DISSECTION OF TRANSCRIPTIONAL PROGRAMS

IN EMBRYONIC AND METASTATIC STEM CELLS

by

Yilong Zou

A Dissertation

Presented to the Faculty of the Louis V. Gerstner, Jr.

Graduate School of Biomedical Sciences,

Memorial Sloan Kettering Cancer Center

in Partial Fulfillment of the Requirements for the Degree of

Doctor of Philosophy

New York, NY

March, 2016

_____   _____

Joan Massagué, PhD   Date

Dissertation Mentor

DEDICATION

I would like to dedicate this thesis to my wife Xi Wang, and my parents Shuhua Ding and Junxin Wang.

ABSTRACT

Gastrulation and the specification of primary germ layers set the body plan of vertebrate development. A wide range of embryo signals including Fgf, Wnt and the TGF-β family member Nodal govern embryo patterning through transcriptionally regulating the master differentiation factors. However, how these various signals and pathways are coordinated spatio-temporally to guide embryo development is poorly understood. In collaboration with postdoctoral fellow Qiong Wang, PhD, I delineated a regulatory network involving the p53 family and the Wnt pathway, acting together with the TGF-β pathway, to drive mouse mesendoderm specification. This network includes two layers of regulation, first, p53 family members that directly control Wnt3 expression, and second, Wnt activated β-Catenin/Tcf3 and Nodal-activated Smad2/3 that are mutually dependent for binding to, and activation of key mesendoderm identity genes. Knockout of all three family members, p53, p63 and p73, shows that these factors are redundantly required for mesendoderm specification as cells exit pluripotency *in vivo* and in culture. *Wnt3* and its receptor *Fzd1* are among a small set of p53 target genes that are specifically activated in this context. Induction of Wnt signaling by p53 is critical for activation of mesendoderm differentiation genes. Globally, I show that Wnt3-activated Tcf3 and Nodal-activated Smad2/3 transcription factors depend on each other for co-occupancy of

enhancer elements in master differentiation loci. Our results reveal a selective interdependence between signal-activated Tcf and Smad transcription factors. Thus, the p53 family governs a regulatory network that integrates essential Wnt-Tcf and Nodal-Smads inputs for mesendoderm differentiation in the early embryo.

To address how Nodal/Activin signaling achieves target gene specificity, I characterized the requirements of Nodal response co-regulators in target gene selection and activation. I describe that *Smad2X*, a splicing isoform of *Smad2* retaining the exon3, is the most potent inducer of mesendoderm differentiation among all *Smad2/3* protein products. Upon Nodal/Activin stimulation, Smad2X is phosphorylated and recruits key transcriptional co-factors including FoxH1, Smad4 and Trim33 to assemble functional transcriptional machinery. Besides protecting Smad2X from GSK3β/Axin mediated phosphorylation and degradation, the 30 amino acid insert encoded by exon3 suppresses the DNA binding ability of Smad2X MH1 domain. The relaxed DNA binding capacity expanded the ability of Smad2X to act as molecular partner with Smad4 and multiple lineage-specific transcription co-factors. This study suggests a new layer of regulatory mechanisms for controlling the target gene specificity of the TGF-β pathway.

My studies in embryonic development and differentiation highlighted the importance of transcriptomic profiling in understanding cell fate regulation.

To apply genome-wide transcriptome analysis to investigations of cellular behavior in developmental processes and diseases in *in vivo* settings, I characterized and optimized the Translating Ribosome Affinity Purification and Sequencing technology (TRAP-Seq), a method that allows gene expression profiling of defined cell types *in situ*. This dissertation describes three studies utilizing TRAP-Seq to investigate the underlying mechanisms of breast cancer tumorigenesis, melanoma drug resistance, and development of brain metastases. In sum, my studies about the transcriptional programs of embryonic and metastatic stem cells provide important insights for our understanding of normal development, tumorigenesis and metastasis.

BIOGRAPHICAL SKETCH

Yilong was born to Shuhua Ding and Junxin Wang on June 11th, 1987 in Jingzhou, Hu-Bei Province, China. He started his education in primary school at age five in his hometown, the Town of *Zhang-Tian-Shi*. After middle school, Yilong traveled to Wuhan to pursue his high school education in the No.1 Middle School attached to Central China Normal University. In 2005 he started college education in Tsinghua University, Beijing and eventually graduated with a Bachelor in Science (B.S.) degree in Biological Sciences and Biotechnology in 2009 with honors. To complete his research project, Yilong stayed another year in his undergraduate thesis lab, the Laboratory of Cellular Autophagy under the mentorship of Dr. Li Yu. On July 22nd, 2010, Yilong moved to the United States of America to pursue his PhD degree under the mentorship of Dr. Joan Massagué at the Louis V. Gerstner Jr. Graduate School of Biomedical Sciences at Memorial Sloan Kettering Cancer Center, New York.

ACKNOWLEDGEMENTS

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF ABBREVIATIONS

Aurka: Aurora Kinase A

BMP: Bone Morphogenetic Protein

ChIP-Seq: Chromatin Immunoprecipitation and Sequencing

CRISPR: Clustered Regularly-Interspaced Short Palindromic Repeats

DKO: *Trp63*$^{-/-}$;*Trp73*$^{-/-}$ double knockout mouse embryonic stem cells

DTC: Disseminated Tumor Cells

ΔNp63, ΔNp73: dominant negative (N-terminal truncated) isoforms of p63, p73

E(6.5): Embryonic day (6.5)

ES: Embryonic Stem (cells)

EB: Embryoid bodies

EGFP: Enhanced Green Fluorescent Protein

FLAG: FLAG octapeptide (epitope tag)

FoxH1: Forkhead Box H1, or Forkhead Activin Signal Transducer 2 (FAST2)

GSK3β: Glycogen Synthase Kinase-3, beta

HA: Human influenza Hemagglutinin (epitope tag)

ICM: Inner Cell Mass

iPS: induced Pluripotency Stem (cells)

L1CAM: L1 Cell Adhesion Molecule

LIF: Leukemia Inhibitory Factor

MEF: Mouse Embryonic Fibroblasts

MH1: MAD Homology 1

MH2: MAD Homology 2

MSC: Mesenchymal Stromal Cells

NFκB: Nuclear Factor, kappa B

qRT-PCR: quantitative Reverse Transcription Polymerase Chain Reaction

RPL10a: Ribosome Protein, Large Subunit 10A

SBE: Smad Binding (DNA) Element

sgRNA: single guide RNA, a chimeric noncoding RNA acts as a Cas9 partner

Smad2X: protein of Smad family member 2, long isoform with exon3 (mRNA, *Smad2X*)

TAp63, TAp73: Transactivating (full-length) p63, p73

TBE: Tcf Binding (DNA) Element

TGF-β: Transforming Growth Factor, beta

TKO: $Trp53^{-/-}$;$Trp63^{-/-}$;$Trp73^{-/-}$ triple knockout mouse embryonic stem cells

TRAP-Seq: Translating Ribosome Affinity Purification and Sequencing

Trim33: Tripartite motif containing 33, also known as Transcription

Intermediary Factor 1 gamma or TIF1γ.

TSS: Transcription Start Site

WT: Wild Type (genotype)

Chapter 1   Introduction

1.1 Embryo development and lineage specification

1.1.1   Mouse germ layer formation

Embryonic patterning is a critical process in vertebrate development. The major steps of mammalian embryo morphogenesis take place during implantation and gastrulation of a blastocyst. Mouse blastocyst is a spherical structure composed of an outside trophectoderm cell layer and an inner cell mass (ICM) (Arnold and Robertson, 2009; Solnica-Krezel and Sepich, 2012; Tam and Behringer, 1997). Cells in the ICM are the origins of the epiblast and the primitive endoderm. These cells maintain the capacity to differentiate into any tissue in the organism. Around Embryonic day 4 (E4.0), ICM cells become a thin layer of epithelial-like cells surrounding a proamniotic cavity. At E5.5-E6.5, the emergence of the primitive streak-a set of cells formed in the posterior region of the embryo, marks the initiation of gastrulation (Lawson et al., 1991; Tam and Behringer, 1997). During gastrulation, the embryos specify the progenitors for the germ cells and the three primary germ layers-the ectoderm, mesoderm and (definitive) endoderm (Arnold and Robertson, 2009; Tam and Behringer, 1997). Germ layer specification and axis patterning in the early embryo requires precise coordination and regulation by key developmental pathways and lineage-specific transcription factors.

At the end of gastrulation, cells allocated to each germ layer develop into specific organs and tissues. Embryonic ectoderm differentiates to all the neuronal tissues, the epidermis and the tooth enamel (Solnica-Krezel and Sepich, 2012). Embryonic mesoderm develops into all the connective tissue, the bone, spleen, heart, kidney, muscle, vasculature, the hematopoiesis system, the gonad and the genital ducts. Embryonic definitive endoderm gives rise to the thymus, the thyroid, the parathyroids, the respiratory tracks including the bronchia and the lung, the gastrointestinal track including the stomach, liver, pancreas and intestine.

1.1.2   Mesendoderm specification

Mesendoderm is a transient structure derived from a group of epiblast cells undergoing epithelial-mesenchymal transition, invagination and migration towards the anterior part of the embryo. Presence of the mesendoderm lineage has been described from C.*elegens* to mouse development (Rodaway and Patient, 2001). Cells in the mesendoderm are bipotent, and are marked by the co-expression of markers of mesoderm (e.g. T/Brachyury) and endoderm (e.g. Gata5 and Foxa2) (Lemaire et al., 1998; Rodaway et al., 1999). These cells can further differentiate into mesoderm and endoderm as gastrulation proceeds (Lawson et al., 1991; Rodaway and Patient, 2001; Rodaway et al., 1999).

The initiation, maintenance and further differentiation of mesendoderm are controlled by the intensity, location and duration of the Wnt and Nodal pathways (Kimelman and Griffin, 2000; Wardle and Smith, 2006). A series of mesendoderm marker genes including *Eomesodermin* (*Eomes*), *Goosecoid* (*Gsc*), *Mixl1*, *Fgf8* and *Foxa2* have been identified as direct Wnt/Nodal target genes (Estaras et al., 2015; Funa et al., 2015). Loss of these genes individually or in combination severely blocks mesendoderm formation, and results in embryonic lethality (Wang and Chen, 2015).

Eomes is a key member of the T-box family transcription factors. It is first identified as a driver of *Xenopus* mesoderm differentiation (Ryan et al., 1996; Ryan et al., 2000). During mouse development, Eomes is initially expressed radially in the proximal epiblast and rapidly becomes confined to the posterior side of the embryo (Ciruna and Rossant, 1999). Loss of Eomes leads to gastrulation arrest due to defects in mesendoderm induction, axis establishment, and epithelium-mesenchyme transition (Russ et al., 2000). Moreover, Eomes is expressed by the extra-embryonic tissues, driving *Lhx1* expression and anterior visceral endoderm formation and trophoblast development (Russ et al., 2000; Arnold et al., 2008; Nowotschin et al., 2013). Besides being a direct target gene of Nodal/Smads, Eomes serves as a molecular partner of Smad2/3/4 to drive expression of a number of master differentiation regulators (Teo et al., 2011).

Gsc belongs to the bicoid subfamily of the paired homeobox transcription factors. It is a master regulator of *Spemann*'s organizer and mesendoderm specification. In *Xenopus*, *Gsc* knockdown by morpholinos induces severe phenotypes such as head malformation, and *Gsc* overexpression leads to duplicate dorsal structures (De Roberts et al., 1992; Sander et al., 2007). In mouse embryos, *Gsc* is expressed in the primitive streak and is involved in the dorsal-ventral axis establishment of the gastrula. Nonetheless, *Gsc*[-/-] mice were born without major developmental defects (Rivera-Perez et al., 1999; Zhu et al., 1998). The normal early development of *Gsc*[-/-] mice is likely a result of compensation from other Goosecoid family members or other primitive streak induction genes.

T/Brachyury is a member of the T-box family transcription factors and a master regulator of mesoderm formation. T/Brachyury is initially expressed in the nascent mesoderm and the adjacent primitive ectoderm cells, and then in the notochord, but not in any adult tissue (Wilkinson et al., 1990; Yanagisawa et al., 1981). Consistently, *T/Brachyury*[-/-] mouse embryos did not develop mesoderm, and exhibited severe defects in notochord formation (Wilkinson et al., 1990; Yanagisawa et al., 1981).

Foxa2 (Forkhead box A2, or hepatocyte nuclear factor 3b (Hnf3b)) is a member of the winged-helix-winged transcription factor family. Its expression is initiated in embryonic mesendoderm and subsequent definitive endoderm,

and sustains in adult endoderm-derived tissues (Weinstein et al., 1994). FoxA2 mutant mouse embryos were arrested around E10-E11 due to notochord formation defects, deficient somite development, and disorganized neural tube (Weinstein et al., 1994). In addition, the gut tube formation was blocked though a few endodermal cells still exist in disorganized manner in *Foxa2*[-/-] embryos (Weinstein et al., 1994). These observations suggest that FoxA2 is a critical player in mesoderm and endoderm formation.

Mixl1 (Mix paired-like homeobox 1) is a member of the Mix/Bix homeobox gene family. Mixl1 expression is first detected in the visceral endoderm of mouse gastrula and then expands to the primitive streak and mesoderm (Hart et al., 2002). Mixl1 mutant mouse embryos were arrested at E8.5-E9.5, the early somite stage due to failure in primitive streak induction, aberrant headfold formation, deficient paraxial mesoderm development and defects in the heart tube and gut tube formation (Hart et al., 2002). In addition, *Mixl1*[-/-] embryonic stem cells failed to contribute to the gut tube in chimeras, supporting the pivotal role of Mixl1 in endoderm gut tube formation.

Fgf8 (Fibroblast growth factor 8; or Androgen-induced growth factor, 'Aigf') is a member of the fibroblast growth factor (FGF) family. During mouse embryo development, Fgf8 is first expressed in the primitive streak at E7.5 and later in the brain, limb and facial tissues (Ohuchi et al., 1994). The presence of Fgf8 in the primitive streak is accompanied by Fgf4 expression, another FGF family

member that shares redundant functions in driving cell migration during primitive streak formation. Loss of Fgf8/Fgf4 together results in gastrulation failure, and blocks mesoderm and endoderm formation (Sun et al., 1999).

1.1.3   Pluripotency and differentiation in embryonic stem cells

Derivation of embryonic stem cells

Our current understanding about the molecular mechanisms of pluripotency and differentiation has been significantly advanced by the isolation of pluripotent embryonic stem (ES) cells from mouse and human embryos (Evans and Kaufman, 1981; Martin, 1981; Thomson et al., 1998). Mouse embryonic stem (mES) cells and human ES (hES) cells undergo unlimited self-renewal and maintain their developmental potential after months of *in vitro* culture. Both mES cells and hES cells exhibit full developmental potential to differentiate into progenitors of all three primary germ layers. Pluripotent stem cells provide convenient and adaptable opportunities for investigating developmental processes, cell therapy and drug responses (De Los Angeles et al., 2015). More recently, reprogramming for induced pluripotency from somatic cells further empowered regenerative medicine (Takahashi et al., 2007; Takahashi and Yamanaka, 2006). To improve the safety and efficiency in utilizing pluripotent stem cells for pre-clinical and clinical applications, a thorough understanding about the molecular mechanisms controlling pluripotency and differentiation is critical.

Maintenance of pluripotency

Pluripotency in ES cells and iPS cells is actively maintained by an interactive, self-sustaining molecular network composed of the Oct4 (Pou5f1, POU domain class 5 transcription factor 1), Sox2 (SRY box 2) and Nanog (the 'OSN' complex) triad and other co-factors such as Sall4 (Sal-like 4), Tcf3 and Esrrb1 (Dunn et al., 2014; Young, 2011; De Los Angeles et al., 2015). This transcriptional circuit activates the expression of their own and of each other, introduces repressive histone modifications (H3K9me3 and H3K27me3) and blocks the transcription of differentiation-driver genes (Young, 2011; Ruthenburg et al., 2007). The molecular pathways required to activate this transcriptional network have been largely delineated over the last decade (Young, 2011; Dunn et al., 2014; De Los Angeles et al., 2015).

As the prototype of pluripotent stem cells, mouse ES cells represents the "naive" state of pluripotency and can be sustained in the presence of leukemia inhibitory factor (LIF), an interleukin (IL)-6 cytokine family member, and bone morphogenetic proteins (BMPs) (Williams et al., 1988; Ying et al., 2003). For mouse ES cells cultured *in vitro*, LIF can be supplied by mouse embryonic fibroblasts (MEF) or through exogenous additions. LIF supports pluripotency through activation of Jak-Stat3 and PI3K/Akt signaling (Niwa et al., 1998; Niwa et al., 2009; Watanabe et al., 2006). Activated Jak-Stat3 pathway stimulates *Klf4* expression, which further triggers *Sox2* transcription (Niwa et al., 2009). In

parallel, active PI3K/Akt pathway drives *Nanog* expression (Niwa et al., 2009). Both Nanog and Sox2 are required for the transcriptional activation of Oct4. Therefore LIF signaling reinforces   the pluripotency transcriptional network.

In the presence of active LIF/JAK/Stat3 signaling, BMP-Smad1/5/8 signaling induces Inhibitor of differentiation 1 (Id1) to block neuroectoderm differentiation, the default differentiation path in the absence of the pluripotency network (Ying et al., 2003). In the absence of LIF, BMP signaling cooperates with Fgf4/Erk1/2 cascade, and induces ectodermal and trophectodermal differentiation in ES cells (Kunath et al., 2007).

In contrast to the "naive" pluripotent state of mouse ES cells, hES cells and mouse epiblast stem cells (mEpiSCs) represent a "primed" pluripotency state that is ready to accept differentiation cues (Davidson et al., 2015; Nichols and Smith, 2009; Buecker et al., 2014; Weinberger et al., 2016). The differences in these pluripotent states are reflected by the distinct molecular pathways that are required for pluripotency maintenance: basic fibroblast growth factor (bFGF), Noggin, BMP4, Activin A and Wnt3a are required to support the maintenance of human ES cell pluripotency while LIF and BMP are dispensable (Beattie et al., 2005; Chen et al., 2014; Dalton, 2013; Noggle et al., 2005).

Induction of differentiation

The maintenance of pluripotency is mainly supported by the effective blockade of differentiation paths, and the transition from pluripotency to differentiation is a fundamental process during embryogenesis. The onset of differentiation is accompanied by the down-regulation of pluripotency factors such as Nanog (Chambers et al., 2003), and the first differentiation events *in vivo* are initiated with the emergence of primitive ectoderm, followed by the formation of primitive streak. Induced differentiation of ES cells *in vitro* largely recapitulates the molecular events in embryo development. ES cells can be induced to differentiate in two approaches, the first is through the generation of embryoid bodies (EBs) (Burkert et al., 1991), which resembles the environment that is present in embryos, the second is to steer ES cells to differentiate towards specific lineages using a specific sequence of cytokines and growth factors. EB differentiation is efficient in inducing multiple lineages that exist in early embryos, however the complexity of cell types dampens the power of this system for molecular studies. On the other hand, cells differentiated through guided steps do not necessarily recapitulate the cellular transitions in embryos. Using a combination of these model systems, a series of signaling pathways has been implicated in guiding pluripotent stem cells to differentiate towards specific lineages.

1.2 The p53 family in development and differentiation

1.2.1 The p53 family

The p53 family transcription factors function as central protectors of genome integrity in adult tissue progenitors and as rheostats for maintaining tissue homeostasis. As the best characterized family member, p53 has been reported to play crucial roles in contexts of stress responses and tumor suppression. In the presence of DNA damage, p53 protein is stabilized and activated. When activated, p53 mediates DNA repair responses, cell cycle arrest, senescence and apoptosis (Stiewe, 2007). It maintains the genome integrity of multicellular organisms by triggering apoptosis in cells that exhibit damages in the genome. Lack of p53 confers resistance to DNA-damaging agents induced apoptosis. In normal cells, Mdm2, an E3 ubiquitin ligase, tightly controls the stability and activity of p53 protein (Momand et al., 1992). Phosphorylation and acetylation are also directly involved in the regulations of p53 functions (Dai and Gu, 2010). Due to its active nature in triggering cytostasis or apoptosis, p53 expression is barely detectable in somatic tissues and is only induced by insults such as DNA damages, and moreover, it is the most prominently mutated gene in human cancers. In the absence of stress, the function of p53 is linked to regulation of development and differentiation.

p63 and p73 are two ancient ancestry of p53. p63 and p73 share high structural and functional similarities with p53 and bind to the same consensus DNA sequence. p53, p63 and p73 share similar secondary structures with a transactivation (TA) domain, a proline-rich (PR) sequence, a highly conserved central DNA-binding domain (DBD) and an oligomerization (OD) domain. The TA domain is involved in the recruitment of transcriptional activators. Each DBD domain recognizes half unit of a consensus symmetric RRRCWWGYYY DNA sequence (Pavletich et al., 1993). The OD is involved in protein tetramer formation (Sturzbecher et al., 1992). p63 and p73 contain two extra domains, a sterile α motif (SAM) and a transcription inhibition domain (TID). SAM mediates protein-protein interactions, whereas TID can interact with the TA domain and represses its transcriptional activity. Transcripts of the p53 family members undergo extensive alternative splicing. Dominant negative isoforms (ΔN isoforms, N-terminal truncated) can be produced by alternative transcription start site skipping the TA domain, or alternative splicing that eliminates the TA region at the N-terminus. Both approaches generate proteins that retain the DNA binding ability but lack transactivation activity. The p53 family share similar functions. For instances, TAp73 and p53 play redundant functions in maintaining genome integrity by suppressing polyploidy and aneuploidy (Talos et al., 2007), and TAp73 and TAp63 are both required for p53-mediated apoptosis (Flores et al., 2002).

