Primary antibodies, for antibody information see Table X, were diluted in blocking buffer and incubated overnight at 4 °C. The next day cells were washed 3 times with DPBS and secondary antibodies were added. For secondary antibody information, see Table X. Secondary antibodies were diluted in blocking buffer and incubated for 1 hour at room temperature protected from light. After removing the secondary antibody solution, 1 µg/mL 4′,6-Diamidino-2-phenylindole dihydrochloride (Sigma) diluted in DPBS was added to each well for 5 minutes. The cells were then washed 3 times in DPBS and coverslips were mounted to the slides with ProLong Gold Antifade (Invitrogen). To take images of the slides, a Leica DM6000B upright microscope and a Zeiss LSM 710 confocal microscope were used at 10X, 20X, and 40X. Images were further processed with Slidebook 6 (Intelligent Imaging Solutions) and videos were made with the same software.

qRT-PCR

To collect cells for qRT-PCR, cells were scraped from the plate using a cell scraper in cold DPBS and pelleted by centrifugation at 1,100 rpm for 4 minutes at room temperature. The E.Z. RNA Isolation Kit (Omega) was used to isolate RNA from cells following the manufacturer's protocol and the RNA was measured and quantified using a Biotek Synergy 2 plate reader. To make cDNA, 1 µg of RNA was used with the Iscript cDNA synthesis kit (Bio-Rad) with the manufacturer's protocol and synthesized with a _____ PCR machine. Synthesized cDNA was diluted in molecular grade water to a final volume of 500 µL. To perform the qPCR, a ViiA7 Real-Tie PCR System (Life Technologies) was used for ΔΔCt qRT-PCR analysis in a 384 well plate containing 5 µL TaqMan Universal PCR Master Mix No AmpErase UNG (appliedbiosystems), 0.5 µL TaqMan primer (Life Technologies), 0.5 µL molecular grade water, and 4 µL synthesized cDNA. See Table X for a list of primers used. Normalization was performed using expression of 18S ribosome. Each reaction was done in triplicate and plotted as the mean ± standard deviation.

Western blotting

To prepare cells for western blotting, cells were scrapped from the plate using a cell scrapper in cold DPBS and pelleting by centrifugation at 1,100 rpm for 4 minutes at room temperature and subsequently flash-frozen in liquid nitrogen and stored at -80 °C as pellets. Cell pellets were resuspended and then lysed for 30 minutes in 50uL of lysis buffer made up of RIPA lysis buffer (Sigma), 1X protease

inhibitor (Roche), 1X phosphatase inhibitor (Calbiochem) and 100 mM dithithreitol. Cells were then incubated on ice for 30 minutes in lysis buffer and centrifuged at 20,000 x g. Supernatant containing protein lysate was transferred to a new microcentrifuge tube. A Bradford assay was used to quantify isolated protein by measuring absorbance at 595nm in a Biotek Synergy 2 plate reader. Sample buffer, containing 950 µL of Laemmli buffer (Bio-Rad) supplemented with 50 µL of 2-mercaptoethanol, was added to the samples in volume equal to the protein lysate.

40 µg of protein from each sample was then loaded in Bolt bis-Tris precast gels (Life Technologies) and separated by electrophoresis at 165V for 35 minutes. Protein was then transferred on to a nitrocellulose membrane using a Bolt transfer system (Life Technologies) at 10V for 60 minutes. Membranes were blocked with 0.5% TBST containing 2% nonfat milk (Bio-Rad) for 60 minutes at room temperature, followed by washes with TBST while rocking. Primary antibodies, found in Table X, were diluted in blocking buffer and incubated with membranes overnight at 4 °C while rocking. The next day, membranes were washed in TBST while rocking, 3 times for 5 minutes each. HRP secondary antibodies, Table X, were diluted in blocking buffer and incubated with membranes for 60 minutes at room temperature while rocking, followed by 3 5-minute washes with TBST. Protein levels were detected following incubation with Amersham ECL detection reagent (GE) for 1 minute, using Amersham hyperfilm (GE).

RNA-Seq

For RNA-Seq, total RNA samples were isolated using E.Z.N.A Total RNA kit and DNase treated to remove potential genomic DNA contamination (Omega, E1091). RNA samples with a RIN > 9 were processed for RNA-Seq library preparation. An average of 30 million paired-end reads with a length of 75 bp were generated per library on a NextSeq platform by Georgia Genomic and Bioinformatics Core. Raw fastq files were mapped to the human genome (hg38) by STAR v2.5.3a using default setting and read counts were obtained in STAR quant-mode. Gene expression analysis was preformed using limma, Glimma and edgeR in R Studio. TMM normalization (edgeR) was performed when conducting across sample comparison and Z score transformed values were used for plotting heatmaps in R package 'pheatmap'. Lineage specific genes were defined by 2-fold difference between selected lineage and all the other lineages. Gene Ontology analysis was performed in Metascape (http://metascape.org/gp/index.html). For principal component analysis, raw read count data matrix is loaded and analyzed by R package 'DESeq2' with its plotPCA function. R package 'ggplot2' is employed to better visualize the result plot. The wild type to LHX1 KO RNA-Seq comparison scatter plot was drawn using R package 'ggplot2' and the cutoff line was set to 2 folds change.