11

1.2.2   p53 in development and differentiation

p53 is present in invertebrates and it is highly conserved throughout vertebrate evolution, though its functions in development and differentiation remain controversial. In mouse embryos, p53 is highly expressed in pluripotent cells from the ICM and the early progenitors of germ layer formation until E10.5 (Bazzi and Anderson, 2014; Danilova et al., 2008; Lutzker and Levine, 1996; Schmid et al., 1991). Its robust expression in embryonic tissues is in direct contrast to its stringently controlled expression in somatic cells.

p53 null mice are viable and fertile (Donehower et al., 1992; Jacks et al., 1994; Purdie et al., 1994; Tsukada et al., 1993), albeit they are prone to develop spontaneous and induced tumors. Somatic cells from $Trp53^{-/-}$ mice exhibit deficient apoptotic responses under DNA damage insults (Donehower et al., 1992). Intriguingly, $Trp53^{-/-}$ ES cells could not induce expression of mesendoderm markers in monolayer differentiation protocols while pluripotency associated genes were down-regulated normally. Nevertheless, $Trp53^{-/-}$ cells in EB differentiation conditions still expressed major mesendoderm markers and contributed to various tissues in chimeric embryos (Shigeta et al., 2013). Moreover, p53 null zebrafish develop normally (Berghmans et al., 2005; Langheinrich et al., 2002). In contrast to the normal early development of p53 mutant mice and zebrafish, p53 knockdown in *Xenopus* led to severe gastrulation defects (Cordenonsi et al., 2003; Sasai et al., 2008; Takebayashi-Suzuki et al.,

2003). What causes the striking species-specific differences in vertebrates is not fully understood.

Studies from pluripotent stem cells have greatly assisted our understanding on the role of p53 in stemness and differentiation. Exogenous expression of p53 stimulates ES cell differentiation (Almog and Rotter, 1997; Dai and Gu, 2010). Embryonic stem cells are refractory to stress-induced p53 mediated signals (Qin et al., 2007; Sabapathy et al., 1997). More recently, it was uncovered that in pluripotent ES cells p53 is inactivated by phosphorylation of Ser212 and Ser312 by Aurora kinase A (Aurka), whereas LIF removal leads to Aurka down-regulation and p53 activation (Katayama et al., 2004; Lee et al., 2012; Liu et al., 2004), a process required for mesoderm gene induction.

The molecular basis for p53 regulated lineage specification is poorly delineated. In mouse ES cells, DNA damage caused by ultra-violet light or doxorubicin treatment induces up-regulation of p53 protein levels within hours (Lin et al., 2005). DNA damage-activated p53 recruits transcriptional repressor mSin3a, which binds to the proximal promoter region of *Nanog* and mediates *Nanog* repression as ES cells exit pluripotency (Lin et al., 2005). Importantly, the suppression of *Nanog* expression is dependent on the phosphorylation of Ser315 of p53 although the executive kinase has not been identified (Lin et al., 2005). High throughput methods have not been employed to discover p53 target genes in differentiating ES cells in an unbiased manner.

13

p53 acts as a substantial barrier for induced pluripotency in somatic cells, suggesting that p53 is involved in cell fate regulation during pluripotency to differentiation transition (Hong et al., 2009; Kawamura et al., 2009; Lee et al., 2012; Li et al., 2009a; Marion et al., 2009; Sarig et al., 2010; Utikal et al., 2009). The action of p53 in restricting induced pluripotency is striking, however, without the molecular mechanisms downstream of this role is poorly understood. It is concerning whether this knowledge could be directly translated into operations in regenerative medicine using iPS cells that generated in the absence of p53.

1.2.3   p63 and p73 in development, differentiation and cancer

One possible explanation for the lack of developmental phenotype in p53 null mice is that p63 or p73 could be playing redundant roles as p53 (Stiewe, 2007). p63 and p73 are substantial regulators of differentiation and development. *Trp63* is present in *Xenopus* and is mainly expressed as ΔNp63 from NF stage 9 to frog, stages of organogenesis. *Trp63* is present and ubiquitously expressed in zebrafish (Lee and Kimelman, 2002; Marcel et al., 2011). Previous report indicated that p63 is expressed since E7.5 in mouse embryos as the TAp63 form, and ΔNp63 is expressed at a much later stage (Stiewe, 2007). Among the p53 family members, p63 null mutants exhibit the most severe developmental defects and suffer from peri-natal lethality. *Trp63*[-/-] mice display abnormal limb formation, craniofacial and epithelium morphogenesis, and aberrant neurogenesis but not apoptotic defects. (Mills et al., 1999; Yang et al., 1999a).

Trp73 is present in zebrafish and is expressed in the zygotes and early larva embryos during development (Pan et al., 2003; Rentzsch et al., 2003; Marcel et al., 2011). *Trp73* expression in adult zebrafish is confined to skin, fin, brain, ovary and testis (Pan et al., 2003; Rentzsch et al., 2003). Homologs of *Trp73* have not been identified in *Xenopus*. p73 null mice are viable, however, they display neurological, pheromonal and immunological defects (Yang et al., 2000). The causes of these defects are unclear. $Trp73^{-/-}$ mice are significantly less prone to develop tumors than $Trp53^{-/-}$ mice. Moreover, *Trp63* and *Trp73* mutations are not frequent in human cancers. These results imply that p63 and p73 are important regulators of differentiation and development, with less likelihood of involvement in apoptosis regulation and cancer development.

Mice that are double or triple deficient in the p53 family have not been fully characterized potentially due to developmental defects that compromises the viability of these mice. However, functional mouse embryonic fibroblasts (MEFs) have been generated from E13.5 embryos with all three double knockout combinations: $Trp53^{-/-};Trp63^{-/-}$, $Trp63^{-/-};Trp73^{-/-}$, and $Trp63^{-/-};Trp73^{-/-}$ (Flores et al., 2002). This attempt strongly suggest that knockout any two of the three p53 family members is not sufficient to cause severe early developmental defects. Strikingly, ectopic expression of dominant negative ΔNp63 isoform inhibits mesoderm formation and ΔNp63 knockdown in early *Xenopus* embryo induces mesoderm in a p53-dependent manner (Barton et al., 2009). Deregulated transgenic expression of the ΔNp73, a p73 isoform that antagonizes all

trans-activating forms of p53 family, results in early mouse embryonic lethality at gastrulation (Erster et al., 2006; Huttinger-Kirchhof et al., 2006; Yang et al., 2000). These evidences strongly suggest that the balance between the transcription inhibitory isoforms (ΔNp63/ΔNp73) and transcription activating isoforms (p53/TAp63/TAp73) is critical for proper development and differentiation. Thus, the role of the p53 family in early development potentially could only manifest in the context of p53/p63/p73 triple knockout.

On the other hand, gain of p53 function through ablation of p53-inhibitory proteins is embryonic lethal. Knockout of other p53 regulators also leads to more severe phenotype than p53-depletion. For instance, knockout of *Mdm2*, a feedback target gene of p53 that regulates the ubiquitination and degradation of p53, leads to embryonic lethality in the mouse (Jones et al., 1995), while Mdm2/ p53 double deficient mice are viable (Montes de Oca Luna et al., 1995). In addition, depletion of Mdm4, another ubiquitin ligase of p53, leads to embryonic lethality, and it can be rescued by p53 depletion as well (Parant et al., 2001). Mdm2 null embryos exhibit increased apoptosis, which could potentially be explained by aberrant activation of the p53 pathway. These embryos display additional embryo patterning defects that are seemingly unrelated to p53's role in cytostasis and apoptosis (Jones et al., 1995). These results support the p53's involvement in the regulation of development and differentiation, although genetic ablation of p53 alone may not be sufficient to cause a developmental defect.

Generating triple knockout of the p53 family is the key experiment to illuminate whether the p53 family members are involved in early embryo development and whether their roles are masked by functional compensation. There are no reported attempts for triple knockout p53/p63/p73 in the mouse. Employing the CRISPR/Cas9 genome-editing technology, Zhong et al reported that *Trp53*$^{+/-}$;*Trp63*$^{+/-}$;*Trp73*$^{+/-}$ triple heterozygous embryos are viable and exhibit no morphological or developmental defects compared with their wild type siblings (Zhong et al., 2015). Crossing *Trp53*$^{+/-}$;*Trp63*$^{+/-}$;*Trp73*$^{+/-}$ mice will have an expected ratio of 1/64 for generating *Trp53*$^{-/-}$;*Trp63*$^{-/-}$;*Trp73*$^{-/-}$ embryos at a Mendelian ratio, a study that required a large number of embryos to be analyzed to obtain statistically significant results. Due to the lack of characterization of p53/p63/p73 triple knockout mice, whether this family is directly involved in development remains elusive. Moreover, whether p53/p63/p73 share redundant functions in differentiation and development remains to be investigated. Beyond early embryo patterning, p53 family members play redundant roles in myogenic differentiation and rhabdomyosarcoma development (Cam et al., 2006). Myoblast-specific expression of ΔNp73α interferes with myogenic differentiation through de-regulated Rb signaling (Cam et al., 2006).

1.2.4   Regulators and mediators of the p53 family in differentiation

Though the downstream targets of p53 in mediating cytostasis and apoptosis have been well characterized, the mediators of p53's functions

in differentiation remain poorly characterized. In ES cells, p53 activated by genotoxic agents induces expression of Wnt signaling pathway components (Lee et al., 2010). In another report, p53 directly binds to the *Nanog* promoter and enhancers and suppresses *Nanog* expression as ES cells exit pluripotency (Lin et al., 2005). Given the plethora of p53 binding sequences and target regions, genome-wide analysis of p53 binding sites and systematic transcriptome analyses are required to dissect the engagement of p53 in mesendoderm differentiation.

In pluripotent stem cells, Aurora kinase A (Aurka) phosphorylates p53 at Ser212 (Serine 215 in human p53), which inhibits the activity and stability of p53 protein. The repression of p53 by Aurka is relieved with the onset of differentiation as Aurka protein level dramatically reduces during this process (Lee et al., 2012). In addition to affecting protein stability, Ser212 phosphorylation also alters the DNA binding and transactivation capacity of the p53 protein (Liu et al., 2004). In normal cells, cell cycle regulators modulate the activity of the Aurka-p53 regulatory axis. However, it remains unclear what directly regulates the activity or abundance of Aurka in ES cell differentiation.

1.2.5 Remaining questions

A number of questions remain in the role of p53 family in differentiation and development: Are p53 family members directly involved in early embryo

18

development? Could the requirement of any of these factors be bypassed in mouse development? How to resolve the discrepancies between the essential roles of p53 in pluripotency and differentiation transition, and the lack of developmental defects in p53 null mice? What are the downstream mediators of p53 in differentiation? What are the roles of p53 in pluripotent embryonic stem cells? How is the activity and stability of p53 regulated during the embryogenesis of different vertebrates? How are the activities of the p53 family members coordinated with other developmental pathways that are active during the same processes?

1.3 TGF-β/Nodal pathway in development and differentiation

1.3.1 Overview of the TGF-β pathway

The transforming growth factor beta (TGF-β) superfamily of growth factors includes the TGF-β family and the bone morphogenetic protein (BMP) family. Nodal, Activin and TGF-β1/2/3 are representative members of the TGF-β family. Receptors for the TGF-β family growth factors are pairs of type I and type II transmembrane serine/threonine kinases. Ligand binding induces the formation of a hetero-tetrameric receptor complex, which phosphorylates the receptor-regulated Smads, Smad2/3 (Massague et al., 2005). Phospho-Smad2/3 form complexes with Smad4 and enter the nucleus to assemble transcriptional complexes (Massague et al., 2005) with various DNA-binding transcription factors. TGF-β pathway could activate or repress target genes, depending on the presence of transcriptional co-activators or co-repressors (Massague et al., 2005).

TGF-β family members are essential regulators of embryonic patterning and organogenesis, cell differentiation, cell cycle and cytostasis, wound healing, and tissue homeostasis. For instances, TGF-β recruits and activates monocytes and stimulates inflammatory response (Ashcroft, 1999). In the adaptive immune system, TGF-β directly modulates immune tolerance by inhibiting the proliferation and activation of effector immune cells (Ruscetti et al., 1993). During

myogenesis, TGF-β inhibits the fusion of myoblasts and blocks the induction of myotube differentiation factors (Olson et al., 1986). In diseases such as cancer, TGF-β suppresses tumorigenesis yet promotes metastasis to various organs (Massague, 2012).

The prototype growth factor, *Nodal*, was first identified through retrovirus induced mutations that could compromise mouse embryo viability (Conlon et al., 1991; Conlon et al., 1994; Zhou et al., 1993). Nodal expression can be first detected at E5.5 in the primitive ectoderm of mouse embryos (Conlon et al., 1994); then it becomes restricted to the proximal posterior ectoderm and primitive streak regions and is further confined to the anterior primitive streak node (Collignon et al., 1996; Conlon et al., 1994; Varlet et al., 1997; Zhou et al., 1993). Nodal binds to the type I activin receptors ActRIB (ALK4) or ActRIC (ALK7) and type II Activin receptors ActRIIA/B (Gu et al., 1999; Oh and Li, 1997; Song et al., 1999). ActRIB mutant embryos exhibit disorganized extra-embryonic epithelium and failed mesoderm induction, and ActRIB null ES cells failed to contribute to posterior primitive streak (Gu et al., 1998). The Activin receptors are essential for egg cylinder growth, gastrulation, and rostral head development in mice (Song et al., 1999).

Nodal is also strongly required for the development of lower vertebrates. In zebrafish embryos, nodal signaling is required for mesoderm and endoderm development. Embryos depleted of nodal exhibit defects in head, trunk

mesoderm and endoderm formation (Feldman et al., 1998; Gritsman et al., 1999). In *Xenopus*, nodal-related factors (Xnrs) are required for the initiation of gastrulation, dorsal mesendodermal specification and head formation (Osada and Wright, 1999). Inhibition of Xnr signaling by dominant negative constructs blocks mesoderm and endoderm induction (Osada and Wright, 1999).

The Nodal/Activin pathway and the TGF-β pathway use distinct receptors but share the downstream effectors Smad2/3, whereas Activin (Inhibin β) is a homolog of Nodal that binds to the same receptors and signals through the same Smad pathway downstream (Ling et al., 1986; Vale et al., 1986). Activin and Nodal share high amino acid sequence homology (Ling et al., 1986; Vale et al., 1986). In *in vitro Xenopus* embryo explant assay, Activin exhibits strong activity in inducing mesoderm (McDowell and Gurdon, 1999). In mouse embryos, Activin expression is detected in the blastocyst but disappears in the epiblasts (Albano et al., 1993). Inhibin β depletion did not alter mouse early embryo development (Matzuk et al., 1995a; Matzuk et al., 1995b). Activin is widely used as a substitute for Nodal in *in vitro* cultured cells to activate the pathway since recombinant Activin protein can be maintained more stably.

The Nodal/Activin signaling cascade is activated in the pre-implantation mouse embryos as early as four-cell stage due to Activin expression (Albano et al., 1993; de Sousa Lopes et al., 2003; James et al., 2005; Paria et al., 1992; Whitman, 1998), and it is involved in the maintenance of pluripotency in epiblast

stem cells (Vallier et al., 2009; Brons et al., 2007; Mesnard et al., 2006). *In vitro*, recombinant Activin A sustains the undifferentiated state of hES cells and mEpiSC (Vallier et al., 2004; James et al., 2005; Brons et al., 2007; Beattie et al., 2005). TGF-β/Activin/Nodal is also required for *ex vivo* mouse blastocyst outgrowth (James et al., 2005). Pharmacological inhibition of Nodal/Activin signaling impairs pluripotency. Activin/Nodal is dispensable for maintaining the pluripotency of mouse ES cells but rather plays a role in driving cell proliferation (Ogawa et al., 2007).

Following implantation, Nodal signaling defines the proximal-distal axis, which further establishes the anterior-posterior axis of the developing embryo. At later stages of embryogenesis, Nodal governs left-right axis asymmetry and the further patterning of the neural and gut tubes (Pauklin and Vallier, 2015). Blocking the Nodal/Activin pathway compromises pluripotency and induces spontaneous differentiation towards the neuroectoderm lineage at E4.5-5.5 (Smith et al., 2008; Vallier et al., 2004; Camus et al., 2006; Mesnard et al., 2006). Further dissection revealed that epiblast Nodal signaling is involved in the induction and maintenance of primitive streak (Conlon et al., 1994; Zhou et al., 1993). Nodal is also involved in the positioning of anterior-posterior (A-P) axis, the left-right axis and the neural plate (Schier, 2003; Schier and Shen, 2000; Arnold and Robertson, 2009). In later stages of development, TGF-β signaling is required for maintaining tissue homeostasis by stimulating the differentiation of tissue resident stem cells (Massague, 2012; Oshimori and Fuchs, 2012).

Nodal signaling is interpreted quantitatively (Shimizu and Gurdon, 1999). Though Nodal is expressed near ubiquitously in the epiblast, the activity of Nodal signaling is precisely controlled by the gradient of Nodal antagonists Lefty1, Lefty2 etc. Ectopic expression of Lefty1/2 inhibits Nodal signaling and compromises the development of the anterior primitive streak derived tissues including the axial and paraxial mesoderm, and the definitive endoderm (Yamamoto et al., 2004). Abrogation of Nodal antagonist expression results in hyperactive Nodal signaling and over-expansion of mesoderm lineage (Iratni et al., 2002; Meno et al., 1999; Perea-Gomez et al., 2002).

1.3.2 Smad2/3/4 mediating Nodal/Activin signaling

Smad proteins are intracellular mediators of the TGF-β superfamily signaling. Three major classes of Smads are involved in this process, the receptor-activated Smads or R-Smads including Smad1/5/8 downstream of BMP signaling and Smad2/3 downstream of TGF-β/Nodal/Activin signaling, the common Smad or co-Smad Smad4, and inhibitory Smads or I-Smads Smad6/7 (Massague et al., 2005). This section focuses on the Smad proteins that are mediating the TGF-β/Nodal/Activin signaling.

Smad2 and Smad3 are the major downstream effectors of the TGF-β family. Upon receptor-mediated phosphorylation, Smad2 and Smad3 associate with Smad4 and enter the nucleus to activate transcriptional targets on the chromatin (Massague et al., 2005) (Figure 1-1). Smad2 and Smad3 both contain

an N-terminal MH1 domain, a linker region and a C-terminal MH2 domain, and share 92% of homology in amino acid sequences (Massague et al., 2005). Smad2 differs with Smad3 by the presence of a 10 amino acid motif at the very N-terminus and an alternatively spliced exon3 encoding a 30 amino acid insert in the MH1 domain. The MH1 domain typically mediates DNA binding, while the MH2 domain mediates homophilic or heterotypic protein-protein interactions. The linker region is the least conserved and contains multiple regulatory nodes (Massague et al., 2005).

Smad4, first identified as a tumor suppressor gene *deleted in pancreatic cancer* (*Dpc4*), is the only co-Smad for both the TGF-β and the BMP pathways. Similar to Smad2/3, Smad4 is composed of a highly conserved DNA-binding MH1 domain, a linker region and an MH2 domain that mediates protein-protein interactions. High frequency mutations occur on the interface of protein-protein interactions in Smad4 MH2 domain in cancerous tissues (Shi et al., 1998). In the absence of TGF-β stimulation, Smad4 proteins constantly undergo cytoplasm-nucleus shuttling through the Importin/CRM1 protein nuclear import/export molecular machinery (Pierreux et al., 2000; Randall et al., 2002; Watanabe et al., 2000). Though the R-Smad-Smad4 oligomers could serve as a nucleating platform for a diverse array of transcriptional regulators (Massague et al., 2005), the precise role of Smad4 in mediating TGF-β/BMP driven transcriptional responses has not been fully elucidated. One possible explanation for the unique requirement of Smad4 for transactivation is the presence of the

Smad-activation domain (SAD) in the C-terminal part of Smad4 linker region (Chacko et al., 2001; Chacko et al., 2004; de Caestecker et al., 2000). SAD could induce a unique conformation in Smad4 MH2 domain, rendering Smad4 a potent partner for p300/CBP transcription co-activators (Chacko et al., 2001; Chacko et al., 2004; de Caestecker et al., 2000).

1.3.3 Smad2 and Smad3: similarities and differences

Within the *Smad2* gene, the exon3 can be selectively spliced. This splicing event produces two isoforms: a short isoform without the exon3 (*Smad2* for its transcript) and a longer isoform retaining the exon3 ('*Smad*2X' for its transcript in this dissertation (Yagi et al., 1999; Takenoshita et al., 1998). The shorter isoform resembles the biochemical features of Smad3, whereas Smad2X differs in many aspects (Yagi et al., 1999). The amino acid sequence encoded by exon3 is strictly conserved across vertebrates. Though the biochemical impact of exon3 on Smad2X is poorly understood, previous reports implicated the exon3 as developmentally dispensable since mice exclusively expressing the short Smad2 isoform are viable and fertile (Dunn et al., 2005).

Developmental functions

The *Smad2X* isoform is the major transcript produced from the *Smad2* locus, and it is ubiquitously expressed throughout the epiblast at E6.5 (Waldrip et

al., 1998; Dunn et al., 2005). Complete abolishment of the *Smad2* gene products leads to embryonic lethality between E6.5-E8.5, exhibiting severe defects in gastrulation, mesoderm and definitive endoderm formation (Nomura and Li, 1998; Weinstein et al., 1998). In *Smad2*[-/-] embryos, expression of mesoderm markers *T/Brachyury*, *Fgf8*, *Lim1*, *Bmp4* is completely blocked and the proamniotic cavity is filled with ectodermal cells (Nomura and Li, 1998; Weinstein et al., 1998). These phenotypes resemble that in *Nodal*[-/-] embryos (Conlon et al., 1994; Zhou et al., 1993), confirming that *Smad2* gene products mediate signaling downstream of Nodal. Indeed, *Smad2*[+/-];*Nodal*[+/-] double heterozygous mutants failed to establish left-right axis and display gastrulation defects (Nomura and Li, 1998). A subset of *Smad2*[+/-] heterozygous embryos exhibited gastrulation defects, mandibular reductions and missing eyes (Nomura and Li, 1998), suggesting that the function of *Smad2* gene products is dose-dependent. In *Xenopus*, exogenous *Smad2X* expression could induce ectopic mesoderm formation (Baker and Harland, 1996).

A few other studies reported less severe phenotypes in *Smad2* mutants using different targeting strategies (Brennan et al., 2001; Heyer et al., 1999; Tremblay et al., 2000; Waldrip et al., 1998). It was later recognized that the second targeting approaches resulted in a 28KDa N-terminally truncated Smad2 protein (*Smad2*[ΔMH1/ΔMH1] for the genotype, and Smad2(C) for the remnant protein product). Smad2(C) is translated at Met241 and contains a partial linker region and a full MH2 domain (Das et al., 2009). In addition, Smad2(C) can be

phosphorylated in response to TGF-β treatment and can be incorporated into Smad4 and FoxH1 containing transcriptional complexes (Das et al., 2009).

Smad2$^{\Delta MH1/\Delta MH1}$ mouse embryos survived one day longer than full loss of *Smad2* function mutants (Heyer et al., 1999; Waldrip et al., 1998). The epiblast of *Smad2*$^{\Delta MH1/\Delta MH1}$ mutant embryos adopt a mesodermal fate, and are able to develop normal yolk sacs as well as fetal blood cells (Waldrip et al., 1998). In wild type (WT) embryos, T/Brachyury expression is induced at the proximal region of the epiblast since early gastrulation and expands to the head process and notochord at late gastrulation. In contrast, *Smad2*$^{\Delta MH1/\Delta MH1}$ mutant embryos induce T/Brachyury expression transiently but uniformly in the epiblast, which disappears rapidly as gastrulation proceeds (Waldrip et al., 1998). In chimeras that were generated via injecting *Smad2*$^{\Delta MH1/\Delta MH1}$ mouse ES cells into WT blastocysts, *Smad2*$^{\Delta MH1/\Delta MH1}$ cells retained their ability to contribute to mesoderm though they failed to contribute to the definitive endoderm (Tremblay et al., 2000). Consistently, injection of Smad2(C) cDNA potently induced ectopic mesoderm marker expression in *Xenopus* embryos, recapitulating the effect of full-length *Smad2X* cDNA injection (Baker and Harland, 1996). Therefore the major defect in *Smad2*$^{\Delta MH1/\Delta MH1}$ mutants is the failure to establish the anterior-posterior axis. Moreover, *Smad2*$^{\Delta MH1/\Delta MH1}$ mutant blastocysts could not host the normal development of WT ES cells due to extra-embryonic membrane defects. These results together suggest that the N-terminal (MH1 domain and part of the linker region) is dispensable for the transcriptional activity of Smad2X,

28

yet it is indispensable for embryo development potentially due to its intrinsic regulatory (inhibitory) roles on Smad2/2X MH2 domain. Interestingly, the left-right asymmetry is unaffected in $Smad2^{\Delta MH1/\Delta MH1}$ mutant embryos (Tremblay et al., 2000), suggesting that Smad3 could potentially compensate this function of Smad2X in inducing left-axis asymmetry genes.

In direct contrast to the severe developmental lethality in *Smad2* mutants, homozygous *Smad3* mutant mice are viable and developmentally normal, though adults exhibit immunological defects, forelimb malformation, reduced enamel mineralization, impaired ovary development and osteoarthritis (Yang et al., 1999b; Ashcroft et al., 1999). The contrasting viability of *Smad2* null and *Smad3* null mice is surprising given how similar structurally and functionally these two proteins are.

In hES cells, *Smad2* gene products rather than Smad3 mediates the pluripotency-maintenance function of Nodal/Activin (Sakaki-Yumoto et al., 2013). Smad2X is selectively recruited to the *Nanog* promoter in hES cells (Sakaki-Yumoto et al., 2013; Beyer et al., 2013). Knockdown *Smad2/2X* gene products in human pluripotent stem cells induce spontaneous differentiation towards the neuroectoderm, mesoderm and trophectoderm lineages but not the definitive endoderm lineages (James et al., 2005). It is unclear how *Smad2* gene products (presumably by Smad2X) but not Smad3 mediate pluripotency and differentiation.
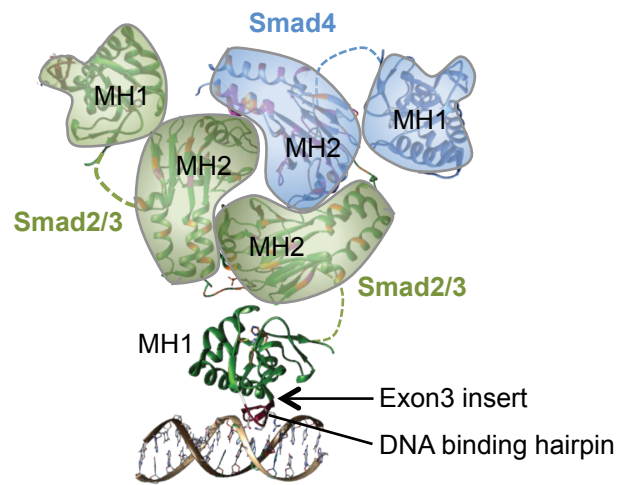
Figure 1-1.   Scheme of a trimeric Smad2/3/4 complex. Each Smad2/3/4 protein is composed of a conserved MH1 domain and a conserved MH2 domain, connected by a flexible linker region. The MH1 domain of Smad3/4 directly binds to 'Smad binding element' (SBE) sequences. *Smad2* gene includes an additional exon (exon3). The 30 amino acid insert encoded by exon3 is located next to the DNA-contacting β-hairpin in the MH1 domain. The MH2 domain mediates homo- or hetero- oligomerization of Smads, as well as interactions between Smads and other co-factors.

Molecular functions

Majority of the Smad transcription factors are capable of binding to specific DNA sequences known as Smad binding elements (SBE) (Dennler et al., 1999; Dennler et al., 1998; Massague et al., 2005; Zawel et al., 1998). Smad3 MH1 domain contacts its target DNA sequence at the major groove through an 11-residue β-hairpin (Shi et al., 1998). Each unit of MH1 domain recognizes a CAGA motif or its reverse complement (Shi et al., 1998). The binding affinity of Smad3 MH1 domain for palindromic SBE sequences is 1.14×10^{-7} M whereas the that of Smad4 MH1 domain for the same sequence is 2.6×10^{-7} M, suggesting that the DNA binding specificity and affinities of Smad3 and Smad4 are highly similar (Shi et al., 1998). Since this palindromic sequence was derived from an *in vitro* screening with randomly synthesized oligos, other potential sequences were not characterized and the physiological relevance was not determined.

The DNA-binding feature of Smad2X MH1 domain has been controversial. Based on the close proximity of the exon3 insert to the DNA contacting β-hairpin, Smad2X MH1 was predicted to be incapable of binding DNA due to potential allosteric hindrance (Figure 1-1) (Shi et al., 1998). However, Smad2X possesses many unique molecular functions that are divergent from Smad2/3. For instance, Smad2X is a potent activator of luciferase reporters driven by Activin response element (ARE, *Xenopus Mix.2* promoter sequence containing Forkhead binding motif and SBEs) in the presence of FoxH1 and Smad4, whereas Smad2/3 is

more active on Smad binding element reporter that contains three consecutive CAGAC sequences (Piek et al., 2001; Yagi et al., 1999). Since plasmid reporters lack the epigenetic environment that is found for target genes on the chromatin, the differential preference in target promoters/enhancers by Smad2X and Smad2/3 needs to be verified by ChIP-Seq analysis *in vivo*.

Biochemically, MH1 domain interacts with MH2 domain and inhibits the function of the MH2 domain in the basal state (Hata et al., 1997). Without TGF-β stimulation, endogenous Smad2X is detected both in the cytoplasm and in the nucleus (Xu et al., 2002). Smad2X-MH2-β-galactosidase chimeric protein is constitutively accumulated in the nuclei of *Xenopus* embryos (Baker and Harland, 1996). Moreover, the MH2 domain is the only Smad2X region that is recognized by the nuclear import machinery (Xu et al., 2000). These data together suggest that the MH1 domain inhibits the nuclear importing of Smad2X in the basal condition through masking the MH2-nuclear import machinery interface.

Smad2/2X/3 functions in mature cells

The functional differences between Smad2X and Smad2/3 are not limited to ES cell pluripotency and differentiation. In TGF-β induced Th17 cell differentiation, T-cell specific depletion of *Smad2* gene but not *Smad3*, blocks TGF-β-IL6 signaling crosstalk, and suppresses Th17 cell differentiation (Malhotra et al., 2010; Yoon et al., 2015a). In MEFs, while SERPINE1 can be induced

by either Smad2X gene or Smad2/3, Matrix metalloproteinase-2 (MMP2) is selectively dependent on *Smad2* gene products and *c-Fos*, *Smad7* and *TGFβ1* are dependent on Smad2/3 (Piek et al., 2001). Moreover, Smad3 impedes wound healing in mouse epithelia (Ashcroft et al., 1999). Oncoprotein Evi-1 inhibits Smad3 and represses TGFβ signaling (Kurokawa et al., 1998). The underlying mechanisms behind these functional differences remain unknown.

Distinctive regulation of Smad2X and Smad2/3 activity

Gel filtration analysis with endogenous proteins revealed that Smad2X and Smad2/3 display distinct oligomeric states (Jayaraman and Massague, 2000). In the un-stimulated state, majority of endogenous Smad2X exists as a monomer whereas Smad2/3 is present as multiple oligomeric states and Smad4 exists as homo-oligomer (Jayaraman and Massague, 2000). TGF-β treatment induces the formation of high molecular weight complexes for both Smad2X and Smad2/3 (Jayaraman and Massague, 2000). The distinct oligomeric states between Smad2X and Smad2/3 are dictated by the presence or absence of the 30 amino acid insert encoded by exon3 within the MH1 domain (Jayaraman and Massague, 2000). However, how the exon3 insert alters the oligomeric state of Smad2X remains elusive.

The involvement of Smad2/3 but not Smad2X in an oligomeric state under basal condition is consistent with their functional differences. Without

stimulation, Smad3 binds to a number of proteins that Smad2X cannot interact with, such as GSK3β, Axin, CKIγ2 (Guo et al., 2008a; Guo et al., 2008b). Smad2/3, but not Smad2X could be phosphorylated by GSK3β, which triggers protein ubiquitination and proteasome-dependent degradation (Guo et al., 2008a). The interaction between Smad3 and GSK3β/Axin could only take place in the absence of TGF-β mediated Smad3 phosphorylation. The GSK3β phosphorylation site on Smad3 is Thr66-a residue located two amino acids upstream of the exon3 insert (Guo et al., 2008a). T66V mutation is sufficient to prevent from GSK3β mediated phosphorylation and degradation. It is conceivable that the presence of exon3 insert on Smad2X will block its interaction with Axin/GSK3β and protects Smad2X from degradation. Indeed, inserting the exon3 sequences into the MH1 domain of Smad3 could revert the Axin/GSK3β mediated protein phosphorylation and degradation (Guo et al., 2008a). Whether this regulation plays a more active role in mediating TGF-β response has not been investigated. Moreover, Smad3 interacts with PP2A (PR65) in hypoxia through the MH1 domain (Heikkinen et al., 2010), whereas Smad2X MH1 domain did not interact with PP2A (PR65). This study suggests that the exon3 could potentially provent Smad2X from numerous interactions with the functional significance unclear.

1.3.4   Transcriptional co-factors of Smad2/2X/3 in mesendoderm differentiation

Due to the relatively low DNA binding specificity and affinity of the MH1 domains, Smad proteins often cooperate with a wide range of lineage-specific transcription factors to execute transcriptional responses. These co-factors are exemplified by Oct4/Sox2/Nanog in pluripotent stem cells, the Mix family (e.g. Mixl1) and a forkhead family transcription factor FoxH1 (or 'FAST2') in mesendoderm differentiation, MyoD in myocyte differentiation, and Pu.1 in macrophages (Mullen et al., 2011; Chen et al., 1996; Chen et al., 1997; Germain et al., 2000; Massague, 2012). The diversity of co-factors and target genes shed light on how each cell lineage interprets the TGF-β pathway activation differently with similar core effectors. However, it is unclear whether Smads and co-factors engage at different steps or different branches of the downstream cascade and regulate distinct target genes.

Smad4

Smad4 transcripts can be detected at high levels in pluripotent cells, and it is ubiquitously expressed in the epiblast of E6.5-E7.5 embryos. While *Smad4*[+/-] embryos are developmentally normal, *Smad4*[-/-] embryos die shortly after implantation (between E6.5 and E8.5), exhibiting little or no elongation of the extra-embryonic portion of the egg cylinder, failure in primitive endoderm and

mesoderm induction, and anterior patterning defects (Sirard et al., 1998; Yang et al., 1998). The severe developmental defects in *Smad4^-/-* embryos are consistent with the active participation of Smad4 in mediating both the BMP and TGFβ pathways, two fundamental signaling modules involved in pluripotency regulation and germ layer specification. *Smad4^-/-* embryos exhibit growth retardation and *Smad4^-/-* ES cells display proliferation defects (Sirard et al., 1998; Yang et al., 1998). Whether Smad4 is involved in mediating mitogenic signals in addition to transmitting differentiation cues in the epiblast remains to be explored.

FoxH1

FoxH1 is a winged helix transcription factor (Chen et al., 1996; Chen et al., 1997; Liu et al., 1999a), and contains a highly conserved DNA-binding forkhead domain that recognizes consensus sequences represented by TGTTGATTGG. FoxH1 and Smad2X form transcriptional complex and binds to regulatory DNA elements such as "Activin Response Element" (ARE) (Chen et al., 1996), a 50bp sequence identified in *Xenopus Mix.2* promoter. FoxH1 is widely expressed in the epiblast but the expression diminishes after gastrulation (Weisberg et al., 1998; Saijoh et al., 2000). FoxH1-deficient mouse embryos exhibit aberrant patterning of the anterior primitive streak and mesendoderm (Hoodless et al., 2001; von Both et al., 2004; Yamamoto et al., 2001). Zebrafish fast1 mutants display defects in dorsal midline cell types and disruption of left-right asymmetry (Pogoda et al., 2000; Sirotkin et al., 2000). FoxH1 interacts with Smad2 through

a common Smad interaction motif (SIM) (Germain et al., 2000; Randall et al., 2002). During OCT3/4, SOX2, KLF4 and c-MYC induced reprogramming for pluripotency, human somatic cells transiently undergo a primitive streak-like state (Takahashi et al., 2014). FoxH1-depletion prevents cells from experiencing this state, and diminishes the efficiency of reprogramming (Takahashi et al., 2014). These studies underscore the central role of FoxH1 in specifying the primitive streak and mesendoderm lineage.

Yap/Taz

Yes-associated protein1 (Yap1, Yap), together with its functional homolog Taz (transcriptional co-activator with a PDZ-binding domain) are transcriptional effectors downstream of the Hippo pathway. Lacking intrinsic DNA binding property, Yap/Taz cooperate with DNA-binding TEAD transcription factors for target gene recognition. Yap/Taz is involved in the maintenance of pluripotency in human ES cells (Beyer et al., 2013; Lian et al., 2010), and ectopic Yap/Taz activation directly induces "naive" pluripotency in human ES cells (Qin et al., 2016). Loss of Yap and Taz together in mouse leads to early embryonic lethality at morula stage (Nishioka et al., 2009). Recent reports suggested that Yap counteracts with Activin A signaling and represses mesendoderm gene induction through recruitment of negative elongation factor (NELF) and blockade of transcription elongation (Estaras et al., 2015).

1.3.5　　Epigenetic regulators of Nodal response

The pluripotency to differentiation transition is associated with dramatic changes in the epigenetic landscape and chromatin architecture (Dixon et al., 2015; Chen and Dent, 2014). The transcriptional output of Nodal pathway activation in ES cells is highly dependent on the local chromatin environment of target genes (Kim et al., 2011). The versatile and flexible genome-wide occupancies of Smads in various cell types reflect the impact of chromatin environment on accessibility by transcription factors (Mullen et al., 2011; Tufegdzic Vidakovic et al., 2015). A number of epigenetic regulators including Trim33, p300/CBP and Dpy30 have been reported to be Smad2/2X/3 binding partners (Massague, 2012), and play important roles in early development (Morsut et al., 2010; Xi et al., 2011; Bernstein et al., 2006; Creyghton et al., 2010; Bertero et al., 2015).

Trim33

Trim33 (Tripartite motif containing 33, also known as transcription intermediary factor 1 gamma or TIF1γ) is a multifunctional protein that contains both a PHD domain and a Bromo-domain, reading dual H3K9me3-H3K18ac modifications on histones. In differentiating ES cells, it forms molecular complexes with Smad2/2X/3, which subsequently binds to promoters and enhancers of mesendoderm differentiation genes (Xi et al., 2011). The interaction

between Trim33 and Smad2/2X/3 is independent of the Smad2/2X/3-Smad4 interaction. Trim33-depletion elicits mesendoderm differentiation defects in pluripotent stem cells, and growth arrest coupled with lethality in early mouse embryos (Morsut et al., 2010; Xi et al., 2011). During erythroid differentiation, Trim33 mediates transcription elongation in response to TGF-β stimulation (Bai et al., 2010; He et al., 2006). It is unclear whether Trim33 is also involved in transcriptional elongation regulation of mesendoderm marker genes.

p300/CBP

p300/CBP (p300 or 'EP300', E1A binding protein, 300KDa; CBP, CREB-binding protein) are histone acetylation enzymes. p300/CBP mediates acetylation on lysine 27 of histone H3, a modification commonly associated with active enhancers (Bernstein et al., 2006; Creyghton et al., 2010). Moreover, p300/CBP recruits RNA polymerase and general transcription activator complexes. In response to TGF-β stimulation, p300/CBP interacts with Smad2/2X/3 to facilitate target gene activation (Pouponnot et al., 1998).

Dpy30

Dpy30 is a subunit of the COMPASS (Complex Protein Associated with Set1) methyl-transferase complexes (Bertero et al., 2015; Jiang et al., 2011). In human ES cells, Activin-stimulated Smad2/2X/3 recruits the DPY30-COMPASS

complex to the core promoters of Activin target genes, resulting in histone H3 lysine 4 tri-methylation and transcriptional activation (Santos-Rosa et al., 2002). Depending on the context, the Smad-COMPASS machinery regulates the expression of both pluripotency and differentiation genes. Dpy30-depleted mouse embryos arrest at pre-gastrulation stage due to failure in maintaining pluripotency and self-renewal in the ICM (Bertero et al., 2015).

1.3.6   Target genes of Nodal pathway

Lineage-specific transcription factors/effectors

Nodal/Activin/TGF-β pathway activates a myriad number of target genes in development, differentiation and tissue homeostasis. For instance, a number of cell lineage markers have been identified as direct target genes of Nodal/ Activin including primitive streak and mesendoderm markers *Eomes*, *Gsc*, *Mixl1* and *Foxa2* (refer to **session 1.1.2**)(Estaras et al., 2015; Funa et al., 2015; Schrode et al., 2014). FoxH1-dependent Nodal target gene *Pitx2* is a bicoid-type homeobox transcription factor involved in left-right axis specification (Faucourt et al., 2001; Piedra et al., 1998; Ryan et al., 1998; Shiratori et al., 2001; Yoshioka et al., 1998). Moreover, TGF-β stimulates the expression of a wide range of transcription factors and functional effectors: cell cycle regulators p27 and p57 (Lee et al., 1995; Polyak et al., 1994a; Polyak et al., 1994b), pancreatic lineage-specification factor Sox4 and Klf5 (David et al., 2016), epithelial-mesenchymal transition regulators Snail1/2 and Zeb1/2 (Yang and Weinberg, 2008)

(Massague, 2012), T_{reg} cell specific transcription factor Foxp3 in (Chen et al., 2003), angiogenesis inhibitor Thrombospondin (Thbs1/2) (Negoescu et al., 1995), extracellumar matrix protein fibronectin (Fn) (Wikner et al., 1988), and matrix metalloproteinase 2 (Mmp2) (Lin et al., 2000), etc.

Nodal antagonists and intracellular feedback regulators

Lefty1 and Lefty2 (Left-Right Determination Factor 1 and 2) are members of the TGF-β family yet they are functionally distinct as they are incapable of forming homo- or heterodimers. Lefty1/2 act as extracellular antagonists of TGF-β family growth factors (Cheng et al., 2000). During embryo development, Lefty1/2 sequesters Nodal from interacting with its receptor, or binds to the Nodal co-receptors Cripto/Cryptic and prevents active ligand-receptor complex formation (Chen and Shen, 2004). The antagonism between Lefty1/2 and Nodal signaling determines the embryo midline formation, establishes the left-right asymmetry, and impacts the A-P axis patterning (Bisgrove et al., 1999; Chen and Shen, 2004; Yamamoto et al., 2004; Cheng et al., 2000).

Smad7, Skil and SnoN are Smad negative feedback regulators that are uniformly induced by Nodal/TGF-β signaling in most cell types (Massague et al., 2005). Smad7, the inhibitory Smad for the TGF-β pathway branch, is conserved from drosophila to vertebrates (Hayashi et al., 1997; Nakao et al., 1997). The MH2 domain of Smad7 resembles that of Smad2/2X/3/4, and Smad7 competitively binds to TGF-β family receptors. In addition, Smad7 interacts with

41

the Smurf2 ubiquitin ligase and promotes the ubiquitination and degradation of TGF-β receptors. Both actions of Smad7 lead to attenuation of TGF-β signaling. Skil and SnoN are able to bind the same DNA sequences as Smad2/2X/3/4 and act as potent feedback regulators of Nodal/TGF-β signaling in various contexts (Macias-Silva et al., 2002; Mizuide et al., 2003; Stroschein et al., 1999). Skil recruits N-CoR/mSin3/HDAC repressor complexes to TGF-β target genes through interacting with Smad2/2X/3/4 (Macias-Silva et al., 2002; Mizuide et al., 2003; Stroschein et al., 1999).

1.3.7   TGF-β/Nodal pathway in cancer

The TGF-β pathway is a major tumor suppressor in adult tissues and it is extensively mutated in human cancers. Most cancers disable the tumor suppressor function of TGF-β through mutations or deletions of TGF-β receptors or Smad4. The tumor suppressive effect of TGF-β signaling is attributed to its ability in inducing cytostatic responses, however, recent investigation uncovered that conflicts between master differentiation transcription factors and oncogenic signals is the major cause of apoptosis in pre-malignant pancreatic progenitors (David et al., 2016). Lineage-specific transcription factors actively participate in maintaining tissue homeostasis in adult tissues. Moreover, the differentiation trajectory potentially impacts disease development. For instance, while TGF-β/ Activin/Nodal signaling is a key signaling module in guiding the formation and differentiation of the mid-gut and hind-gut, interrogation of the mutation pattern of

TGF-β pathway components in various human cancers revealed that mutations are highly enriched in tumors that are definitive endoderm origin, including colorectal cancer, pancreatic adenocarcinoma and stomach adenocarcinoma (Figure 1-2). This observation suggests that differentiation cues could potentially impede malignant transformation in adult tissues.

1.3.8 Remaining questions about Nodal pathway in development

TGF-β regulates a plethora of biological processes and each cell type responds to this family of growth factors with different physiological consequences. There are a few questions that are intriguing and important: How are the seemingly opposing functions of Activin/Nodal signaling in maintaining pluripotency and stimulating mesendoderm differentiation coordinated? One hypothesis is that Smads switch molecular partners from Nanog-Oct4-Sox2 to Tcf/β-Catenin. As the list of Smad2/3 interacting partners accrue, is there a uniform mechanisms that determines the interaction between the TGF-β pathway and the cell-type specific factors? How does TGF-β activate lineage-specific target genes and a set of universal target genes in the same cells?