ATAC-Seq preparation

For ATAC-Seq, 50,000 cells in single-cell suspension were prepared as described in a previously published ATAC-Seq protocol (Buenrostro, Jason D., et al. 2015) and treated in 50 μl of Tn5 transposase containing transposition reaction mixture (Nextera DNA Sample Prep Kit, Illumina) and 0.01% Digitonin (Thermo Fisher) at 37 °C for 30 minutes. Genomic DNA was extracted by Zymo DNA clean&concentrator-5 kit (Zymo Reseaarch) and subjected to preliminary PCR amplification using indexed primers and NEB Next High-Fiedelity 2X PCR master mix. Aliquots of preliminary PCR production were then amplified using quantitative PCR to determine the number of additional PCR cycles needed for one-third of maximal amplification. The additional cycles were applied to the remaining preliminary PCR amplicon accordingly and then purified by Mag-Bind RXNPure plus beads from Omega Bio-tek. The purified final amplicons were sent to Georgia Genomic and Bioinformatics Core for fragment-size distribution analysis using Bioanalyzer to confirm similar fragment-size distribution of each samples and then sequenced in Illumina NextSeq platform.

ATAC-seq bioinformatic analysis

For analysis of ATAC-Seq data, raw fastq files were aligned to UCSC hg38 human genome with Bowtie2 (bowtie2 -p 16 --local -N 1 --phred33 -x /hg38_Bowtie2_Index *.fastq -S ADSC_Rep2_ATAC.sam). PCR duplicates, mitochondria DNA, unmapped reads and Y chromosome DNA were removed by

samtools. For visualization of ATAC-Seq data in Integrated Genomics Viewer, mapped bam files were converted to Bigwig format in Deeptools using its bamcoverage function. Peak calling were conducted in MACS2 with parameter '-g hs -B --keep-dup 1 --nomodel --shift -75 --extsize 150 -q 0.01'. The resultant narrow peak files from MACS2 output were merged in HOMER with its mergePeaks function with parameter '-d 100'. Tag density of each merged peak was calculated in HOMER by using makeTagDirectory and annotatePeaks.pl. The obtained tag counts were normalized to RPKM (read count per kilobase of peak width) in R. For later clustering analysis, distal peaks of merged peak dataset were selected by picking peaks with distance larger than 3Kb to the closest transcription start site. The filtered peaks were then used in K-mean clustering analysis by using scale (package 'scales') and kmean function in R. Peaks of each individual cluster were then used for motif analysis in HOMER using its findMotifsGenome.pl function.

To test the relevance between lineage specific ATAC-Seq peaks and gene expression, each ATAC-Seq peak was assigned to gene basing on HOMER peak annotation result. The expression of corresponding genes in each lineage was extracted from TMM normalized RNA-Seq log2 transformed CPM (read count per million) matrix of 4 cell types (H9ESC, H9PM, H9IM and H9ISL1). The statistical analysis was performed using one-way ANOVA and visualized in Graphpad Prism 8.

Rescue of LHX1

To rescue IM differentiation in LHX1-KO cells, LHX1 was exogenously and transiently expressed in differentiating LHX1-KO cells. To express LHX1 in cells, LHX1 cDNA (Dharmacon) was cloned into a Cag-GFP vector resulting in a Cag-LHX1-GFP plasmid (LHX1-GFP). LHX1-GFP was transfected into LHX1-KO cells at day 2 of intermediate mesoderm differentiation using Lipofectamine Stem Reagent as described above. At day 3 of differentiation, fresh medium was added on top of medium containing transfection mix. Cells were collected at day 4 of differentiation, pelleted, washed and sorted into GFP- and GFP+ populations on a Becton Dickinson FACSMelody. Both populations were analyzed by qPCR for IM markers (Figure 3.1).

EdU labeling

For cell cycle analysis of intermediate mesoderm cells, both wildtype and LHX1-KO cells were differentiated to intermediate mesoderm and analyzed using the Click-iT EdU Alexa Fluor 647 Flow Cytometry Assay Kit (Thermo Fisher). Cells were labeled with EdU for 60 minutes prior to collection. After labeling, cells were harvested from plates using Accutase, pelleted and fixed using the fixative included in the kit according to protocol. Cells were then labeled with Click-iT reaction cocktail according to protocol. Cells were stained for DNA content using 1 µL of FxCycle Violet (Thermo Fisher) per 1 million cells and analyzed on a CytoFLEX S Flow Cytometer (Beckman Coulter).