% 1 2 5 10 20 30 40 50

Cancer Type	TGFBR1	TGFBR2	SMAD2	SMAD3	SMAD4	Mutant cases	Total cases	%	Pathway Mutation rate % 0 20 40 60 80
Colorectal_hypermutated						28	35	80	
Pancreatic adenocarcinoma						17	42	40	
Stomach adenocarcinoma						59	219	27	
Colorectal_non−hypermutated						42	189	22	
Head and neck squamous cell carcinoma						48	302	16	
Uterine corpus endometrial carcinoma						29	240	12	
Lung adenocarcinoma						14	129	11	
Lung squamous cell carcinoma						20	178	11	
Ovarian serous cystadenocarcinoma						31	311	10	
Skin cutaneous melanoma						21	228	9	
Breast Invasive Carcinoma						43	760	6	
Prostate adenocarcinoma						4	82	5	
Lower Grade Glioma						9	218	4	
Kidney renal cell carcinoma						17	411	4	
Bladder Cancer						3	97	3	
Cervical Squamous cell Carcinoma						1	36	3	
Glioblastoma Multiforme						7	235	3	
Kidney renal papillary cell carcinoma						2	112	2	
AML						2	187	1	
Thyroid carcinoma						2	399	0.5	
Cancer Cell Line Encyclopedia						264	881	30	

Figure 1-2. Mutation frequencies of the TGF-β pathway components in various human cancer types. Sequencing samples for each cancer type were retrieved from cBioportal (MSKCC), a database that deposits samples sets from The Cancer Genome Atlas (TCGA, 2014). Colorectal cancer samples are subdivided into hypermutated group and non-hypermutated group based on the level of micro-satellite instability.

1.4 Wnt pathway in development and differentiation

1.4.1   The Wnt pathway

The mammalian Wnt family includes 19 cysteine-rich, secreted glycolipoproteins. Most Wnt proteins are approximately 40KDa in molecular weight, and are lipid-modified to assist secretion, and act in a paracrine fashion (Clevers and Nusse, 2012). The central mediator of Wnt signaling is β-Catenin. In the absence of Wnt, β-Catenin is constantly phosphorylated and degraded via activities of the Axin complex, which include Axin-the scaffold protein, casein kinase 1 (CK1), glycogen synthase kinase 3 (GSK3) and adenomatous polyposis coli (APC). In the presence of Wnt, Wnt ligands bind to their receptors Frizzled (Fzd) and co-receptors Lrp5/6, which recruit the scaffold protein Dishevelled (Dvl). This interaction leads to Lrp5/6 phosphorylation and activation, followed by the recruitment of Axin to the Wnt-Fzd-Lrp5/6 ligand-receptor complexes (Clevers and Nusse, 2012). Relocation of Axin causes destruction of the Axin complex and stabilization of β-Catenin. β-Catenin released from inhibition translocates to the nucleus, interacts with Tcf/Lef transcription factors and activates transcription of Wnt downstream target genes (MacDonald et al., 2009). Mice deficient in β-Catenin, Porcupine or Lrp5/6 suffer from embryonic lethality (Biechele et al., 2011; Fu et al., 2009; Hsieh et al., 2003; Huelsken et al., 2000; Kelly et al., 2004).

1.4.2   Wnt in development

The Wnt pathway is involved in the development of all model organisms examined to date. Due to potential functional compensation, knockout alleles of each Wnt family member did not cause severe developmental defects. As the first identified Wnt family member, Wnt1 transcripts emerge at E10.5 in the diencephalon area and Wnt1 protein performs a critical role in midbrain formation (McMahon and Bradley, 1990; McMahon et al., 1992; Nusse and Varmus, 1982; Thomas and Capecchi, 1990). Wnt3 expression initiates in the proximal epiblast of the egg cylinder, and is then detected in the primitive streak and mesoderm. *Wnt3$^{-/-}$* mouse embryos fail to form anterior-posterior axis, primitive streak and mesoderm, and die prior to E10.5 (Liu et al., 1999b), suggesting that Wnt3 is an indispensable regulator of primitive streak and mesendoderm specification. Further study with visceral endoderm-specific Wnt3 depletion suggests that the extra-embryonic tissue is an important source of Wnt3 that guides the temporal establishment of primitive streak in the epiblast (Yoon et al., 2015b). In parallel, epiblast-specific depletion of Wnt3 compromises the maintenance but not the initiation of primitive streak formation (Tortelote et al., 2013).

1.4.3   Tcf/Lef transcription factors and development

The Tcf/Lef family is the major DNA-binding partner of β-Catenin. Drosophila and C.*elegans* have one single tcf gene whereas mammalians

contain four Tcf/Lef members, Tcf1 (Tcf7), Tcf3, Tcf4 and Lef1. Each factor plays overlapping and distinctive roles. Alternative splicing and usage of alternative transcription start sites further diversifies the protein members of this family (Arce et al., 2006; Hoppler and Kavanagh, 2007). Most Tcf/Lef factors contain a high-mobility group (HMG) DNA-binding domain, and recognize similar consensus DNA sequences represented by TCAAAG. Tcf3 is ubiquitously expressed in embryonic structures at E6.5 whereas Tcf4 is only expressed since E10.5 in di- and mesencephalon and the intestinal epithelium (Korinek et al., 1998). Though Tcf4 predominantly engages in maintaining intestinal epithelium homeostasis, it is unlikely to be involved in gastrulation and germ-layer specification due to its limited expression (Clevers and Nusse, 2012). Tcf1 and Lef1 are co-expressed in many embryonic tissues.

1.4.4   Wnt pathway in pluripotency and differentiation

In pluripotent stem cells, the roles of Wnt signaling in regulating pluripotency and differentiation are contradictory. Inhibition of GSK3 is a widely used approach to maintain the pluripoteny and self-renewal of embryonic stem cells (Sato et al., 2004). However, inhibition of Wnt signaling in mouse ES cells by competitive inhibitor DKK1 blocks mesoderm differentiation (Lindsley et al., 2006). Treatment of human ES cells with recombinant Wnt3a results in down-regulation of pluripotency markers and differentiation toward the mesendoderm lineage. In pluripotent conditions, Tcf3 cooperates with the

Nanog-Sox2-Oct4 triad to support self-renewal and inhibit differentiation (Cole et al., 2008). As differentiation is permitted, Tcf3 acts as a gatekeeper for the onset of mesendoderm differentiation. Genetic ablation of Tcf3 in haploid stem cells blocks differentiation and enhances pluripotency (Leeb et al., 2014). *Tcf3[-/-]* null mouse ES cells self-renew robustly in LIF and are refractory to differentiation cues (Wray et al., 2011).

Wnt pathway activates feedback regulatory genes such as Frizzled, Tcf1, Lef1, Dkk1, and Axin2 in most Wnt-responsive cell types. In addition, Wnt signaling activates a myriad of target genes that differ in various contexts. In gastrulation, Wnt stimulation activates expression of Eomes, Xbra and Mixl1, etc in *Xenopus* and stimulates mesoderm formation (Tao et al., 2005). Wnt3 activated β-Catenin could directly stimulate Nodal expression through binding to an intronic enhancer (Norris and Robertson, 1999; Vincent et al., 2003). In human colon cancer cells, hyperactive Wnt signaling stimulates expression of c-Myc and drives tumorigenesis (He et al., 1998).

1.5 Crosstalk between signaling pathways

A delicately regulated signaling network controls the transition from pluripotency to differentiation in the embryos or ES cells. Multiple signaling pathways have been implicated in this process, including the TGF-β/BMP pathway, Wnt pathway, Fgf/MAPK pathway and PI3K/AKT signaling (Pauklin and Vallier, 2015; Wang and Chen, 2015; Young, 2011). Crosstalk between these pathways are often mediated by common players, physical interactions and transcriptional regulations.

1.5.1   TGF-β and Wnt crosstalk

The TGF-β and Wnt signaling pathways, acting through Smad and Tcf transcription factors, respectively, regulate the expression of lineage identity genes that control stem cell pluripotency and differentiation (Brennan et al., 2001; Clevers and Nusse, 2012; Massague, 2012). The two pathways are present as complete signaling modules even in simple metazoans that lack receptor tyrosine kinases and other major signaling pathways (Clevers and Nusse, 2012; Macias et al., 2015; Massague, 2012). Loss of Smad2/2X, Smad4 or Wnt3, and Tcf3 in embryos led to arresting at gastrulation stage and further lethality. The effects of TGF-β and Wnt are context dependent and often intertwined, as is evident in embryo development (Clevers and Nusse, 2012; Glinka et al., 1996; Massague, 2012; Oshimori and Fuchs, 2012), mammalian ES cell differentiation (Estaras et al., 2015; Funa et al., 2015; Labbe et al., 2000), and induced pluripotency

reprogramming (Ho et al., 2013; Ichida et al., 2009; Maherali and Hochedlinger, 2009; Marson et al., 2008). During mesendoderm lineage specification, a large set of cell lineage markers has been identified as direct target genes of Nodal/ Activin and Wnt pathway including primitive streak specification factors *Gsc*, *Eomes* and *Foxa2* (Estaras et al., 2015; Funa et al., 2015; Labbe et al., 2000).

The crosstalk between these two pathways can be mediated by physical interactions between Smad and Tcf/β-Catenin effectors and co-binding to regulatory DNA regions adjacent to target genes (Estaras et al., 2015; Funa et al., 2015; Labbe et al., 2000). Smad3 and Lef1 physically interact in response to TGF-β and Wnt signaling (Labbe et al., 2000). This interaction is mediated through both the MH1 and MH2 domains of Smad3 and the 324-334 amino acids in Lef1 (Labbe et al., 2000). Introducing a de novo SBE sequence into the TOP-flash luciferase reporter converts its promoter into a TGF-β-responsive promoter (Labbe et al., 2000). This molecular and functional cooperation provides a platform for the crosstalk between TGF-β and Wnt pathways. Using ChIP-Seq, RNA-Seq and GRO-Seq, Estaras et al identified a list of key developmental genes including *EOMES*, *MIXL1* and *T/BRACHYURY* that are synergistically activated by SMAD-FOXH1 complex and WNT-β-CATENIN cascade (Estaras et al., 2015). Nonetheless, how the TGF-β and Wnt signaling inputs are coordinated remains unknown.

Wnt signaling switches the function of Activin signaling from "pro-pluripotency" to "pro-differentiation". In the absence of Wnt3a, Activin/Nodal

signaling acts synergistically with activated PI3K/AKT signaling to promote the expression of pluripotency genes including *Nanog* (Singh et al., 2012). Up-regulated Wnt3a and down-regulated PI3K/AKT activity, together guided the target genes of ACTIVIN-SMAD to mesendoderm differentiation genes (Singh et al., 2012; Beyer et al., 2013).

1.5.2   TGF-β and p53 crosstalk

Defects of p53-deficient *Xenopus* embryos resemble Nodal-deficiency for lack of mesoderm, suggesting that these two pathways crosstalk during early embryo development. As suggested by Cordenonsi et al., and Takebayashi-Suzuki et al., p53 directly interacts Nodal-driven Smad2/2X/3 and binds to Mix2.promoter (Cordenonsi et al., 2003; Takebayashi-Suzuki et al., 2003). However, the cooperation between p53 and Smad2/2X/3 was described through *in vitro* binding assays with synthesized DNA oligos and purified proteins. No demonstration in *in vivo* settings has been reported, and it remains unclear whether the synergistic p53-Smad2/2X/3 cooperation applies to other target genes genome-wide.

1.5.3   Wnt and p53 crosstalk

DNA damage agents induced p53 activation in pluripotent stem cells induces the expression of Wnt pathway components (Lee et al., 2010).

Whether this regulation is involved in cell fate regulation has not been addressed. Moreover, the diverse functions of Wnt signaling in pluripotency and differentiation, the direct consequences of Wnt induction by p53 will be subjected to the physiological context. In colorectal cancer, Wnt signaling is involved in maintaining the self-renewal of crypt stem cells. Hyper-activated Wnt signaling through APC mutations is a major driver of colorectal cancer (TCGA, 2012; Dow et al., 2015). In these cells, p53 mutation is another pre-requisite for malignant transformation (TCGA, 2012). How the loss of p53 signaling facilitates Wnt-driven cancer development remains to be explored.

1.6 Metastatic stem cells

While normal tissues often exhibit strong hierarchical structures, the development of metastasis is a major exception that breaks the normal tissue structure restrictions. The emergence of metastases is frequently a lethal complication for cancer patients. Metastasis development from metastatic stem cells is a multi-step process. Invasive cancer cells first leave the primary site, intravasate into the blood or lymphatic circulation, disseminate and extravasate to host organs over the body (Nguyen et al., 2009; Obenauf and Massague, 2015; Massague and Obenauf, 2016; Oskarsson et al., 2014). After entering the target organs, disseminated tumor cells (DTCs) are challenged by multiple environmental factors including therapeutic agents, attacks from innate and adaptive immune system, lack of growth factors and cytokines etc. A small proportion of invaded cancer cells will survive, remaining latent or developing into overt metastasis (Obenauf and Massague, 2015; Massague and Obenauf, 2016; Oskarsson et al., 2014). Though the process of metastasis development is inefficient, there are very few effective treatments against metastatic cancers.

Tumors originating in different tissues metastasize to distinct target organs. Each organ differs in their cellular composition and microenvironment. For instance, for metastatic colonization in the brain, cancer cells will be faced with a tightly controlled blood-brain barrier that prevents extra-cranialfacial cells from entering the brain (Massague and Obenauf, 2016). Moreover, cancer cells

needs to overcome the pro-apoptotic insults exerted by plasminogen activators (Valiente et al., 2014). The bone microenvironment is marked by the continuous cycle of osteoclasts and osteoblasts. The osteolytic cycle generates numerous growth factors, cytokines and enzymes that require cancer cells to possess special signaling modalities to boost survival (Massague and Obenauf, 2016).

In the host organs, cancer cells are stimulated by stromal signals constantly and establishing interactions with the stromal components is the key for cell survival. Stimulants from the stromal include physical restrictions, growth factors and cytokines, proteases, extracellular components, connective tissue cells, etc. These factors not only contribute to the initiation and growth of metastasis, but also to the resistance to therapeutic agents. A clearer understanding about how disseminated tumor cells co-opt with the stromal components is key to develop therapeutic agents that blocks metastases.

A large part of our current knowledge about metastasis formation is developed from mouse modeling of metastatic cancers. Mouse modeling offers the advantage of characterizing cellular processes *in situ*. To dissect the molecular events dictating metastasis development, unbiased gene expression characterization of both the cancer and stromal components is in urgent need. However, very few tools could allow high throughput molecular characterization *in situ* and most current approaches involve flow cytometry based sorting, a laborious process that is likely to introduce alterations in the signaling pattern

of target cells. Novel technologies that permits in situ characterization of target cell types at the molecular level will significantly advance our understanding and handling about metastatic diseases.

Chapter 2   Materials and Methods

2.1 Molecular cell biology

2.1.1   Cell line maintenance and differentiation

E14Tg2a.IV mouse ES cells were maintained on gelatin (0.1%, Millipore, ES-006-B) coated plates in LIF-supplemented medium at 37°C with 5% CO_2 (Xi et al., 2011). Basic ES cell medium included 80% Knockout DMEM (Life Technologies, 10829-018), 15% Fetal Bovine Serum (HyClone, SH30071), 50U Penicillin and 50ug/ml Streptomycin (Cellgro, 30-001-CI), 1% Non-essential Amino acids (Life Technologies, 11140-050), 1% L-Glutamine (Life Technologies, 25030081), 100µM β-Mercaptoethanol (Sigma-Aldrich, M6250), 10^3 U/ml mouse LIF (Gemini Bio-Products, 400-495). Mouse ESCs $Trp53^{+/+}$ and $Trp53^{-/-}$ cell lines were kindly provided by Jing Huang (Li et al., 2012) and maintained in N2B27 LIF+2i (PD0325901, 1µM, Cayman; CHIR99021, 3 µM, Cayman) (Ying et al., 2008) medium. EB formation and differentiation were carried out as described by ATCC.

2.2 Plasmids, lentivirus and chemicals

Lentiviral infections and plasmid transfections were performed as previously described (Xi et al., 2011). The p53^{S212A} and p53^{S212D} plasmids (kindly provided by I. Lemischka), (Lee et al., 2012), Human TAp63α cDNA (Vigene Bioscience, CH842945) and human TAp73α cDNA (kindly provided by E. Flores) were sub-cloned into the pLVX-EF1α-IRES-mCherry vector (Clontech) and a FLAG- or HA- tag was added accordingly. p63 and p73 mutants were generated by site-directed mutagenesis using pairs of mutagenic primers: p63^{S285A}, 5'-cccat-cacaggaagacagg<u>c</u>tgtgctggtaccttatga-3'; p63^{S285D}, 5'-cccatcacaggaagacagg<u>a</u>tgt-gctggtaccttatga-3'; p73^{S235A}, 5'-caccggcaggcagg<u>c</u>cgtcgtggtgccc-3'; and p73^{S235D}, 5'-caccggcaggcagg<u>a</u>cgtcgtggtgccc-3'. To generate plasmids for doxycycline-in-ducible vectors of mouse *Wnt3* and *Tcf3*, the ORFs of *Wnt3* (42276,Addgene) and *Tcf3* (kindly provided by H. Nguyen) were cloned into pLVX-Tight-Puro vector (Clontech), separately. In addition, the CMV promoter present in plasmid pLVX-Tet-On was replaced with a pGK promoter to avoid epigenetic silencing in embryonic stem cells.

Wnt inhibitors IWP-2 (10536, Sigma-Aldrich) and XAV939 (X3004, Sigma-Aldrich) were used at 2.5 μM, and DKK-1 (5439,R&D) at 100ng/ml. SB431542 (Tocris, #1614) was used at 10μM. Recombinant human activin A (R&D Systems, 338-AC) was used at 50ng/ml. Recombinant mouse Wnt3a (R&D Systems, 1324-WN-010) was used at 150ng/ml. shRNA targeting vectors were

obtained from Mission TRC shRNA library (Sigma-Aldrich).

2.2.1 qRT-PCR analysis

For RNA extraction, $2x10^6$ ES or EB cells, or mouse embryos were collected at the indicated times and processed with the PrepEase RNA spin kit (Affymetrix). 500ng total RNA from each sample was used for cDNA synthesis with Transcriptor First Strand cDNA Synthesis Kit (Roche). Quantitative PCR was performed on a ViiA 7 Real-Time PCR System (Life Technologies). *B2m* was used as internal control for calculating relative expression.

Taqman primers used were: Mm00437762_m1 *B2m,* Mm01731290_g1 *Trp53* (all isoforms) Mm00660220_m1 *Trp73* (all isoforms), Mm01263634_g1 *Trp73* (isoform α), Mm01263882_m1 *Trp73* (ΔNp73 isoform α&ζ), Mm00495793_ m1 *Trp63* (all isoforms), Mm01150797_m1 *Trp63* (isoforms α-**γ**), and Mm01169470_m1 *Trp63* (ΔNp63 isoforms α-**γ**).

Sequences of synthesized primers used for qRT-PCR assays (designed with mouse genome mm9): *Gsc*-forward, TTGCACAGACAGTC-GATGCTACT, reverse, TCGTTGCTTTCTCGACCCC; *Mixl1*-forward, CG-GTTCTGGATCATCTCTCAA, reverse, TACCGAGAACAAGCCAGCAGT; *Skil*-forward, GACAGGGAGGCCGAGTATG, reverse, CCGCTCCTGTCT-GAGTTCAT; *T/Brachyury*-forward, TCCTCCATGTGCTGAGACTTGT, reverse, CCAAGAGCCTGCCACTTTG; *Eomes*-forward, GCGCAT-

GTTTCCTTTCTTGAG , reverse, GGTCGGCCAGAACCACTTC; *Foxa2*-forward, TACCCAGGGGGCTATGGT, reverse, CCCGCTTTGTTCGTGACT; *Nanog*-forward, TCTTCCTGGTCCCCACAGTTT, reverse, GCAAGAATAGTTCTCGGGATGAA , *Smad7*-forward, GACAGCT-CAATTCGGACAACA , reverse, CAGTGTGGCGGACTTGATGA ; *Wnt3*, TGGGCCTGTCTTGGACAAA , reverse, GCGATGGCATGCACGAA; *Axin2*-forward, TGACTCTCCTTCCAGATCCCA, reverse, TGCCCACACTAGGCTGACA.

2.2.2   Chromatin immunoprecipitation

Cells were cross-linked with 1% formaldehyde (Sigma-Aldrich) at 37°C for 10-15 min and quenched with 0.125 M glycine for 5 min at room temperature. ChIP was performed as previously described (Xi et al., 2011). Samples were incubated with 3–5 µg of antibody bound to 60 µl Dynabeads protein G (Life Technology), then incubated overnight at 4 °C. 1-2% pre-cleared chromatin prior to primary antibody addition was kept as input DNA. Magnetic beads were washed, chromatin was eluted, and reverse cross-linked ChIP DNA was dissolved in 10 mM Tris pH 8.0 buffer for further analysis. For ChIP-qPCR, immunoprecipitated DNA was analyzed by qRT-PCR, and the amplification product was expressed as percentage of the input, or then normalized to the control experiment for each condition. PCR primer pairs used to amplify the unrelated control or promoter regions of indicated genes are listed below. 1)

Gsc +6kb enhancer: 5'-CTAGGCTCCTAAACCAACAACA-3' (forward) and 5'-CGTCTGACACATCGGTTCATTA-3' (reverse); 2) *Eomes* -10kb enhancer: 5'-GCCCAGCGGGATGTTAAT-3' (forward) and 5'- AGGAGGAGCTATCT-GCTAGAC-3' (reverse); 3) *Eomes* +9kb enhancer: 5'-GCTATCTGCAGAC-GGCTTAAA-3' (forward) and 5'-AAATGACCCTCCCAGCTAGA-3' (reverse); 4) *Foxa2* -37kb enhancer: 5'-AAATGTGTCACCCAAGGCATTT (forward) and 5'-TTACCAGGTCATCAGTCTCAGC-3' (reverse); 5) *Foxa2* -50kb enhancer: 5'-TCCCAAGTGTTCTGTCCTGAAA-3' (forward) and 5'-CAAGGGGAGTCA-CAGGAAGC-3' (reverse); 6) *Foxa2* -53kb enhancer: 5'-CTCAGGTGGGCAAA-CAGTATCT-3' (forward) and 5'-AAAATCCCCATCCAAGTCAGCT-3' (reverse); 7) *Wnt3* -1.5kb: 5'-GAGGATCGGGCTAGGAACTCG-3' (forward) and 5'- GT-GTAAGGAGAGCAAGGAACTGG-3' (reverse); 8) *Axin2* promoter primers (Yi et al., 2011) were used as reported. Genomic positions of these primers are given relative to the transcription start site (upstream, -; downstream, +). Antibodies used for ChIP were: Smad2/3 (8685S, Cell Signaling Technology), Lef1 (sc-8591X, Santa Cruz Biotechnology), Tcf1 (2203S, Cell signaling), Tcf3 (sc-8635X, Santa Cruz Biotechnology), Tcf4 (sc-8631X, Santa Cruz Biotechnology), HA (11867423001, 3F10, Roche), FLAG (F3165, Sigma-Aldrich), and p53 (P53-505, Leica Biosystems and sc-6243, Santa Cruz).