SIX2-EGFP cell line generation

To generate SIX2-EGFP knock-in reporter ESCs, a TALEN-based homologous recombination strategy was used to insert a 2A-EGFP-PGK-Neo cassette downstream of the stop codon of the endogenous *SIX2* gene according to Li et al 2016(Li et al., 2016)(Figure 3.2). TALEN and donor plasmids were provided by Zhongwei Li lab at University of Southern California. 10 µg of both TALENs and donor plasmids were transfected into both WA09 ESCs and LHX1-KO ESCs via Lipofectamine Stem Reagent. 4 days following transfection, neomycin selection was used to select for cells that have incorporated the 2A-EGFP-PGK-Neo cassette. Following selection, polyclonal cells were single-cell sorted into 96-well plates using a Beckman Coulter MoFLo XDP. Colonies were passed to a 24 well plate and collected for genomic DNA extraction and genotyping. Clones were genotyped by PCR using a primer in the endogenous *SIX2* locus and inside of the 2A-EGFP-PGK-Neo cassette insert. The presence of a band indicated successful recombination and successful clones were further validated via Sanger sequencing. Three lines for both wildtype and LHX1-KO cells were frozen and one line for each was used for further studies.

Luciferase assay for *PAX8* enhancer

To test the functionality of a putative *PAX8* enhancer element, a genomic region downstream of the *PAX8* gene was chosen based on a unique ATAC peak in IM cells that contains two LHX1 motifs (Figure 3.3). Two DNA sequences were synthesized, one with the wildtype sequence, and one with the LHX1 motifs

mutated at the putative DNA binding motif where 3 base pairs were changed to interrupt binding. These genomic regions were cloned into pGL4.23[***luc2***/minP] vectors as described above. These LHX1-luciferase plasmids were transiently transfected into differentiating wildtype IM cells at day two as described in "Rescue of LHX1". Cells were also transfected with a Renilla expression plasmid (Promega) at day 2 to normalize for transfection efficiency differences between replicates and samples. Cells were collected at day 4 and assayed for luciferase activity using a Dual-Luciferase Reporter Assay System kit (Promega) and analyzed on a BioTek Synergy 2 plate reader. Luficerase activity was a function of raw luciferase activity divided by Renilla activity.

CRISPR-CAS9 GENOME EDITING

The overview of the CRISPR-Cas9 genome editing strategy is to design and clone guide RNAs (gRNAs) to target a specific genomic locus, transfection of Cas9 and gRNAs into cells, drug selection for a polyclonal population of edited cells, clonal isolation and expansion of edited cells, and validation of properly edited cells resulting in a knockout cell line. Each step will be explained in detail as follows.

Guide RNA design and build

To design a strategy for CRISPR-Cas9 knockout, gene loci from the NCBI database (www.ncbi.nlm.nig.gov) to the SnapGene software to determine sites for gRNA binding. The strategy for knockout of the human LHX1 gene was to delete a ~250 base pair segment of the second exon in which part of the LIM domain, using a gRNA on each side of the knockout region. To design gRNAs, the genomic region of the second exon of the human LHX1 gene was uploaded to the Optimized CRISPR Design web-based portal (crispr.mit.edu/), which analyzed the uploaded genomic regions for access to protospacer adjacent motifs (PAMs) and assigned a score based on the frequency of off target binding throughout the genome(Ran et al., 2013).

20 base pair gRNA sequences with off target scores above 90 were selected and purchased as DNA oligos and were cloned into a pSpCas9(BB)-2A-GFP plasmid (Addgene plasmid #48138), which will be referred to as pCas9-GFP, using BbsI overhangs. The form of the sgRNA oligos contains the

sequence CACCGGNNNNNNNNNNNNNNNNNNN, the N's being the gRNA sequence, and CCNNNNNNNNNNNNNNNNNNNNCAAA, the N's being the gRNA reverse compliment. The oligos were annealed via the T4 PNK (New England Biolabs) and ligated into the pCas9-GFP vector following BbsI digestion at 16 °C for 60 minutes using ligation mix (Takara). Ligated plasmids were transformed into Mix & Go Competent Cells (Zymo Research) and plated on carbenicillin resistant bacterial plates to isolate clones. Clones were sequenced (Eurofins Scientific) to validate insertion of gRNAs.

Transfection of gRNA-Cas9 plasmids

Once gRNA-pCas9-GFP plasmids were generated and validated they were transfected into WA09 hESCs using Lipofectamine Stem Reagent (Invitrogen) following the protocol for transfection of cells in mTeSR1 medium. To transfect cells in a 100mm dish, 10 µg of DNA of each gRNA-Cas9 plasmid diluted in Opti-MEM I Reduced Serum Medium (Thermo Fisher) was mixed with 64 µL of Lipofectamine Stem Reagent diluted in Opti-MEM I and incubated for 15 minutes at room temperature. The transfection mix was then added dropwise to a 100mm dish of hESCs cultured in mTeSR1 medium at ~ 50% confluency. After 24 hours, fresh medium was added on top of the medium containing the transfection mix.

Selection for polyclonal edited cells

Selection started 72 hours after transfection using mTeSR1 medium supplemented with puromycin to select for cells that stably integrated the gRNA

constructs into the genome. Under selection conditions, cells were passaged after 4-6 days to maintain proper cell confluency. This process was repeated into progressively smaller cell culture plates until cell death stopped and cells began to amplify. Cells were amplified and cells from a 60mm plate were frozen in liquid nitrogen as a stock of polyclonal edited cells labeled “LHX1-KO”. Genomic DNA was extracted from the LHX1-KO cells and PCR was performed using primers inside and outside of the deleted region of the second exon, as well as a primers both outside of the edited region. The polyclonal population showed both bands indicating that some successfully edited clones were present.

Clonal cell line generation

Isolation of clones was accomplished via single cell sorting using a Beckman Coulter MoFlo XDP into each well of a 96-well plate containing Mouse Embryonic Fibroblasts (MEFS, Gibco) in 200 µL per well of CDM supplemented with 10 µM Y-27632 (ROCK inhibitor, Stem Cell Technologies) and 10% KSR. Medium was changed every other day using a volume of 100 µL per well. Wells were examined under a microscope at 10X two weeks after sorting and wells containing hESC colonies were passed via Accutase to a 96-well plate coated with Geltrex. Once confluent cells were passed to a 24-well plate and cultured in mTeSR1 medium.

Validation of edited cell lines

Clones were collected from 24-well plates and genomic DNA was isolated. PCR was performed using a Phusion High Fidelity PCR kit (New England Biolabs) with primers inside and outside of the edited region and run on a 1% agarose gel.

Clones that were successfully edited lacked a band and were further analyzed by Sanger sequencing to confirm successful editing of the LHX1 second exon. qRT-PCR for LHX1 using Taqman primers inside of the second exon also confirmed loss of the genomic region as indicated by a loss of LHX1 transcript. Three validated clones were frozen in stocks and one clone was used for experimentation in this thesis.

Table 3.1 – Primary antibodies used in these studies

Primary Antibody	Application	Dilution or Concentration	Vendor	Product Number
CDK2	WB	1:2000	Santa Cruz Biotechnology	sc-163
FOXA2	IF	1:100	R&D Systems	AF2400
ISL1	IF	1:100	R&D Systems	AF1837
LHX1	IF, WB	1:200, 1:1000	Santa Cruz Biotechnology	sc-515631
NKX2-5	IF	1:100	R&D Systems	MAB2444
OCT3/4	IF, WB	1:200, 1:1000	Santa Cruz Biotechnology	sc-9081
PAX2	IF, WB	1:200, 1:500	Biolegend	901001
SOX17	IF	1:100	R&D Systems	AF1924

Key: WB – western blot, IF – immunofluorescence, CHIP – chromatin

immunoprecipitation

Table 3.2 – Secondary antibodies used in these studies

Secondary Antibody	Application	Dilution or Concentration	Vendor	Product Number
Alexa Fluor 488 donkey anti-mouse IgG	IF	1:400	Thermo Fisher	A21202
Alexa Fluor 488 donkey anti-rabbit IgG	IF	1:400	Thermo Fisher	A21206
Alexa Fluor 555 donkey anti-mouse IgG	IF	1:400	Thermo Fisher	A31570
Alexa Fluor 555 donkey anti-rabbit IgG	IF	1:400	Thermo Fisher	A31572
Alexa Fluor 647 donkey anti-goat IgG	IF	1:250	Thermo Fisher	A21447
Alexa Fluor 647 donkey anti-mouse IgG	IF	1:250	Thermo Fisher	A31571
Alexa Fluor 647 donkey anti-rabbit IgG	IF	1:250	Thermo Fisher	A31573
Goat Anti-Rabbit Immunoglobulins/HRP	WB	1:2000	Dako	P0448
Rabbit Anti-Goat Immunoglobulins/HRP	WB	1:2000	Dako	P0449
Rabbit Anti-Mouse Immunoglobulins/HRP	WB	1:2000	Dako	P0161

Key: WB – western blot, IF – immunofluorescence, CHIP – chromatin

immunoprecipitation

Table 3.3 – TaqMan primers used in these studies

Gene	Product Number	Gene	Product Number
DLL1	Hs00194509_m1	PAX2	Hs01057416_m1
FOXA2	Hs00232764_m1	PAX3	Hs00240950_m1
FOXC1	Hs00559473_m1	PAX8	Hs00247586_m1
FOXC2	Hs00270951_m1	PITX1	Hs00267528_m1
FOXF1	Hs00230962_m1	OCT4 (POU5F1)	Hs01895061_m1
GATA4	Hs00171403_m1	OSR1	Hs01586544_m1
GATA5	Hs00388359_m1	OTX2	Hs00222238_m1
GATA6	Hs00232018_m1	RNA18S	Hs03928985_g1
GSC	Hs02330376_g1	SIX1	Hs00195590_m1
HAND1	Hs00232764_s1	SOX1	Hs01057642_s1
ISL1	Hs00158126_m1	SOX17	Hs00751752_m1
LHX1	Hs01001500_g1	T	Hs00610080_m1
MIXL1	Hs00430824_m1	TBX5	Hs00361155_m1
NANOG	Hs04399610_m1	TBX6	Hs00365539_m1
NKX2-5	Hs00231763_m1	WT1	Hs01103751_m1