2.2.3 Immunoblotting and immunoprecipitation

Cell pellets were lysed with RIPA buffer (Cell Signaling) and protein

concentrations were determined using the BCA Protein Assay Kit (Pierce). The Nuclear Complex Co-IP Kit (Active Motif) was used for immunoprecipitation. Proteins were separated by SDS-PAGE using Bis-Tris 4-12% gradient polyacrylamide gels in the MOPS buffer system (Life Technologies) and transferred to nitrocellulose membranes (BioRad) according to standard protocols. Membranes were immunoblotted with antibodies against Trim33 (A301-060A-1, Bethyl), Smad2/3 (8685S, Cell Signaling Technology), Smad4 (Ab40759, Abcam), p73 (ab26123, Abcam and sc-7957, Santa Cruz), p63 (sc-25268 and sc-8343, Santa Cruz Biotechnology), p53 (P53-505, Leica Biosystems), Wnt3 (ab52568, Abcam), Nanog (A300-397A, Bethyl Laboratories), Gsc (ab109024, Abcam), Aurka (610938, BD Transduction Laboratories), Nodal (WH0004838M1, Sigma-Aldrich), Tcf3 (sc-8635X, Santa Cruz Biotechnology), HA (11867423001, 3F10, Roche), FLAG (F3165, Sigma-Aldrich), FoxA2 (sc-6554, Santa Cruz Biotechnology), Eomes (ab23345, Abcam) and γ-Tubulin (T6074, Sigma-Aldrich) in Odyssey-TM blocking buffer (LI-COR). Following primary antibody incubation, membranes were washed and probed with IRDye 800CW donkey-anti-mouse IgG (LICOR) or IRDye 680RD goat-anti-rabbit IgG (LI-COR) secondary antibody and imaged using the LI-COR Odyssey system. All western immunoblots were performed independently at least twice. γ-Tubulin was used as a loading control for all experiments.

2.2.4   Genome-editing with CRISPR/Cas9

sgRNAs targeting genomic regions of interest were designed using CRISPR Design Tool (http://crispr.mit.edu/) (Hsu et al., 2013) and synthesized by IDT, Inc. Single cells were sorted onto irradiated MEF feeder for increased viability through FACS 72h post-transfection. Mutant clones were first screened through aberrant melting temperature of qPCR products, then verified by PCR, TA-cloning and Sanger sequencing individually. sgRNA target sequences used are as following: *Trp53* exon4 Cas9n sgRNA sense strand: ACCCTGTCACCGAGACCCC, anti-sense strand: AGGAGCTCCTGACACTCGG; *Trp63* exon4 sgRNA 1: CCGTCACGCTATTCTGTGCG; sgRNA 2; TGGGC-CCGGGTAATCTGTGT; *Trp73* exon4 sgRNA 1, CGGGGTGTAGGGGCTC-GCCG; sgRNA 2: CCGGGGTAGTCGGTATTGGA; *Eomes* -10kb enhancer sgRNA 1: TAACAGTATTAACATCCCGC, sgRNA 2: CTCTCCGCTTTGAT-GTGAGC; *Gsc* +6kb enhancer sgRNA1: CAGCACAGACTGTGTCCTG, sgRNA 2: CAGACCATGTTTTCAAAGC; *Ctnnb1* (β-catenin) N-terminus Cas9n sgRNA sense strand: AGTAGCCATTGTCCACGCAG, anti-sense strand: GTGATTCAGATGCCTGTCTG; FLAG-HA epitope tag oligo DNA template for *Ctnnb1*: CTTAAGTTTTAATGACTTGATGGAATTTTTCAGG-GTACCTGAAGCTCAGCGCACAGCTGCTGTGACACCGCTGCGTGGA-CAATGGACTACAAAGACGATGACGACAAGGCCGCATACCCATACGAT-GTTCCAGATTACGCTGCTACTCAAGGTTTGTGATTCAGATGCCTGTCT-GAGGATCTGCCTCATAGCCCTGCTGC. Annealed sgRNA oligos were

cloned into pX330-U6-Chimeric_BB-CBh-hSpCas9 (pX330, Addgene), pSpCas9(BB)-2A-GFP (pX458, Addgene) or pSpCas9n(BB)-2A-GFP (pX461, Addgene) vectors according to previous protocols (Ran et al., 2013) and transiently transfected into mouse ES cells with Lipofectamine 3000 (Life Technologies). Truncated sgRNA sequences for Smad2X exon3 deletion experiment: left sgRNA, GTGGCTCGTGCTGACCCGTTGGG, right sgRNA, TGGGACCCTAGAGACCGCGTGGG.

2.3 Biochemistry assays

2.3.1 Protein purification

cDNAs of Smad2/2X-MH1 domains (1-155 residues for Smad2 MH1 domain, and 1-185 residues for Smad2X-MH1 domain) were cloned into pFastbac1 baculovirus vector with N-terminal GST tag, or His_6 tag or FLAG tag. Hi-5 insect cells (Life Technologies, Carlsbad, CA) were repetitively inserted with the collected baculovirus and expanded before protein purification. Recombinant proteins were purified by GST-affinity chromatography, anion exchange, and gel-filtration chromatography. FLAG-affinity resin was utilized to affinity purify FLAG-Smad2/2X-MH1 protein. Buffer used for chromatography contains 40mM Tris-HCl, pH8.0, 300mM NaCl, 5% glycerol, 50μM ZnCl2 and 0.5mM TCEP.

2.3.2 Electrophoretic mobility shift assay

Electrophoretic mobility shift assay (EMSA) were performed in reaction buffer containing 20mM Tris-HCl, pH8.0, 125mM NaCl, 2% glycerol, 0.3mM TCEP and 1% BSA (Zhou et al., 2015). Reaction mixtures were prepared with reaction buffer, 10nM FITC-labeled DNA substrates, and increasing amount of recombinant proteins as indicated in each experiment. Reactions were incubated on ice for 45min. Electrophoresis was performed on 4% non-denaturing native PAGE gels for 40min at 4 °C.

2.4 Mouse and embryo analysis

2.4.1 Generation of chimeric embryos

mCherry expressing single ES cell colonies were picked and micro-injected 3 days after culture on irradiated MEF feeder layers. 10-15 ES cells from each group were injected into E3.5 blastocysts (C57BL/6N Taconic) on a Nikon (Eclipse-Ti) microscope equipped with Narishige micromanipulators. Injected blastocysts were cultured in KSOM /AA (Millipore) at 37°C in an atmosphere of 5% CO_2 for 2-3 h until blastocyst cavity expansion and implanted into the uterine horns (10 embryos per horn) of E2.5 pseudopregnant females (CD-1: Charles River) using standard protocols. Chimeric embryos were recovered at E7.25, E7.75, E8.5 and E10.5.

2.4.2 Immunofluorescence

For immunofluorescence (IF) embryos were fixed in 4% PFA in PBS 20 min at room temperature and then washed with 0.1% Triton in PBS. ISH was performed using antisense riboprobes as previously described (Nowotschin et al., 2013). Immunofluorescence was carried out as previously described (Nowotschin et al., 2013). Primary antibodies used: FoxA2 (1:100; sc-6554, Santa Cruz and ab108422, Abcam), T/Brachyury (1:100; AF2085, R&D Systems), Eomes (1:500; ab23345, Abcam). Secondary Alexa-Fluor conjugated

antibodies (Life Technologies) were used at a dilution of 1:1000. DNA was visualized using Hoechst-33342 (5 µg/mL, Molecular Probes). For cryosections, fixed embryos were taken through a sucrose gradient, embedded in O.C.T. (Tissue-Tek) and sectioned at 12 μm on a cryostat (CM3050S, Leica).

2.4.3   Image data acquisition, processing and quantitation

Wide field images were collected with Zeiss Axiocam MRc/m CCD cameras mounted on a Leica MZ165FC microscope. Laser-scanning confocal images were acquired using a Zeiss LSM880. Raw data were processed using ZEN software (Zeiss) and assembled in Photoshop CS6 (Adobe).

2.4.4   TRAP-Seq with tumor samples

To investigate the gene expression changes specifically of drug-sensitive tumours during vemurafenib treatment, or gene expression changes of resistant cells exposed to a regressing tumour microenvironment, A375 and A375R cells, respectively, were modified to express eGFP-RPL10a. Tumours derived from implanted A375-eGFP-RPL10a and A375R-eGFP-RPL10a cells were homogenized and processed with the TRAP protocol as previously described (Zhang et al., 2013; Heiman et al., 2008; Doyle et al., 2008) with the following modifications: fresh tumour was homogenized with a Model PRO 200 homogenizer at speed 5 for four cycles of 15 s, RNasin Plus RNase inhibitor

(Promega, N2615) was used as RNase inhibitor, and anti-eGFP antibody coated sepharose beads (GE Healthcare) were used for immunoprecipitation. Polysome-associated RNA was purified with RNAqueous micro kit (Life Technologies, AM1931). Ribogreen and the Agilent BioAnalyzer technologies were used to quantify and control the quality of RNA; 500 ng RNA (RNA integrity number (RIN) > 8.5) from each sample was used for library construction with TruSeq RNA Sample Prep Kit v2 (Illumina) according to the manufacturer's instructions. The samples were barcoded and run on a Hiseq 2000 platform in a 50-base-pair (bp)/50-bp or 75-bp/75-bp paired-end run, using the TruSeq SBS Kit v3 (Illumina). An average of 40 million paired reads was generated per sample.

2.5 Bioinformatics analysis

2.5.1   Statistical analysis

Quantitative data are expressed as mean± s.e.m. Statistical significance was determined using a two-tailed Mann-Whitney test using Prism 6 software (GraphPad Software) unless otherwise indicated (* $p<0.05$, ** $p<0.01$, *** $p<0.001$, n.s.= not significant).

2.5.2   RNA-Seq and data analysis

Total RNA purified from mouse ES cells and EBs was quantified by Ribogreen and quality assessed by Agilent BioAnalyzer 2000. 500ng RNA with integrity number (RIN) > 9.5 from each sample was used for library construction with TruSeq RNA Sample Prep Kit v2 (Illumina) according to manufacturer's instructions. Multiplexed sequencing libraries were run on a Hiseq-2500 platform and more than 40 million raw paired-end reads were generated for each sample. For data analysis, reads pairs in FASTQ format (50bp/50bp, or 75bp/75bp) were quality assessed by FastQC v0.11.3 and mapped to mouse genome mm9 (NCBI build 37, July/2007) with STAR2.3.0e (Dobin et al., 2013) using standard settings for paired reads. On average 82% of raw reads were uniquely mapped. Uniquely mapped reads were counted to each gene with HTSeq v0.5.4 with default settings. Read counts were normalized by library size through the "*DESeq2*"

(Anders and Huber, 2010; Love et al., 2014) package deposited in Bioconductor (www.bioconductor.org). Differential gene expression analysis between any two conditions was performed based on a model utilizing the negative binomial distribution (Anders and Huber, 2010; Love et al., 2014). Genes with FDR<0.05, fold change > 2.0 or < 0.5, and average normalized read counts > 10 were called as differentially expressed genes unless otherwise indicated. Basic statistical calculations were performed in R (v3.0.1). Heatmaps for RNA-Seq data were generated with *heatmap.2* function in *gplots* package.

2.5.3   ChIP-Seq and data analysis

For library construction and sequencing, ChIP-Seq DNA samples were quantified and quality assessed by Ribogreen and Agilent Bioanalyzer 2000. DNA fragments range from 200-600bp were selected constructed for ChIP-Seq library with TruSeq ChIP Sample Prep Kit (Illumina) according to manufacturer's instructions. Sequencing libraries were multiplexed and run on a Hiseq-2000 platform.

For mapping and visualization, single end (50bp) or paired-end (50/50bp) FASTQ reads were mapped to mouse genome mm9 (NCBI build 37, Jul/2007) with Bowtie2 with default filtering criteria (Langmead and Salzberg, 2012). Resulted SAM files were converted to BAM files through Samtools 0.1.19 (Li et al., 2009). BAM files were sorted and indexed with Samtools (Li et al., 2009).

To visualize ChIP-Seq data, BAM files were converted to TDF file by IGV Tools 2.3.32(Robinson et al., 2011) using the command "igvtools count -z 5 -w 25 -e 250", specifying the coverage window size to be 25bp and average fragment size of 250bp. TDF files were loaded into IGV genome browser (Robinson et al., 2011) and signal intensities normalized by "1x10^6/total million reads" or Reads Per Million Reads (RPM) to display normalized coverage data tracks. Scale bars indicating length of genomic ranges is provided above each gene track plot.

Peak calling from ChIP-Seq data was performed with MACS 1.4.2 and verified by HOMER (v4.2) (Heinz et al., 2010). The parameters for peak calling included fold change >8, p value < 1e-8 to detect high confidence binding events. Input samples were used as reference controls for background correction. Peaks identified from MACS 1.14.2 are annotated with HOMER (v4.2) using *annotatePeaks.pl* function. Genes are assigned with the "nearest TSS" criteria. Peak region overlap was performed with the *intersect* function from Bedtools 2.17.0 (Quinlan and Hall, 2010). Conservation data for mammals and vertebrates were extracted from UCSC genome browser and plotted together with ChIP-Seq data tracks. Differentially bound peaks between two conditions were identified by *mergePeaks.pl* function in HOMER (v4.2) and validated selectively in the IGV browser. Overlapped peaks were detected by *bed –intersection* tool from bedtools v2.25.0.

Tag density for genomic ranges surrounding defined peak centers were

calculated using *annotatePeaks.pl* function in HOMER (v4.2). log2 transformed tag densities were pre-ranked by peak score. Data matrix from each ChIP-Seq experiment were merged by peak names and plotted for heatmaps in R. White indicates low tag density and blue indicates high tag density in each figure.

DNA motif enrichment analysis was performed with HOMER (v4.2) and verified by MEME (Machanick and Bailey, 2011). BED file of Smad2/3 binding peak regions identified by MACS 1.4.2 and HOMER were used as input for HOMER *findMotifs.pl* tool.

For Gene Ontology analysis, genes adjacent to peaks of interest were used as input for the web tool of Gene Ontology (Gene Ontology Consortium) and results were plotted as bar graph with fold change and adjusted p-value.

2.5.4   Accession numbers

All RNA-Seq and ChIP-Seq data were deposited in the Gene Expression Omnibus database under accession number GSE70486.

Chapter 3   A   p53-Wnt-Nodal   Network   Driving Mesendoderm Differentiation

3.1 Summary

Mesendoderm specification involves multiple signaling inputs including Activin/Nodal and Wnt. However, both pathways have been implicated in conflicting roles in supporting pluripotency as well as guiding differentiation. It remains to be resolved about how the activities of these pathways are coordinated to transcriptionally regulate differentiation. In collaboration with postdoctoral fellow Qiong Wang, PhD, I delineated a regulatory network involving the p53 tumor suppressor family and the Wnt pathway, acting together with the TGF-β pathway, to drive mouse mesendoderm differentiation. Knockout of all three members, p53, p63 and p73, shows that the family is essential for mesendoderm specification as cells exit pluripotency *in vivo* and in culture. *Wnt3* and its receptor *Fzd1* are among a small set of p53 target genes that are specifically activated in this context. Induction of Wnt signaling by p53 is critical for activation of mesendoderm differentiation genes. Globally, I show that Wnt3-activated Tcf3 and Nodal-activated Smad2/3 transcription factors depend on each other for co-occupancy of target enhancer elements in master differentiation loci. Our results reveal a selective interdependence between

signal-activated Tcf and Smad transcription factors. Thus, the p53 family governs a regulatory network that integrates essential Wnt-Tcf and Nodal-Smads inputs for mesendoderm differentiation in the early embryo. This study advances our understanding about the precise regulation of mesendoderm differentiation and establishes the p53 family as central developmental regulators.

3.2 Introduction

The transcription factor p53 is abundantly expressed in mouse embryo germ layer progenitors and in embryonic stem (ES) cells, suggesting a role in early development (Bazzi and Anderson, 2014; Danilova et al., 2008; Lutzker and Levine, 1996; Schmid et al., 1991). This role is perhaps unrelated to the prominent role of p53 in DNA damage responses in the adult. p53 levels in somatic cells are kept low by complex regulatory mechanisms, rising sharply after DNA damage and other stresses to trigger cell cycle arrest, senescence, or apoptotic responses (Bieging et al., 2014; Zilfou and Lowe, 2009). The gene, *TP53* (*Trp53* in mice), is the most frequently inactivated tumor suppressor in human cancers (Bieging et al., 2014; Lane and Levine, 2010; Vousden and Prives, 2009). Despite an extensive current understanding of p53 as a tumor suppressor in the adult, its role in early embryo development remains unknown.

An involvement of p53 in early development is indicated by the failure of p53-depleted *Xenopus* embryos to undergo gastrulation (Cordenonsi et al., 2003; Takebayashi-Suzuki et al., 2003; Wallingford et al., 1997). p53 is required for mesendoderm differentiation of mouse ES cells in monolayer culture, though not in ES embryoid bodies (EBs) in suspension (Shigeta et al., 2013). p53 inhibits the generation of induced pluripotent stem (iPS) cells (Hong et al., 2009; Kawamura et al., 2009; Marion et al., 2009; Utikal et al., 2009). Albeit p53 regulates expression of LIF in the female for embryo implantation (Hu

et al., 2007; Kang et al., 2009), the role of p53 in the embryo proper remains controversial because *Trp53*-null mice develop normally, even if they are cancer-prone as adults (Donehower et al., 1992; Jacks et al., 1994; Purdie et al., 1994; Tsukada et al., 1993).

Functional redundancy of *Trp53* with its other two family members, *Trp63* and *Trp73*, might explain these discrepancies. The transactivating forms of p63 and p73 (TAp63 and TAp73) show structural similarity with p53, bind to the same consensus DNA sequence, and are interchangeable with p53 in certain assays (Dotsch et al., 2010; Flores et al., 2002). However, embryos with homozygous mutations of *Trp63* or *Trp73* (Mills et al., 1999; Yang et al., 1999a; Yang et al., 2000), or double knockouts of p53 family members develop well beyond gastrulation and germ layer specification, and have only late developmental defects (Flores et al., 2002). No triple knockout has been reported.

The mechanism by which p53 family members might control the expression of lineage identity genes in ES cells is also not clear. Germ layer specification genes are directly regulated by various embryo signals, including Wnt and the TGF-β family member Nodal. Nodal drives mesendoderm differentiation during gastrulation (Brennan et al., 2001; Conlon et al., 1991; Conlon et al., 1994; Zhou et al., 1993). It binds to activin receptors and activates Smad2/3 transcription factors that directly regulate mesendoderm differentiation genes (Baker and Harland, 1996; Weinstein et al., 1998). Wnt binds to frizzled (Fzd) receptors to activate β-Catenin and Tcf transcription factors (Clevers

75

and Nusse, 2012). Wnt cooperates with Nodal during mesendoderm induction (Estaras et al., 2015; Funa et al., 2015; Reid et al., 2012), while Tcf interact with Smad2/3 in certain contexts (Labbe et al., 2000; Nawshad et al., 2007). p53 was reported to bind Smad proteins for regulation of TGF-β target genes (Cordenonsi et al., 2003; Takebayashi-Suzuki et al., 2003), but these observations have not been widely confirmed.

Here in collaboration with Qiong Wang, PhD, I use double and triple knockout combinations of p53 family members in ES cells to address the role of p53 during the exit from pluripotency as cells undergo differentiation. I show that the p53 family is essential for mesendoderm specification through the activation of mesendoderm differentiation genes in the embryo and *in vitro*. Though these genes lack p53-binding elements, they contain enhancers that are synergistically co-occupied and regulated by Nodal-activated Smad2/3 and Wnt-activated Tcf3. I demonstrate that the p53 family enables activation of these genes by stimulating Wnt production as ES cells exit the pluripotent state, thus governing the cooperation of Wnt and Nodal for the onset of mesendoderm differentiation.

3.3 Results

3.3.1 The p53 family redundantly drives mesendodermal differentiation

p53 gain of function blocks induced pluripotency in mature cells and favors differentiation in mouse ES cells, yet loss of p53 function does not prevent differentiation and development consistently (Donehower et al., 1992; Hong et al., 2009; Kawamura et al., 2009; Lee et al., 2012; Lin et al., 2005; Shigeta et al., 2013). To investigate the role of p53 in differentiation and development, I modeled the gastrulation process through *in vitro* differentiation of embryonic stem (ES) cells.

Pluripotent ES cells and early stage EBs formed by these cells under differentiation-permissive conditions [suspension culture in leukemia inhibitory factor (LIF)-free media], recapitulate the signaling and transcriptional events of germ layer specification (Nishikawa et al., 1998; Xi et al., 2011). In LIF-free media, mouse ES cells quickly downregulate the pluripotency genes *Nanog*, *Sox2* and *Pou5f1* (Oct4) and gradually induce the expression of mesendoderm marker genes including *Eomes*, *Foxa2, Goosecoid* (*Gsc*), *Mixl1*, *Fgf8*, *T (Brachyury)*, and ectoderm marker genes including *Sox1*, *Fgf5*, *Nes* (Nestin) (Figure 3-1, 3-2). Expression of the mesendoderm markers is driven by autocrine Nodal through Nodal/Activin receptor kinases, and can be blocked with the

specific kinase inhibitor SB431542 (SB) (Ichida et al. 2009) (Estaras et al., 2015; Funa et al., 2015)

Study by Qiong Wang revealed that *Trp53*[-/-] ES cells failed to express the mesendoderm specification genes *Eomes*, *Foxa2* and *Gsc* in monolayer culture, but still expressed these genes in EB conditions (Figure 3-3, 3-4) (Shigeta et al., 2013). The defects in mesendoderm gene induction in *Trp53*[-/-] ES cells in monolayer did not alfter the down-regulation of pluripotency factors such as *Nanog*. In line with the lack of severe differentiation defects in EBs, *Trp53*[-/-] ES cells contributed to various tissues when injected into blastocysts and generated as chimeras (Shigeta et al., 2013).