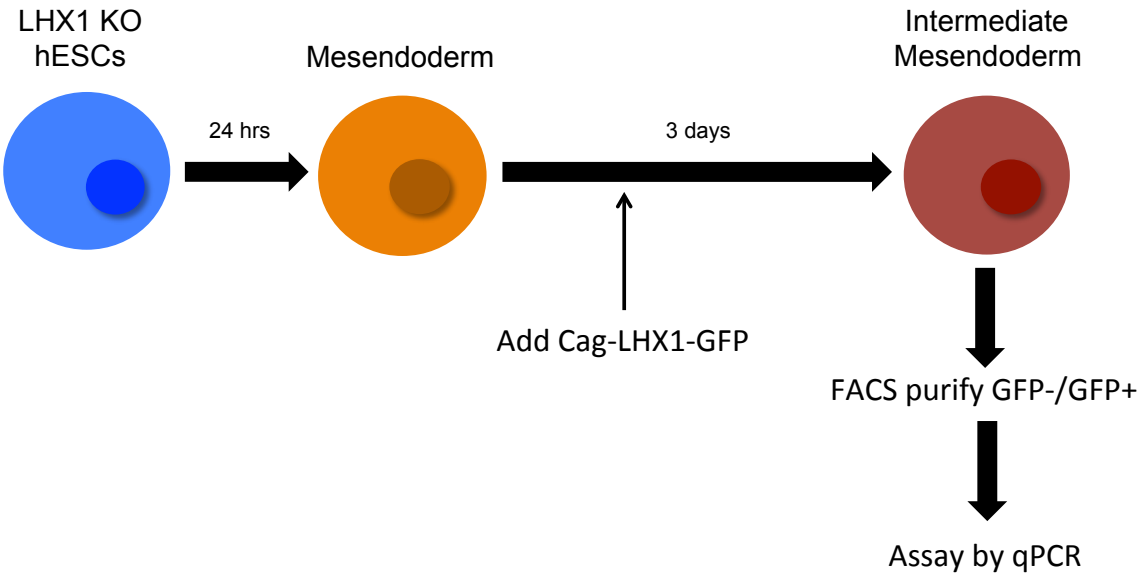


Figure 3.1. LHX1 rescue strategy in differentiating LHX1-KO IM. Differentiating LHX1-KO cells were transfected with a Cag-LHX1-GFP vector at day 2 of differentiation. Cells were collected at day 4, FACS sorted, and analyzed by qPCR.

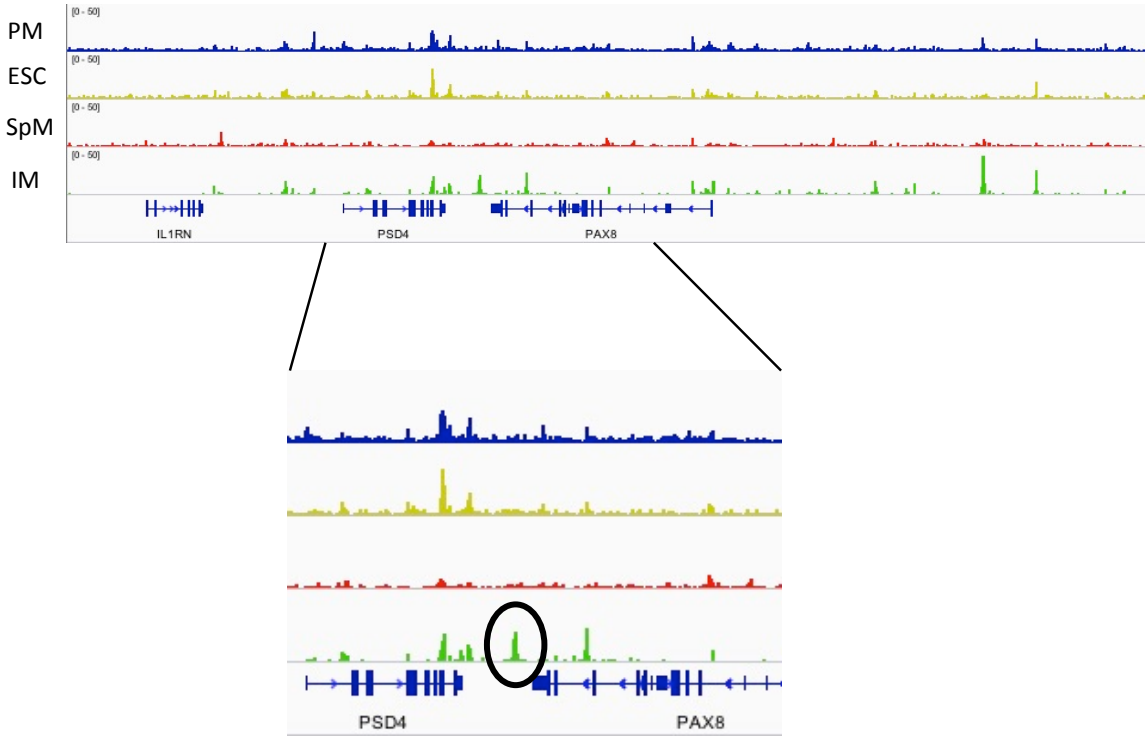


Figure 3.2. Design of luciferase assay for function of LHX1 at *PAX8* enhancer.

ATAC-seq track around the *PAX8* gene shows a peak downstream of the *PAX8* gene in IM that is absent in ESCs, PM, and SpM. This peak contains two LHX1 binding motifs that were mutated for analysis of LHX1 function at this putative enhancer element. The DNA from this peak was cloned into a pGL4.23 luciferase plasmid.

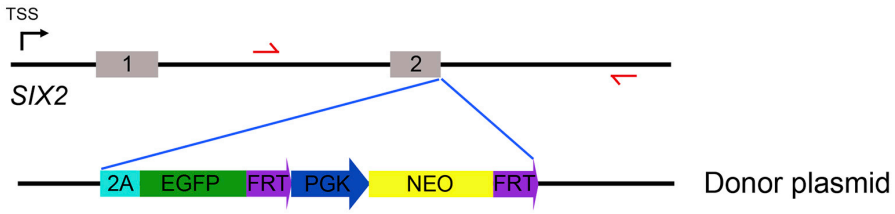


Figure 3.3. Strategy for knocking-in an EGFP cassette at the end of the endogenous *SIX2* gene. A TALEN-based approach was used to knock-in an 2A-EGFP-Neo cassette at the end of the *SIX2* gene in WA09 ESCs and LHX1-KO ESCs. Resulting cell lines are Neo resistant and have GFP fluorescence upon expression of endogenous *SIX2*.

REFERENCES

Berger, R.P., Sun, Y.H., Kulik, M., Lee, J.K., Nairn, A.V., Moremen, K.W., Pierce, M., Dalton, S., 2016. ST8SIA4-Dependent Polysialylation is Part of a Developmental Program Required for Germ Layer Formation from Human Pluripotent Stem Cells. Stem Cells 34, 1742–1752. doi:10.1002/stem.2379

Lam, A.Q., Freedman, B.S., Morizane, R., Lerou, P.H., Valerius, M.T., Bonventre, J.V., 2014. Rapid and efficient differentiation of human pluripotent stem cells into intermediate mesoderm that forms tubules expressing kidney proximal tubular markers. J. Am. Soc. Nephrol. 25, 1211–1225. doi:10.1681/ASN.2013080831

Li, Z., Araoka, T., Wu, J., Liao, H.-K., Li, M., Lazo, M., Zhou, B., Sui, Y., Wu, M.-Z., Tamura, I., Xia, Y., Beyret, E., Matsusaka, T., Pastan, I., Rodriguez Esteban, C., Guillen, I., Guillen, P., Campistol, J.M., Izpisua Belmonte, J.C., 2016. 3D Culture Supports Long-Term Expansion of Mouse and Human Nephrogenic Progenitors. Cell stem cell 19, 516–529. doi:10.1016/j.stem.2016.07.016

Ran, F.A., Hsu, P.D., Wright, J., Agarwala, V., Scott, D.A., Zhang, F., 2013. Genome engineering using the CRISPR-Cas9 system. Nat Protoc 8, 2281–2308. doi:10.1038/nprot.2013.143

Singh, A.M., Sun, Y., Li, L., Zhang, W., Wu, T., Zhao, S., Qin, Z., Dalton, S., 2015. Cell-Cycle Control of Bivalent Epigenetic Domains Regulates the Exit from Pluripotency. Stem cell reports. doi:10.1016/j.stemcr.2015.07.005

CHAPTER 4

DISCUSSION AND CONCLUSIONS

The mammalian kidney arises from the intermediate mesoderm, an early mesoderm subtype located between the paraxial and lateral plate mesoderm in the post-gastrulation embryo(Davidson, 2008). Developmental and molecular regulation of the kidney has been reasonably well characterized, however the very early stages of kidney development, specifically specification of the intermediate mesoderm, have been largely overlooked. LHX1 is a LIM family transcription factor that is required for the proper patterning of the kidney field; specifically, driving the specification of several kidney cell types such as renal progenitors, ureteric bud and nephric vesicle cells(Cirio et al., 2011; Kobayashi et al., 2005; Pedersen et al., 2005; Shawlot and Behringer, 1995). Expression of LHX1 in the intermediate mesoderm has been demonstrated in mouse and human, but the role of LHX1 in intermediate mesoderm specification is unknown.