I investigated if this discrepancy was linked to differences in the expression of p53 family members. Indeed, p53 was expressed both in monolayer and EB conditions, whereas the active form of p73, TAp73, was expressed in EBs but not ES monolayers (Figure 3-5). p63 was below the detection limit in both conditions (Figure 3-5). As revealed by Qiong Wang, shRNA-mediated knockdown of *Trp73* in *Trp53*[-/-] cells (*Trp53*[-/-];*Trp73*[sh] cells) eliminated the expression of *Eomes*, *Foxa2* and *Gsc* in EBs (Figure 3-4, 3-6), whereas expression of *Nanog* and the Smad negative feedback regulator *Skil* were unperturbed by the absence of p53 and p73 (Figure 3-4, 3-6).

Together with Qiong Wang, we performed RNA-Seq analysis of day 4

(d4) EBs derived from *Trp53*[+/+];control[sh] (*Trp53*[+/+];Ctrl[sh]) or *Trp53*[-/-];*Trp73*[sh] cells to systematically assess the changes in cell fate in response to p53/p73-depletion. *Trp73* mRNA knockdown in the *Trp53*[-/-];*Trp73*[sh] cells is confirmed in the RNA-Seq counts (Figure 3-7). Analysis results showed reduced expression of 140 genes in the *Trp53*[-/-];*Trp73*[sh] EBs compared to the *Trp53*[+/+];Ctrl[sh] EBs, whereas 154 genes were upregulated (Figure 3-7). Gene Ontology analysis of the down-regulated genes revealed significant enrichment for genes associated with gastrulation, mesoderm and endoderm formation (Figure 3-7). The down-regulated genes included the known Nodal-regulated mesendoderm genes *Eomes*, *Foxa2*, *Gsc, Mixl1*, *Gata6*, *Cxcr4* and *Fgf8* (Figure 3-7).

The down-regulated genes in *Trp53*[-/-];*Trp73*[sh] cells are also enriched with Activin response genes in human pluripotent stem cells differentiating towards the primitive streak state (Estaras et al., 2015; Funa et al., 2015). We wondered whether Activin pathway activation could rescue the expression of these genes. Addition of exogenous activin induced the expression of these genes in d3 *Trp53*[+/+];Ctrl[sh] EBs, but failed to do so in *Trp53*[-/-];*Trp73*[sh] EBs (Figure 3-8). *Nanog* and the Smad negative feedback regulators *Skil* and *Smad7* remained responsive to Activin in the absence of p53 and p73 (Figure 3-8). Notably, the p53/p73-depleted cells adopted a neuroectoderm fate, as determined by expression of *Pax6, Tubb3, Gbx2, Sox1* and *Onecut2* (Figure 3-7). This switch is consistent with adoption of an alternative fate upon Nodal inhibition (Smith et al., 2008; Vallier et al., 2004). These results suggest that p53/

p73-deficiency did not compromise the transcriptional activity of Activin signaling in the ES cells; instead, it selectively inhibited a subset of Activin response genes. p53 and p73 act redundantly to enable Nodal-dependent mesendoderm specification of ES cells. It remains to be investigated about the mechanisms that cause the bifurcation of Nodal target genes into p53/p73-dependent and p53/p73-independent genes.

Together with Qiong Wang, we investigated this effect in ES cells that were depleted of all members of the p53 family by triple knockout using CRISPR/Cas9 (Cong et al., 2013; Mali et al., 2013). High specificity sgRNAs targeting *Trp53*, *Trp63* and *Trp73* genomic loci were transiently transduced with *Cas9* into WT ES cells (Figure 3-9). One *Trp53*$^{+/+}$;*Trp63*$^{-/-}$;*Trp73*$^{-/-}$ (double knockout, DKO) clone and two *Trp53*$^{-/-}$;*Trp63*$^{-/-}$;*Trp73*$^{-/-}$ (p53/63/73 triple knockout, TKO) clones were selected and verified by sequencing and western blot analysis (Figure 3-9, 3-10). In TKO cells, p53, p63 and p73 proteins were depleted due to frameshift mutations in exon4 at the N-terminus (Figure 3-10). *T/Brachyury*, *Foxa2*, *Eomes* and *Gsc* expression was detected in p63/p73 DKO EBs but not in TKO EBs, whereas the expression dynamics of *Nanog* and *Skil* remained unaffected (Figure 3-11).

Collectively, our data obtained using two independent approaches suggest that the p53 family is required for mesendoderm specification of ES cells. All actively expressed p53 family members need to be depleted to block

mesendoderm differentiation defect regardless of under the p53-only monolayer condition or the p53/p73-expressing EB condition. Moreover, the genome-wide transcriptome analysis provides a mechanism whereby the p53 family functions through the modulation of Nodal/Activin signaling.
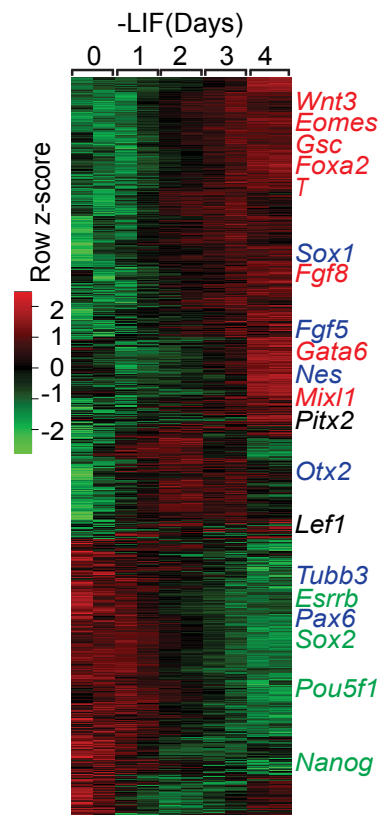
Figure 3-1.   Heatmap presentation of RNA-Seq transcriptomic analysis of ES cells and EBs at the indicated times after placement in LIF-free media. Row z scores of genes that were differentially expressed after d0 are plotted. Representative pluripotency-associated genes (*green*), early mesendoderm lineage marker genes (*red*), and early ectoderm marker genes (*blue*) are highlighted. Two biological replicates at each time point were analyzed.
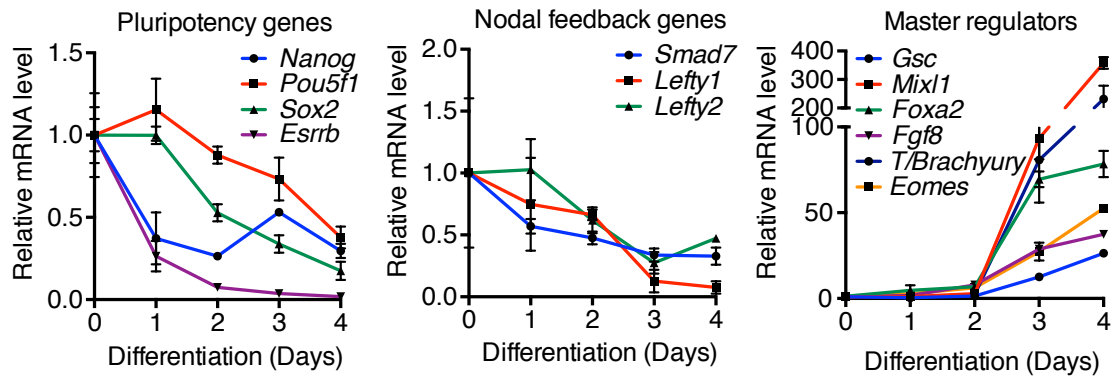
Figure 3-2.   qRT-PCR analysis of the mRNA expression dynamics of selected genes: pluripotency-associated factors *Nanog*, *Pou5f1*, *Sox2* and *Esrrb* (*left*), Nodal feedback genes *Smad7*, *Lefty1* and *Lefty2* (*middle*), and Master mesendoderm differentiation regulators *Gsc*, *Mixl1*, *Foxa2*, *Fgf8*, *T/Brachyury* and *Eomes* (*right*) during 4 day time course EB differentiation.
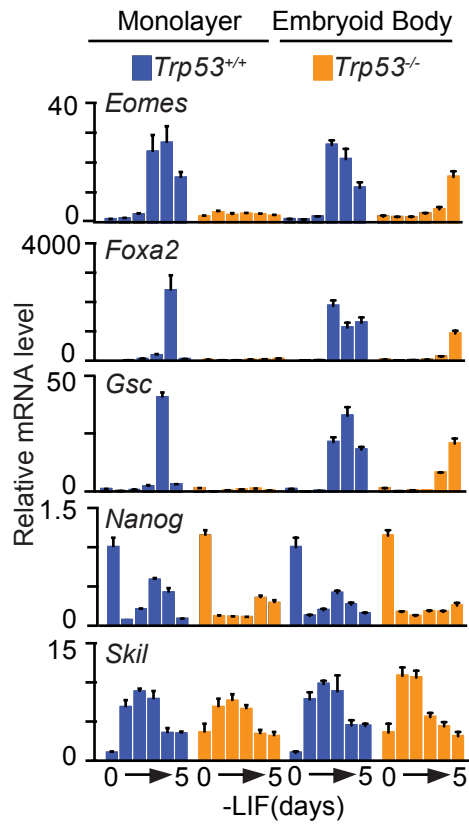
Figure 3-3.   qRT-PCR analysis of mRNA levels of the indicated genes in *Trp53*[+/+] and *Trp53*[-/-] cells under EB suspension or ES cell monolayer differentiation conditions. *n*=3 for each condition. Error bars represent s.e.m..
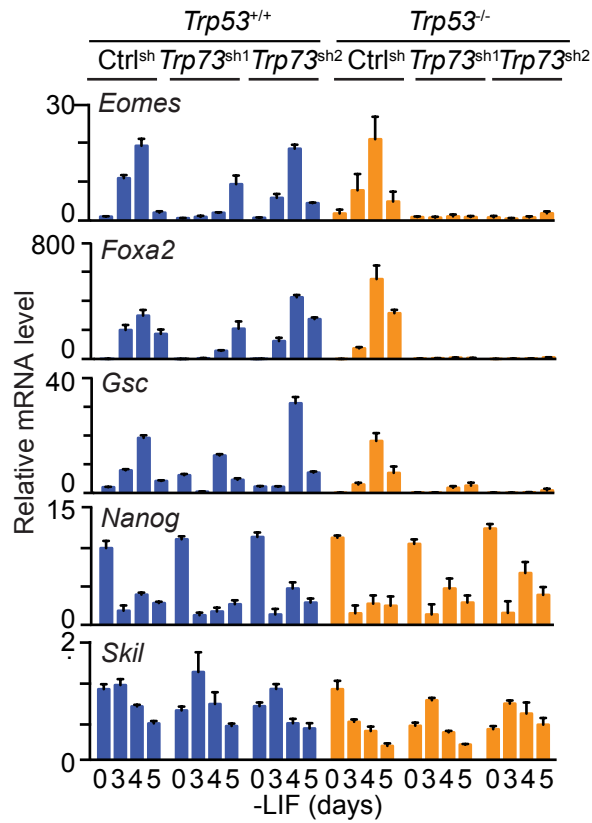
Figure 3-4. Levels of the indicated mRNAs during EB differentiation of *Trp53*[+/+] or *Trp53*[-/-] cells that were transduced with shRNA vectors targeting control (*luciferase*) or *Trp73* mRNA. Two independent shRNA vectors were used for targeting *Trp73* mRNA. Error bars represent s.e.m.. Experiment was performed in triplicate and a representative result is presented. (Courtesy of Qiong Wang, PhD)

A

mRNA level of p53 family members (RPM)

-LIF(Days)	D0	D1	D2	D3	D4
Trp53	432.17	549.29	330.37	260.36	299.97
Trp63	0.07	0.18	0.08	0.14	0.16
Trp73	15.78	26.88	29.87	17.77	5.72
Gapdh	150.60	159.20	219.83	256.60	191.32

B
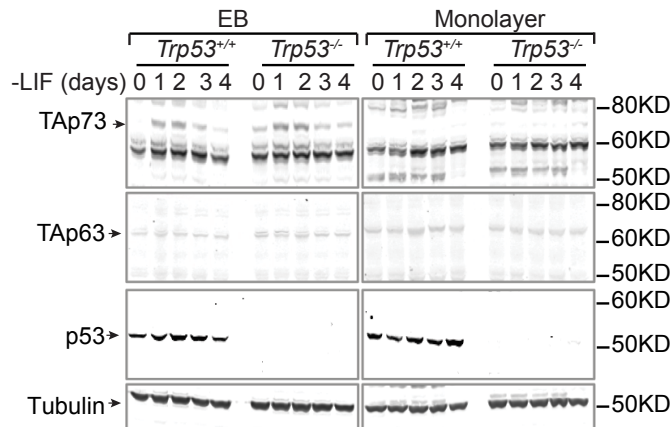
Figure 3-5. Relative expression of the p53 family members in monolayer ES cells or EBs. (**A**) mRNA expression levels (RNA-Seq) of *Trp53*, *Trp63*, *Trp73* and *Gapdh* in d0 to d4 EBs. Read counts were normalized to reads per million reads (RPM). Mean counts of two biological replicates are presented. (**B**) Western immunoblotting analysis for p53, p63 and p73 protein levels in *Trp53*[+/+] or *Trp53*[-/-] cells over 4 days in EB or monolayer differentiation conditions. Tubulin was used as loading control. Experiments were performed in triplicate and representative results are presented.
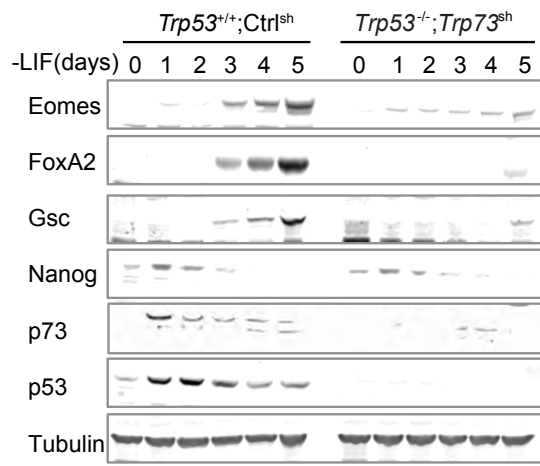
Figure 3-6.   Western immunobloting analysis of indicated proteins in control (*Trp53*$^{+/+}$;Ctrl^{sh}) and p53/p73-depleted (*Trp53*$^{-/-}$;*Trp73*^{sh}) cells over 5 days under EB differentiation conditions. Tubulin was used as loading controls. Experiment was performed in triplicate and a representative result is presented.

A

Trp53[+/+];Ctrl[sh] vs Trp53[-/-];Trp73[sh]
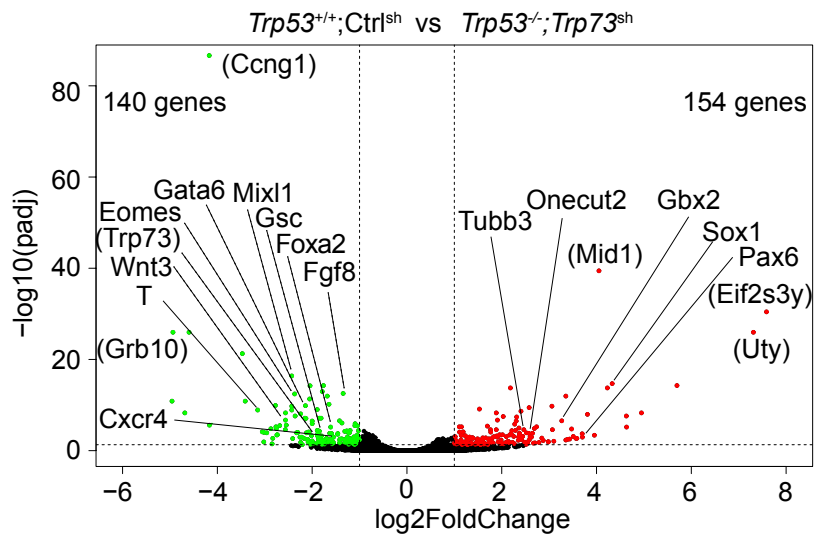
B

Top GO Biological Processes	p-Value
Germ layer formation	9.90E-08
p53 signaling pathway	1.50E-06
Mesodermal cell migration	2.17E-04
Endoderm formation	4.14E-04
Axon extension	1.08E-03
Anterior-Posterior specification	2.40E-03

Figure 3-7. RNA-Seq analysis of day-4 EBs derived from wild type and p53/p73-depleted cells. (**A**) Volcano plot of RNA-Seq transcriptomic data sets of day-4 EBs derived from control (Trp53[+/+];Ctrl[sh]) and p53/p73-depleted (Trp53[-/-];Trp73[sh]) cells. *green*: genes down-regulated in p53/p73-depleted cells (fold change <0.5, p < 0.05); *red*: genes up-regulated in p53/p73-depleted cells (fold change >2, p < 0.05). Lineage specification genes for mesendoderm or ectoderm are indicated. Genes in parentheses are significantly differentially expressed but not known to have germ layer specification functions. Two biological replicates for each condition were analyzed. (**B**) Gene Ontology analysis of the genes that were down-regulated in Trp53[-/-];Trp73[sh] cells comparing to Trp53[+/+];Ctrl[sh] cells. The most significantly enriched Biological Process GO terms with are listed with the corresponding p values.
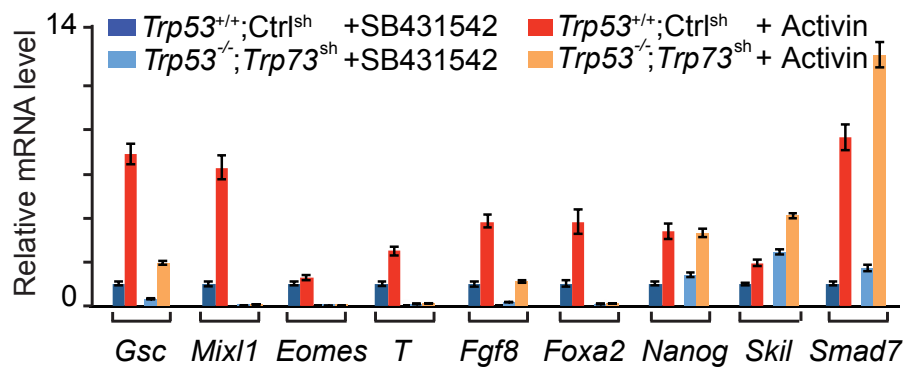
Figure 3-8. mRNA levels of mesendoderm marker genes (*Gsc*, *Mixl1*, *T*, *Eomes*, *Fgf8* and *Foxa2*), pluripotency-associated gene (*Nanog*) and Nodal feedback genes (*Skil* and *Smad7*) in d3 EBs that were treated with SB431542 or Activin A (Activin) for 2 h. Error bars represent s.e.m.. Experiment was performed in triplicate and a representative result is presented. (Courtesy of Qiong Wang, PhD)

A **p53/p63/p73 knockout by CRISPR/Cas9**

$Trp53^{+/+};Trp63^{+/+};Trp73^{+/+}$ (WT) mESC

$Trp63$ exon4 sgRNAs
$+Trp73$ exon4 sgRNAs

$Trp53^{+/+};Trp63^{-/-};Trp73^{-/-}$ (DKO) clone

$Trp53$ exon4 sgRNAs

$Trp53^{-/-};Trp63^{-/-};Trp73^{-/-}$ (TKO) clones #1 & 2

B
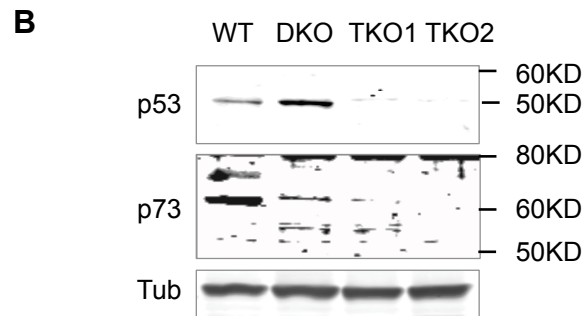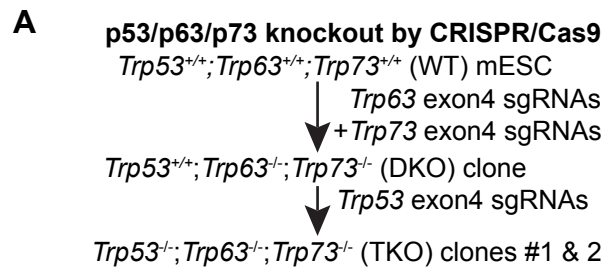
Figure 3-9.   CRISPR/Cas9-mediated p53, p63, p73 triple knockout and validation. (**A**) Scheme for the strategy of CRISPR-mediated *Trp53/Trp63/Trp73* knockout. sgRNAs were designed to target the exon4 of each gene locus. (**B**) Western immunoblotting analysis for p53 and p73 in wild type (WT) or mutant clones (DKO, TKO#1, TKO#2) under pluripotent conditions. p63 protein expression is beyond the detection limit of immunoblotting. Tubulin was used as loading control.

Figure 3-10. CRISPR sgRNA sequences and the genomic sequences of *Trp63* exon4 locus (**A**), *Trp73* exon4 locus (**B**) and *Trp53* exon4 locus (**C**) in mutant clones DKO, TKO#1 and TKO#2.
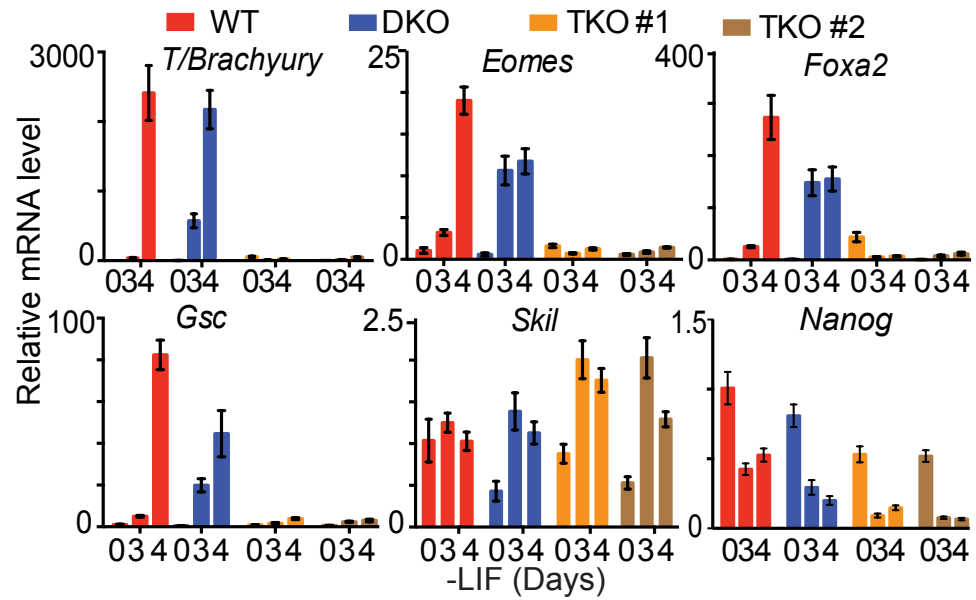
Figure 3-11. mRNA expression analysis (qRT-PCR) of mesendoderm marker genes *T/Brachyury*, *Eomes*, *Foxa2, Gsc*, Nodal feedback gene *Skil*, and pluripotency-associated gene *Nanog* in WT, DKO, and TKO#1 and #2 cells during EB differentiation. Error bars represent s.e.m. (Courtesy of Qiong Wang, PhD)

3.3.2   p53/p63/p73 triple knockout leads to early defects in mouse embryo development

p53 null mice are viable while p53-depletion is embryonic lethal in *Xenopus* embryos (Cordenonsi et al., 2003; Donehower et al., 1992; Takebayashi-Suzuki et al., 2003), which lack TAp63 and TAp73. *In vitro* differentiation assays suggest that TAp63 and TAp73 are redundant factors for p53 in mesendoderm differentiation. To extend these observations *in vivo*, I determined the expression levels of p53 family members by qRT-PCR in mouse embryos ranging from embryonic day E6.5 until E11.5, corresponding to stages of gastrulation and extensive organogenesis. p53 was expressed at E6.5 and thereafter until E10.5 (Figure 3-12). Transactivating isoforms of p63 (TAp63α-Y) and dominant negative isoforms of p63 (ΔNp63α-Y) were below the qRT-PCR detection limit in E6.5-E7.5 embryos but were highly expressed from E8.5 (Figure 3-12). Transactivating isoforms of p73 (TAp73α) and dominant negative isoforms of p73 (ΔNp73α and ζ) were moderately expressed at E6.5, followed by a decrease between E7.5-E8.5 and re-elevation after E9.5 (Figure 3-12). Thus in contrast to *Xenopus* embryos that only express p53, mouse embryos dynamically express all three members of p53 family during gastrulation, raising the possibility of redundant functionality.

Together with Sonja Nowotschin, PhD, Anna-Katerina Hadjantonakis, PhD, and Sang Yong Kim, PhD, we micro-injected wild type (WT), DKO or TKO

ES cells labeled with mCherry into wild type blastocysts to generate chimeric embryos. Chimeras were dissected at E10.5, corresponding to midgestation, and analyzed for the contribution of mCherry+ cells to major organs (Figure 3-13). Regardless of the level of chimerism, which was classified as strong, intermediate or weak according to mCherry fluorescence, all 16 WT control chimeras and another 8 DKO chimeras proceeded to midgestation and exhibited normal morphology (Figure 3-13) (Flores et al., 2002). By contrast, 4 chimeras with strong or intermediate TKO cell contribution exhibited profound retardation and resembled E8.0 embryos, whereas weak TKO cell contribution permitted development to slightly later stages (Figure 3-13).

Based on the morphology of embryos recovered at E10.5 we estimated that the onset of a defect occurred at around E7.5-E8.5, corresponding to the time of gastrulation when mesoderm and definitive endoderm germ layers are formed. In support of this possibility, at the late headfold to early somite stage (~E7.75), 6 out of 6 strong TKO chimeras exhibited morphological defects symptomatic of defective gastrulation (Figure 3-13). At E8.5 (6-10 somite stage), 11 out of 18 TKO chimeras exhibited defects in the primitive streak region at the posterior end of the embryo, a kinked neural tube, a paucity of adjacent mesoderm, cardiac defects, as well as defects in the headfold region (Figure 3-13).

We performed serial sections to determine whether TKO cells fail to

contribute to specific regions of the embryo. Notably, E8.75 TKO chimeras exhibited an abnormal primitive streak and gut endoderm morphology, and a round neural tube rather than a flat neural plate observed in WT chimeras (Figure 3-14). Moreover, TKO cells exhibited elevated contribution to ectopic neural protrusions only observed in TKO chimeras, but reduced contribution to gut endoderm, confirming that p53/p63/p73-depletion inhibits endoderm differentiation and favors a neuroectodermal fate (Figure 3-14). E7.25 TKO chimeras exhibited an accumulation of cells at the primitive streak, which bulged into the amniotic cavity, indicative of a failure in gastrulation (Figure 3-14) (Migeotte et al., 2011; Sun et al., 1999; Viotti et al., 2014). These results suggested that strong TKO chimeras failed to execute normal gastrulation.

To verify the lineage-specific defects, we performed immunofluorescence analysis of FoxA2, a nascent mesendoderm marker, and T/Brachyury, a primitive streak, mesoderm and midline marker (Monaghan et al., 1993; Wilkinson et al., 1990). FoxA2 expression was interrupted in the midline of E7.75 TKO chimeras (Figure 3-15), reminiscent of defects observed in mutants of *Lhx1,* a downstream target of Nodal-Smad signaling (Costello et al., 2015). In E7.25 chimeras mCherry+ TKO cells failed to induce FoxA2 expression whereas surrounding wild type cells expressed high levels of FoxA2 (Figure 3-15), reinforcing the conclusion that TKO cells were unable to trigger mesendoderm specification. Though T/Brachyury expression was not abolished in TKO cells of E8.75 chimeras, T/Brachyury positive cells were aberrantly organized at the primitive

streak and midline structures (Figure 3-14). Taken together, our analyses of chimeric embryos strongly suggest that the p53 family plays a critical role in mesendoderm specification.
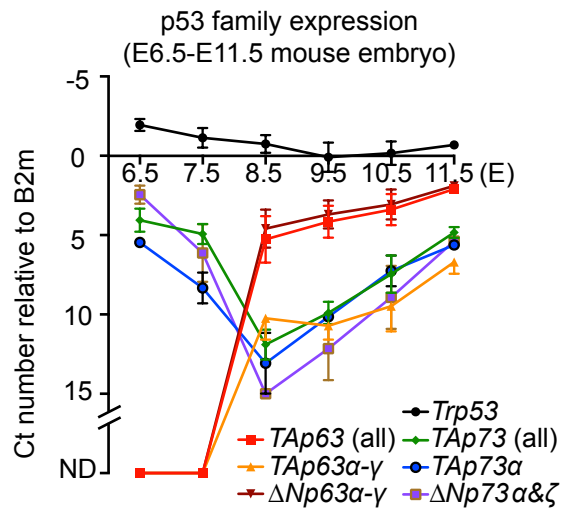
Figure 3-12. mRNA expression levels (qRT-PCR) of p53 family isoforms in E6.5-E11.5 mouse embryos. Y-axis represents qPCR Delta Ct number relative to *B2m*. *n*=3. ND: not detectable, level below the sensitivity limit of qPCR.
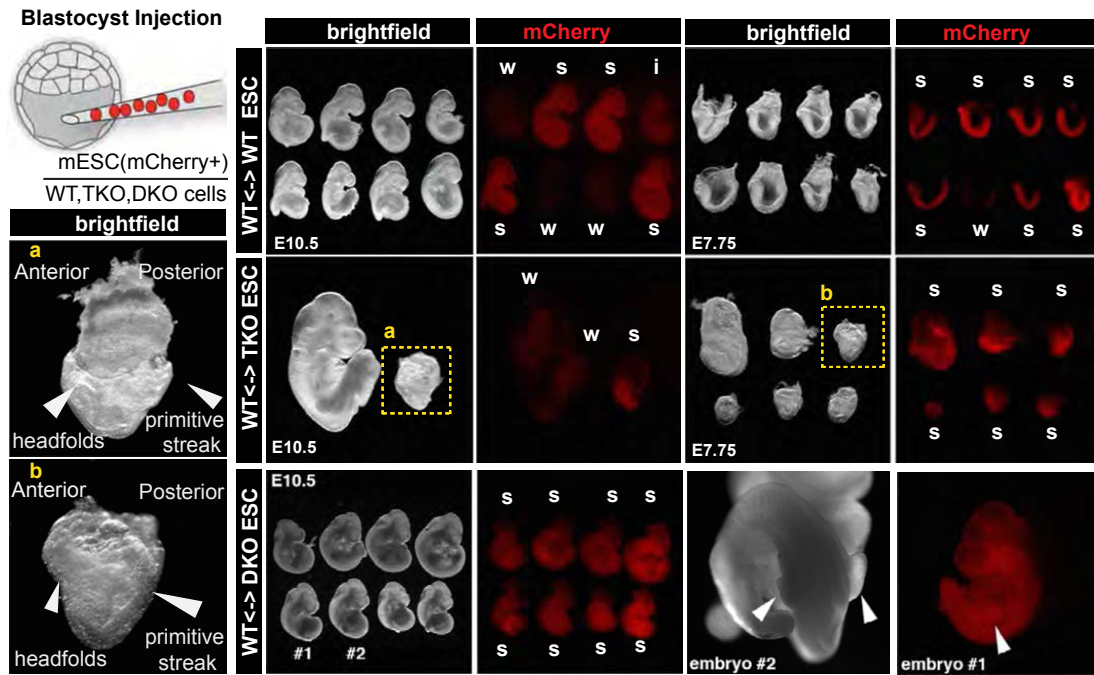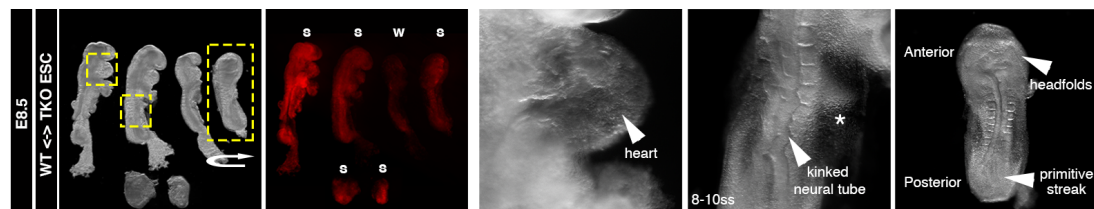
Figure 3-13. (**A**) Scheme of blastocyst injection, brightfield and red fluorescence (mCherry) images of embryo chimeras, comprising control (WT), $Trp53^{-/-}$;$Trp63^{-/-}$;$Trp73^{-/-}$ triple KO (TKO), and $Trp63^{-/-}$;$Trp73^{-/-}$ double KO (DKO) ES cells, dissected at E10.5 and E7.75. Level of chimerism of each embryo is classified as strong (s), intermediate (i) or weak (s) based on mCherry fluorescence. Panels a and b depict high magnification bright field images of strong TKO chimeric embryos recovered at each stage, but arresting around the same time, highlighting developmental delay and defects in the primitive streak and headfold regions. (**B**) Brightfield and red fluorescence (mCherry) images of embryo chimeras comprising $Trp53^{-/-}$;$Trp63^{-/-}$;$Trp73^{-/-}$ triple KO (TKO) ES cells dissected at E8.5. Panels on the right depict high magnification bright field images highlighting developmental delay and aberrant morphology, including defects in the heart and neural tube. Arrowheads, related embryonic structures.
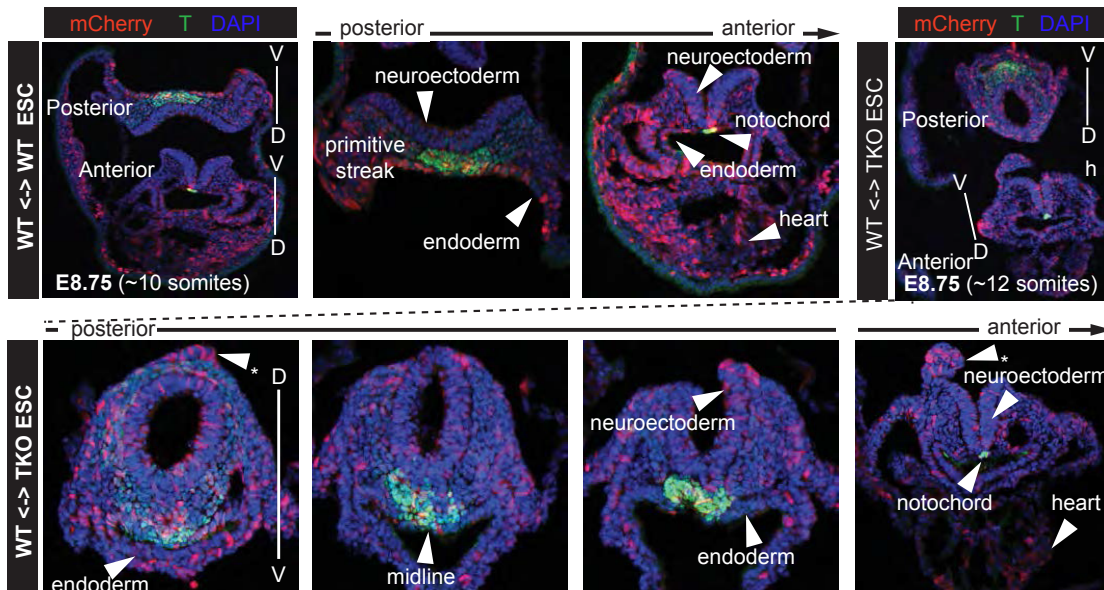
Figure 3-14. Transverse sections and T/Brachyury expression analysis in TKO chimeras at E8.75. 3D reconstructions of z-stacks of confocal images of transverse sections of control and TKO ES cell containing embryo chimeras recovered at E8.75 (~10 somite stage wild-type) depicting mCherry (*red*, ES cell descendants), T/Brachyury expression (*green*) and nuclear (stained by DAPI, *blue*) localization. By contrast to control chimeras which exhibited normal morphology, and having uniform and extensive distribution of mCherry cells, TKO ES cell chimeras exhibited morphological defects including an abnormally round neural tube, dorsal neuroectodermal protrusions (asterisks), non-uniform ventral midline, and flattened gut endoderm exhibiting an absence of mCherry cells.
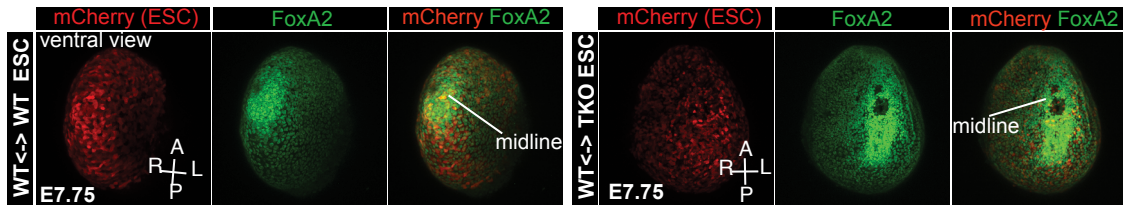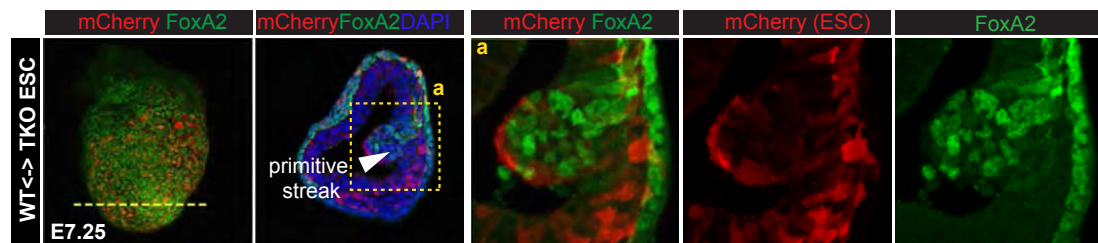
Figure 3-15. Immunofluorescence analysis of FoxA2 expression in TKO chimeras. (**A**) 3D reconstructed views of whole mount embryo chimeras comprising control (WT) or TKO ES cells dissected at E7.75 and imaged for mCherry (*red*, ES cell descendants) and FoxA2 expression (*green*). TKO ES cell chimeras exhibit an interrupted midline (*right*), which is consistent with a defect at gastrulation. Note that the unaffected FoxA2-positive cell population on the embryo's surface is visceral endoderm, which is not pluripotent epiblast-derived. (**B**) Confocal microscope images taken of serial sectioned embryo chimeras comprising control and TKO ES cells dissected at E7.25 depicting mCherry (*red*, ES cell descendants), FoxA2 expression (*green*) and nuclear (stained by DAPI, *blue*) localization. TKO ES cell chimeras exhibit an accumulation of cells at the primitive streak and protrusion into the amniotic cavity (*left*), which is consistent with a defect at gastrulation. FoxA2 is expressed in wild type but not mCherry-positive TKO mutant cells within the expanded primitive streak protrusion.

3.3.3 The p53 family selectively enables Smad binding to mesendoderm genes

The preceding results argue strongly that p53 family is indeed involved in mouse development and act redundantly. However, how the activity of this family is coordinated with other active developmental pathways at the same stage remains unknown. To investigate the mechanistic basis for the role of p53 in the activation of mesendoderm specification genes, we considered previous reports that p53 interacts directly with Smad2/3 on target promoters (Cordenonsi et al., 2003; Takebayashi-Suzuki et al., 2003). However, no interaction of p53 family members with Smads 2, 3 or 4 in ES cells or EBs was observed by immunoblotting or mass spectrometry analysis of Smad-associated proteins (data not shown). With Qiong Wang we performed p53 and Smad2/3 ChIP-Seq analysis in EBs that were treated with activin or SB431542. Genome-wide, the p53 binding sites were enriched for the p53 consensus binding sequence (Figure 3-17), and included the typical p53 target loci *Mdm2* and *Cdkn1a* (Figure 3-18). However, my genome-wide analysis revealed that the binding patterns of p53 and Smad2/3 did not overlap (Figure 3-16). I noticed that there was no significant p53 binding within 100kb centered on the *Eomes*, *Foxa2*, *Gsc* or *Mixl1* transcription start sites (TSS) (Figure 3-18), regions that contained multiple Smad2/3 binding sites (Figure 3-19). Moreover, Smad2/3 binding to chromatin was stimulated by Activin whereas p53 binding was not (Figure 3-16, 3-18, 3-19). I concluded that p53 acts as a determinant of Nodal action in ES cells without

physically contacting Smad target loci.

Together with Qiong Wang, I investigated the effect of p53 on the binding of Smad2/3 to target enhancers. To avoid potential clonal variability introduced by CRISPR/Cas9 selection, the *Trp53*[-/-];*Trp73*[sh] and *Trp53*[+/+];Ctrl[sh] EBs were utilized. Activin stimulated the binding of Smad2/3 to sites in *Eomes*, *Foxa2*, *Gsc*, *Mixl1* and *Smad7* in control d3 EBs (Figure 3-19). I revealed that p53/p73-depletion inhibited Smad2/3 binding to a subset of these sites in the differentiation genes but not to sites in *Smad7* (Figure 3-19). Based on global analysis, I classified Smad2/3 binding peaks as p53/p73-dependent and p53/p73-independent peaks (Figure 3-20). Loci with p53/p73-independent peaks included the Nodal feedback regulators *Smad7*, *Tdgf1 (Cripto)* and *Nodal* (Figure 3-20), and loci with p53/ p73-dependent peaks were enriched for mesendoderm specification genes (Figure 3-20). Notably, *Gsc* is an exception that Smad2/3 binding to its +6kb enhancer is p53/p73-dependent, whereas binding to the proximal promoter is p53/p73-independent (Figure 3-19).
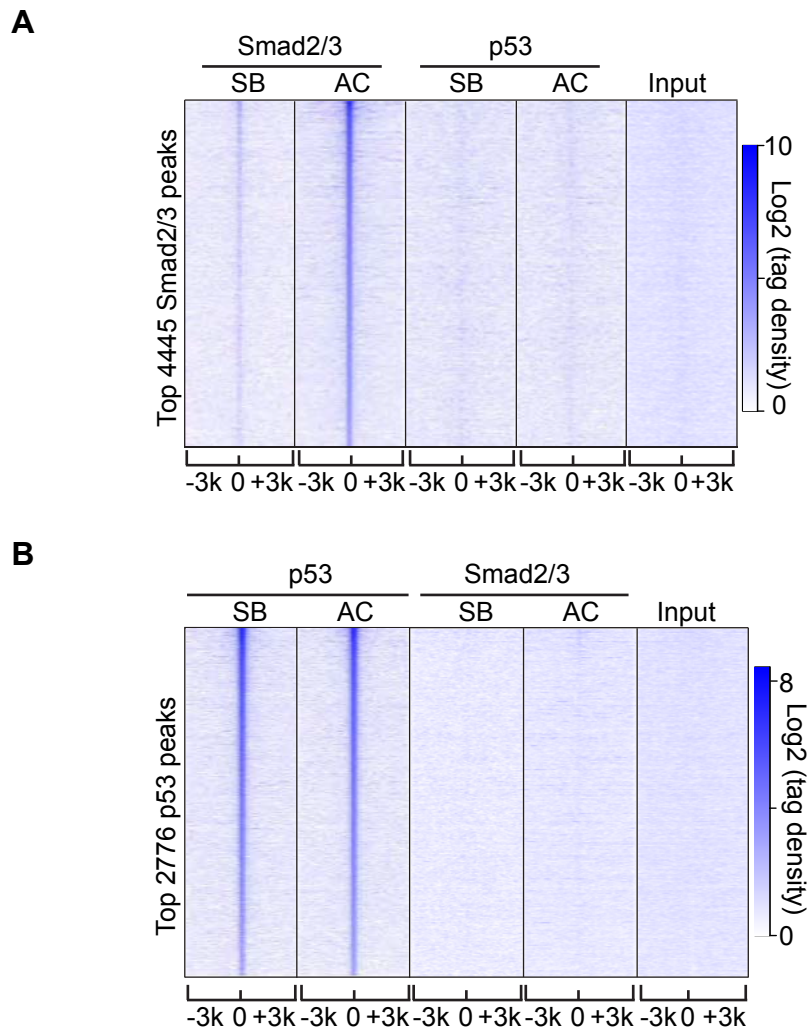
Figure 3-16. Comparison of Smad2/3 and p53 genome-wide ChIP-Seq occupancy. (**A**) Heatmap of ChIP-Seq tag densities for Smad2/3 and p53 within -3kb/+3kb genomic regions surrounding the centers of 4445 high-confidence Smad2/3 binding sites in wild type d3 EBs treated with SB431542 (SB) or activin (AC). Color key represents log2 tag density distribution. (**B**) Heatmap of ChIP-Seq tag densities for Smad2/3 and p53 within -3kb/+3kb genomic regions surrounding the centers of 2776 high-confidence p53 binding sites in wild type d3 EBs treated with SB431542 (SB) or activin (AC). Color key represents log2 tag density distribution.