In our studies we have outlined LHX1 as a master regulator of intermediate mesoderm formation from human embryonic stem cells. LHX1 is required for proper formation of the intermediate mesoderm from human ESCs, as knockout cells are unable to express key lineage markers. When LHX1 is re-introduced into differentiating knockout cells, expression of IM genes is rescued. Importantly when we exogenously express LHX1 in ESCs under self-renewal conditions, IM gene expression is induced. Taken together, this suggests that

LHX1 is central to the IM differentiation program, being both necessary and sufficient to induce the expression of key IM lineage genes. This is the first report that positions LHX1 in a central transcriptional role, further highlighting the importance of this factor in human kidney development. Furthermore, these finding are consistent with the previous report describing a complete loss of the kidney field and reproductive tract in mice(Shawlot and Behringer, 1995). Since the intermediate mesoderm gives rise to all of the cell types of the adult kidney, it is not surprising that the loss of the kidney field observed in LHX1 deficient mice is a result of the inability to form the intermediate mesoderm. Implication of LHX1 in intermediate mesoderm formation is also consistent with the observed loss of the reproductive tract in LHX1 deficient mice(Shawlot and Behringer, 1995). LHX1 has also been shown to be required for the proper development of the reproductive tract(Kobayashi et al., 2004), and our report shows evidence that this phenotype is a result of the inability of LHX1 deficient mice to form the reproductive tract's and kidney's shared and earliest precursor, the intermediate mesoderm.

Open chromatin analysis, paired with RNA-seq, and combined with motif discovery analysis, has been used previously for discovery of regulatory factors, master regulators, and gene regulatory networks controlling complex processes such as tissue specification, tumor development, tissue development, etc. (Baek et al., 2017; Davie et al., 2015; Hosoya et al., 2018; Tripodi et al., 2018). In our reports, this analysis led to a long list of genes that may be involved in regulating IM formation. LHX1 was chosen for further investigation because of its statistical

significance, specificity of expression in the IM lineage, and known expression in IM and kidney cell types. LHX1 however, was not the only interesting gene that was highly ranked in our bioinformatic analysis. Several *HOX* genes, as well as *TBX* genes were also highly ranked. Due to time constraints with the project, we chose to investigate what we felt was the most interesting candidate, LHX1. It would be very interesting to further investigate some of the other highly ranked candidates and study how they regulate IM formation and how/if they interact with LHX1. This analysis could lead to a more complete picture of the hierarchy of transcriptional regulation in IM differentiation. Weissman et al. described cell fate decisions of cardiac and paraxial mesoderm using bioinformatic analysis(Loh et al., 2016), however, our work is the first report describing transcriptional and open chromatin analysis in all three early mesoderm subtypes. We envision that our analysis could serve as a resource to other research groups for investigation of the role of other protein families' role in IM formation. RNA-seq analysis of IM-specific genes could lead to interesting candidates for investigation of proteins other than transcription factors that may be involved in IM specification such as signaling components and chromatin architectural proteins. Furthermore, the data presented here could serve as a resource for studying cell fate decisions in splanchnic or paraxial mesoderm. Many studies describe the formation of downstream cell types but very few have detailed the formation of very early embryonic cell types such as the early mesoderm.

The exact mechanism through which LHX1 regulates the IM transcriptional program is not completely defined, although we have

demonstrated that LHX1 likely regulates *PAX8* expression through binding at a *PAX8* distal element. This mode of regulation is consistent with previous reports describing regulation of developmental gene expression through LHX1 enhancer binding(Costello et al., 2015; Lui et al., 2017). Our results suggest that LHX1 binds to enhancer elements around key IM lineage genes to regulate their expression during IM differentiation. Consistent modes of regulation across development programs bring up the question; which binding partners does LHX1 associate with at enhancer elements? There is likely a protein or complex that LHX1 recognizes at enhancers to bind and regulate expression of associated genes. It will be interesting to further investigate this putative complex that LHX1 forms at enhancers, which should lead to a better understanding of the role of LHX1 at enhancers and could lead to targets that may improve IM differentiations.

Future goals of this work include further dissection of the LHX1 interaction network to gain a more complete understanding of the breadth of the transcriptional network that is controlled by LHX1. This could be accomplished by bioinformatic analysis of LHX1 genome-wide binding; to determine the gene networks that LHX1 drives. Furthermore, a future goal of this work is to use the transcriptome and ATAC-seq data to design improvements to the differentiation protocol, which will result in greater differentiation efficiency. Currently, efficiency of *in vitro* IM differentiation is between 70-80%. Better understanding of the transcriptional regulation, signaling environment, and protein interaction networks will lead to improvements in differentiation efficiency. These improvements in the

earliest progenitor cell population should lead to higher efficiency in downstream kidney differentiations as well, opening the door for more effective disease modeling, organoid generation, and therapeutic applications.

Discovery of regulatory factors involved in specification of cell types using sequencing-based methods or functional methods is an unbiased approach for finding key regulatory components that drive cell fate specification. Further studies using unbiased functional methods such as CRISPR-CAS9 screening(Shalem et al., 2014; Wang et al., 2014; Zhou et al., 2014) to investigate how cells make cell fate decisions can add to the depth of analysis and provide more invaluable resources for the understanding of cell differentiations, which will lead to improvements in efficiency, more precise disease modeling, and therapeutic targets.

REFERENCES

Baek, S., Goldstein, I., Hager, G.L., 2017. Bivariate Genomic Footprinting Detects Changes in Transcription Factor Activity. Cell Rep 19, 1710–1722. doi:10.1016/j.celrep.2017.05.003

Cirio, M.C., Hui, Z., Haldin, C.E., Cosentino, C.C., Stuckenholz, C., Chen, X., Hong, S.-K., Dawid, I.B., Hukriede, N.A., 2011. Lhx1 is required for specification of the renal progenitor cell field. PLoS ONE 6, e18858. doi:10.1371/journal.pone.0018858

Costello, I., Nowotschin, S., Sun, X., Mould, A.W., Hadjantonakis, A.-K., Bikoff, E.K., Robertson, E.J., 2015. Lhx1 functions together with Otx2, Foxa2, and Ldb1 to govern anterior mesendoderm, node, and midline development. Genes & development 29, 2108–2122. doi:10.1101/gad.268979.115

Davidson, A.J., 2008. Mouse kidney development. StemBook. doi:10.3824/stembook.1.34.1

Davie, K., Jacobs, J., Atkins, M., Potier, D., Christiaens, V., Halder, G., Aerts, S., 2015. Discovery of transcription factors and regulatory regions driving in vivo tumor development by ATAC-seq and FAIRE-seq open chromatin profiling. PLoS Genet. 11, e1004994. doi:10.1371/journal.pgen.1004994

Hosoya, T., D'Oliveira Albanus, R., Hensley, J., Myers, G., Kyono, Y., Kitzman, J., Parker, S.C.J., Engel, J.D., 2018. Global dynamics of stage-specific transcription factor binding during thymocyte development. Sci Rep 8, 5605. doi:10.1038/s41598-018-23774-9

Kobayashi, A., Kwan, K.M., Carroll, T.J., McMahon, A.P., Mendelsohn, C.L., Behringer, R.R., 2005. Distinct and sequential tissue-specific activities of the LIM-class homeobox gene Lim1 for tubular morphogenesis during kidney development. Development 132, 2809–2823. doi:10.1242/dev.01858

Kobayashi, A., Shawlot, W., Kania, A., Behringer, R.R., 2004. Requirement of Lim1 for female reproductive tract development. Development 131, 539–549. doi:10.1242/dev.00951

Loh, K.M., Chen, A., Koh, P.W., Deng, T.Z., Sinha, R., Tsai, J.M., Barkal, A.A., Shen, K.Y., Jain, R., Morganti, R.M., Shyh-Chang, N., Fernhoff, N.B., George, B.M., Wernig, G., Salomon, R.E.A., Chen, Z., Vogel, H., Epstein, J.A., Kundaje, A., Talbot, W.S., Beachy, P.A., Ang, L.T., Weissman, I.L., 2016. Mapping the Pairwise Choices Leading from Pluripotency to Human Bone, Heart, and Other Mesoderm Cell Types. Cell 166, 451–467. doi:10.1016/j.cell.2016.06.011

Lui, N.C., Tam, W.Y., Gao, C., Huang, J.-D., Wang, C.C., Jiang, L., Yung, W.H.,

Kwan, K.M., 2017. Lhx1/5 control dendritogenesis and spine morphogenesis of Purkinje cells via regulation of Espin. Nat Commun 8, 15079. doi:10.1038/ncomms15079

Pedersen, A., Skjong, C., Shawlot, W., 2005. Lim 1 is required for nephric duct extension and ureteric bud morphogenesis. Developmental biology 288, 571–581. doi:10.1016/j.ydbio.2005.09.027

Shalem, O., Sanjana, N.E., Hartenian, E., Shi, X., Scott, D.A., Mikkelsen, T.S., Heckl, D., Ebert, B.L., Root, D.E., Doench, J.G., Zhang, F., 2014. Genome-scale CRISPR-Cas9 knockout screening in human cells. Science 343, 84–87. doi:10.1126/science.1247005

Shawlot, W., Behringer, R.R., 1995. Requirement for Lim1 in head-organizer function. Nature 374, 425–430. doi:10.1038/374425a0

Tripodi, I.J., Allen, M.A., Dowell, R.D., 2018. Detecting Differential Transcription Factor Activity from ATAC-Seq Data. Molecules 23, 1136. doi:10.3390/molecules23051136

Wang, T., Wei, J.J., Sabatini, D.M., Lander, E.S., 2014. Genetic screens in human cells using the CRISPR-Cas9 system. Science 343, 80–84. doi:10.1126/science.1246981

Zhou, Y., Zhu, S., Cai, C., Yuan, P., Li, C., Huang, Y., Wei, W., 2014. High-throughput screening of a CRISPR/Cas9 library for functional genomics in human cells. Nature 509, 487–491. doi:10.1038/nature13166