103

Motif enrichment in p53 ChIP-Seq: EB-AC		
Motif	**factor**	**logP-value**
	p53/p63/p73	-3.111e+03
	Tcfcp2l1	-5.590e+01

Figure 3-17. Enriched transcription factor binding motifs in p53 bound regions in day-3 EBs treated with Activin A for 2h.
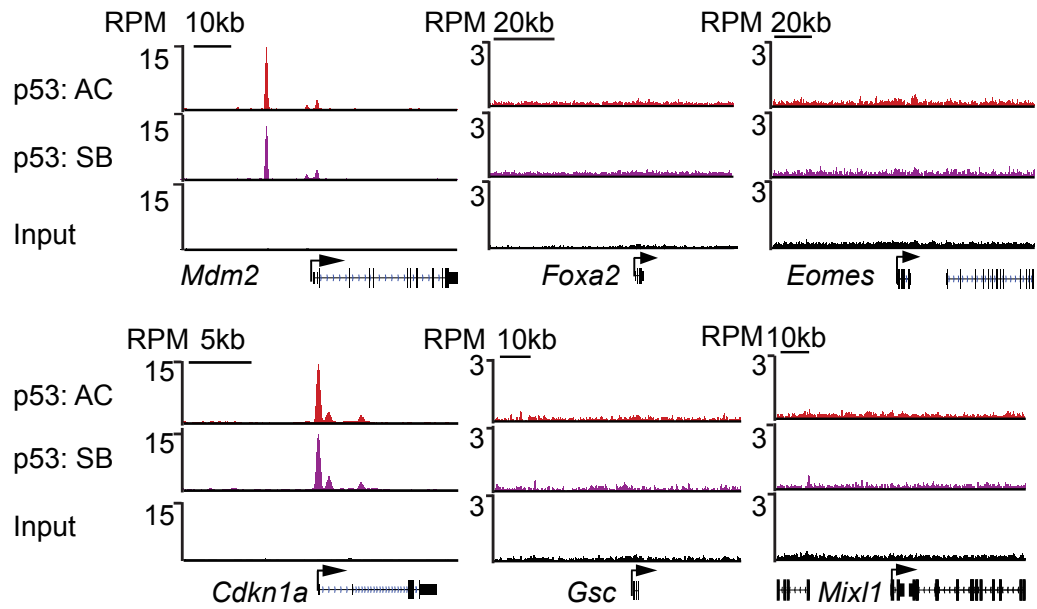
Figure 3-18. Gene track view for p53 ChIP-Seq data or input control at the *Mdm2*, *Cdkn1a* (p21Cip), *Foxa2*, *Eomes*, *Gsc* and *Mixl1* loci. Day-3 EBs were treated with SB431542 (SB) or Activin A (AC) for 2 h. Gene bodies are schematically represented at the bottom of each track set.

A
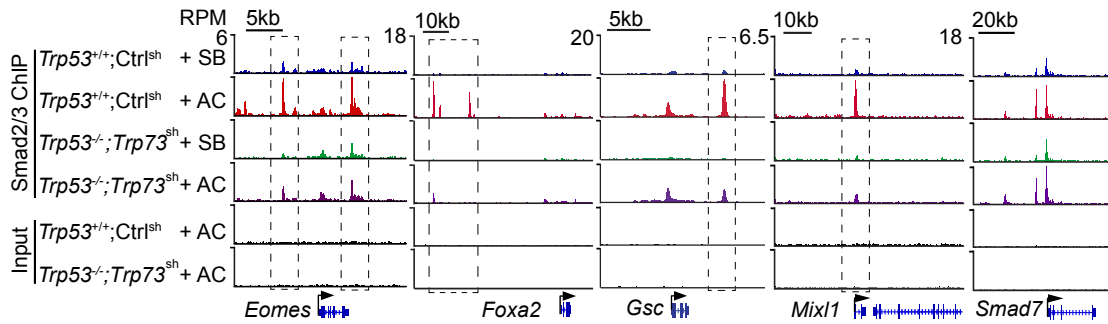
B
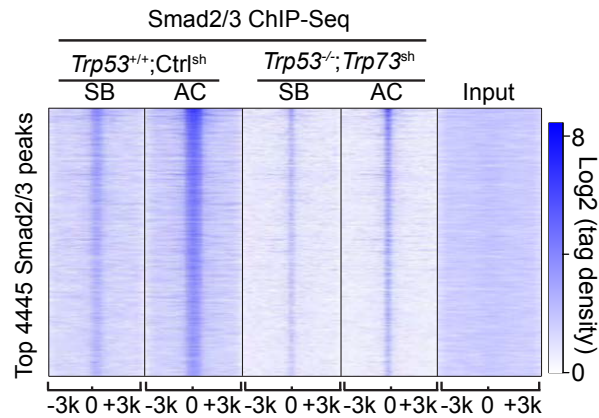
Figure 3-19. Smad2/3 ChIP-Seq analysis in wild type and p53/p73-depleted cells. (**A**) Gene track view for Smad2/3 ChIP-Seq data in the indicated loci, cell lines, and conditions. Gene bodies are schematically represented at the bottom of each track set. Dashed boxes indicate p53/p73-dependent Smad2/3 binding sites. (**B**) Heatmap of Smad2/3 tag densities within -3kb/+3kb genomic regions surrounding centers of 4445 high-confidence Smad2/3 binding sites, in control (*Trp53*[+/+];Ctrl[sh]) or p53/p73-depleted (*Trp53*[-/-];*Trp73*[sh]) d3 EBs that were treated for 2 h with SB431542 (SB) or activin (AC). Color key represents log2 tag density distribution.

A

B
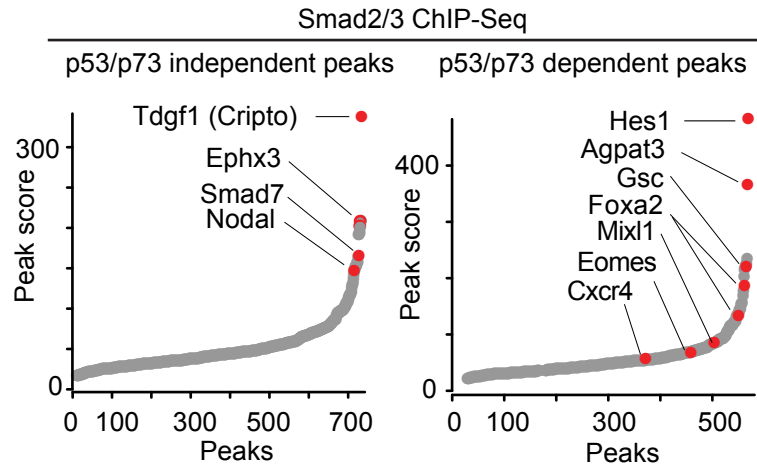
Figure 3-20. (**A**) Peak score plot for the most significant p53/p73-independent (*left*) and p53/p73-dependent (*right*) Smad2/3 binding sites (fold change >2.0 and p value <0.01). Genes adjacent to the highest intensity peaks are indicated. (**B**) Enriched known transcription factor binding motifs in p53/p73-independent Smad2/3 binding sites (*left*) and p53/p73-dependent Smad2/3 binding sites (*right*).

3.3.4 The p53 family controls Tcf-sensitive Nodal target genes

My transcription factor motif analysis revealed that both the p53/ p73-dependent and independent Smad binding regions are rich in binding elements for Smad (SBEs) and the Smad partners FoxH1 (Chen et al., 1996; Chen et al., 1997; Liu et al., 1997) and HEB (Yoon et al., 2011). Interestingly, a Tcf3/4 binding motif (TBE) was significantly enriched only in p53-dependent Smad2/3 binding sites and not in p53-independent sites (Figure 3-20). SBE (CAGAC/T) and TBE (TCAAAG) motifs are located approximately within ~50 bp of each other in the regulatory regions of *Eomes*, *Foxa2*, *Gsc*, *Mixl1*, *Fgf8*, and *T/ Brachyury* (Figure 3-21, 3-22, 3-23).

I used CRISPR/Cas9 to generate focal deletions of neighboring SBEs and TBEs in the *Eomes* -10kb enhancer and the *Gsc* +6kb enhancer in ES cells (Figure 3-21, 3-22). These deletions caused a significant reduction in *Eomes* or *Gsc* induction by endogenous signals or exogenous activin (Figure 3-21, 3-22). *Smad7* induction by activin remained intact, arguing that the mutant clones were capable of Nodal/Activin signal transduction (Figure 3-21). Collectively, these findings suggest that p53 regulates a component that is selectively required for activation of Wnt-responsive Nodal target genes.

The Tcf family in mammals includes four members: Lef1, Tcf1 (Tcf7), Tcf3 (Tcf7l1) and Tcf4 (Tcf7l2). ES cells predominantly express Tcf3 and Tcf1 (Figure

3-24). Tcf3 is essential for embryo development (Merrill et al., 2004) and ES cell exit from pluripotency (Leeb et al., 2014; Pereira et al., 2006; Wray et al., 2011). ChIP analysis showed that Tcf3 is the most highly enriched Tcf family member in the *Eomes* -10kb enhancer, *Foxa2* -50kb enhancer and *Gsc* +6kb enhancer. Tcf3 binding was further stimulated by Activin (Figure 3-24). Tcf1 and Tcf3 were highly enriched in the *Axin2* promoter, a known Wnt/Tcf target gene (Jho et al., 2002; Lustig et al., 2002; Yan et al., 2001) (Figure 3-24). Moreover, Tcf3 ChIP-Seq analysis showed a high concordance with Smad2/3 binding sites at mesendoderm gene enhancers (Figure 3-25). The binding of Tcf3 and Smad2/3 to these enhancers increased significantly in differentiating EBs compared to pluripotent ES cells (Figure 3-25). 70% (3131/4445) of Smad2/3 binding sites genome-wide exhibited Tcf3 co-occupancy, revealing an extensive Smad2/3-Tcf3 cooperation on the chromatin (Figure 3-25). Collectively these results suggest that Smad2/3 and Tcf3 converge on *cis*-regulatory elements of p53-dependent mesendoderm specification genes.
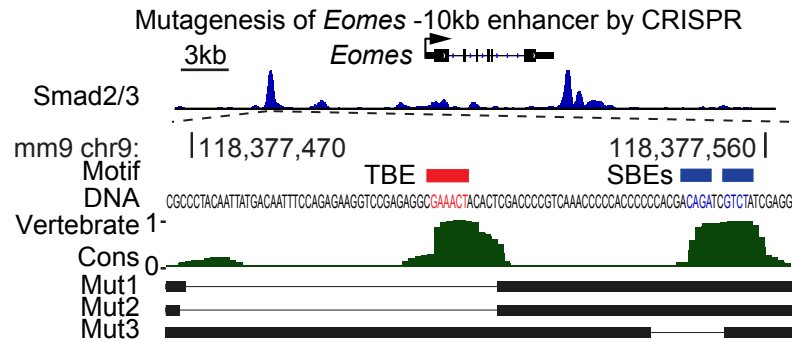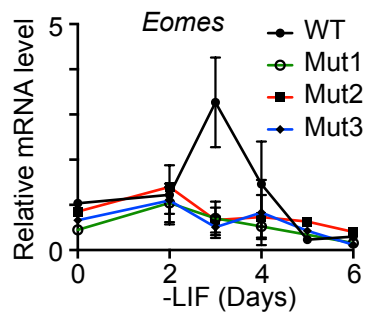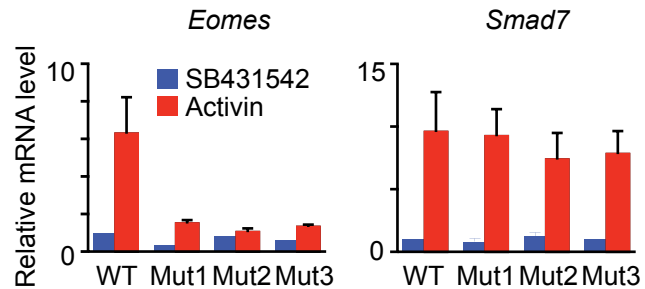
Figure 3-21. (**A**) Scheme of CRISPR/Cas9-mediated mutagenesis of the *Eomes* -10kb distal enhancer region. Gene track view for Smad2/3 ChIP-Seq tag density at *Eomes* locus at the top panel highlights the targeted region. The DNA sequence of targeted region and the positions of Smad binding elements (CAGAC/T) (SBE, *blue*) and Tcf binding element (TCAAAG) (TBE, *red*) are indicated. Middle panel, evolutionary conservation scores in vertebrates. Bottom panels, black horizontal bars represent the remaining genomic regions in mutant (Mut1-3) clones whereas the thin lines indicate deleted regions. (**B**) qRT-PCR analysis of *Eomes* mRNA expression in d0 to d6 EBs derived from WT or *Eomes* -10kb enhancer mutant clones. Gene expression level is normalized to d0 WT samples. *n*=3. Error bar represents s.e.m.. (**C**) qRT-PCR analysis of *Eomes* and *Smad7* expression in Activin or SB431542 treated d3 EBs derived from WT or *Eomes* -10kb enhancer mutant cells. *n*=3. Error bar represents s.e.m..
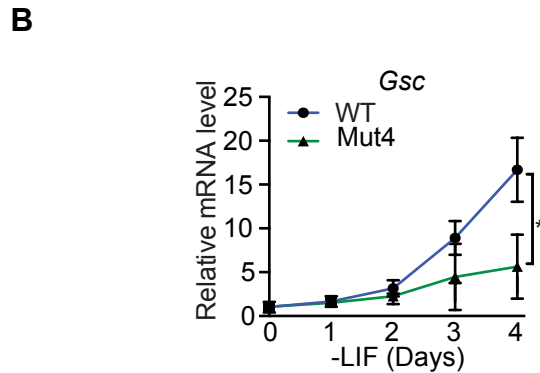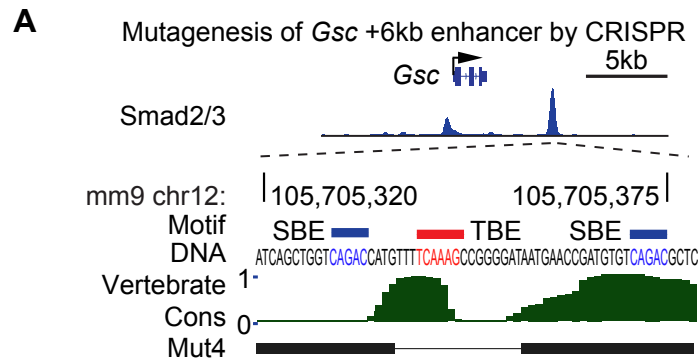
Figure 3-22. (**A**) Scheme of the CRISPR/Cas9-mediated mutagenesis of the *Gsc* +6kb distal enhancer region. Each panel is illustrated similarly as Figure 3-21A. (**B**) qRT-PCR analysis of *Gsc* mRNA expression in wild type (WT) or *Gsc* +6kb distal enhancer mutant clone. Gene expression level is normalized to d0 WT samples. *n*=3. Error bars represent s.e.m.. * *p*<0.05, Mann-Whitney test.
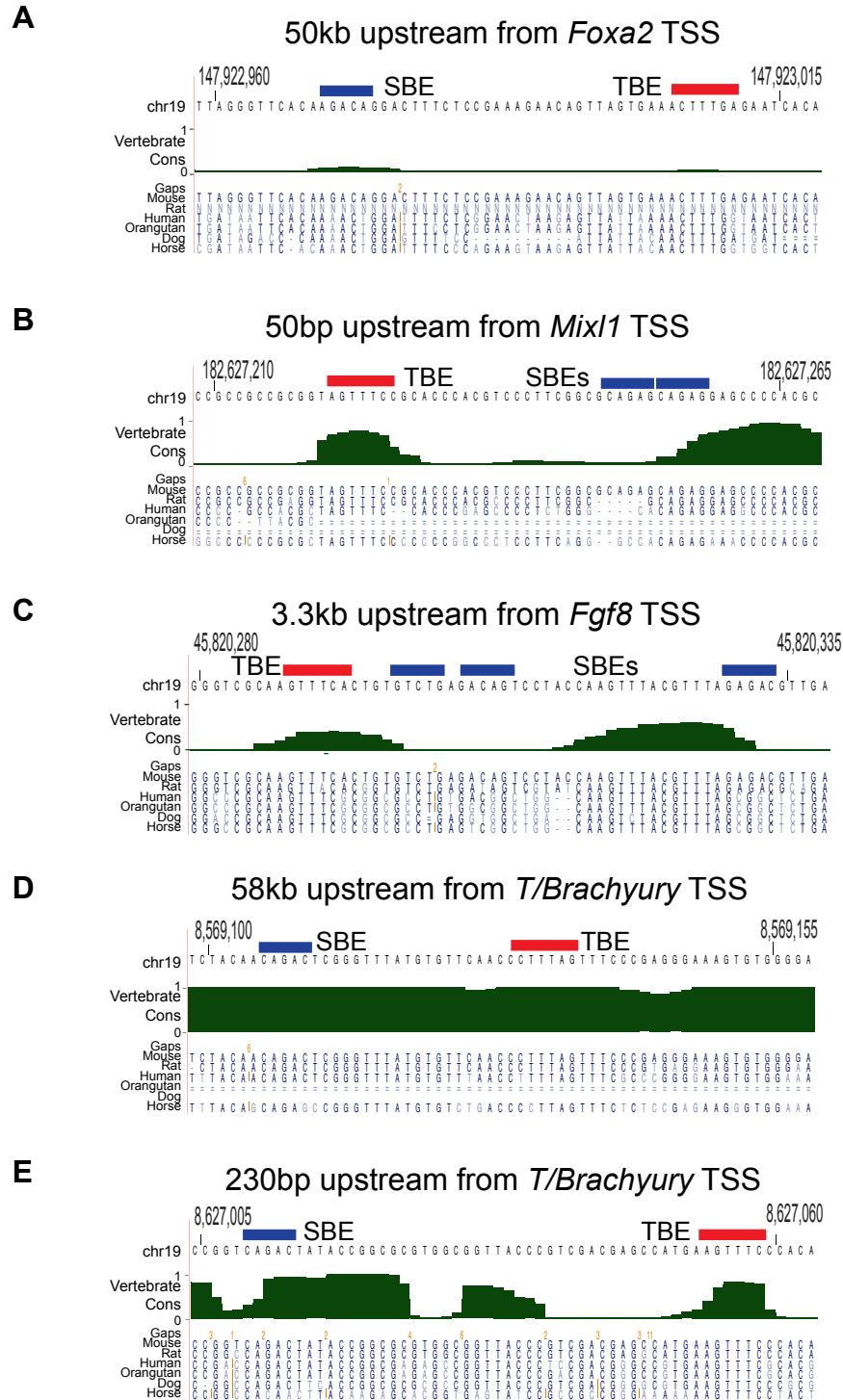
Figure 3-23. Evolutionary conservation (Vertebrate Conservation Score) and genomic sequence alignments of regions that are 50kb upstream from *Foxa2* transcription start site (TSS) (**A**), 50bp upstream from *Mixl1* TSS (**B**), 3.3kb upstream from *Fgf8* TSS (**C**), 58kb upstream from *T/Brachyury* TSS (**D**) and 230bp upstream from *T/Brachyury* TSS (**E**).
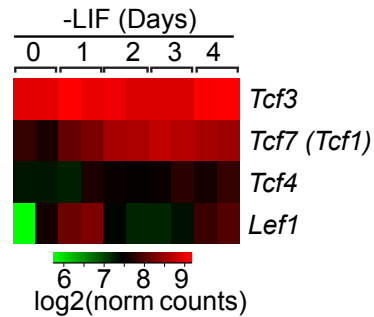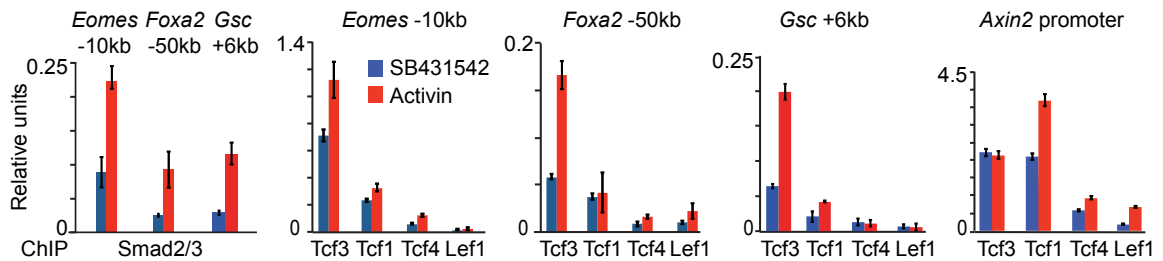
Figure 3-24. mRNA expression and chromatin binding analysis of Tcf/Lef family members. (**A**) Heatmap of Tcf family members *Tcf3*, *Tcf7* (*Tcf1*), *Tcf4*, *Lef1* mRNA expression levels (RNA-Seq) from d0 to d4 of EB differentiation. Scale represents the log2 normalized read counts. (**B**) Analysis of Smad2/3, Lef1, Tcf1 (Tcf7), Tcf3, and Tcf4 binding (ChIP-qPCR) to the *Eomes* -10kb, *Foxa2* -50kb, *Gsc* +6kb enhancers, and *Axin2* proximal promoter in d3 EBs treated with SB431542 or activin for 2 h. Error bars represent s.e.m. (Figure **B**: Courtesy of Qiong Wang, PhD)

A
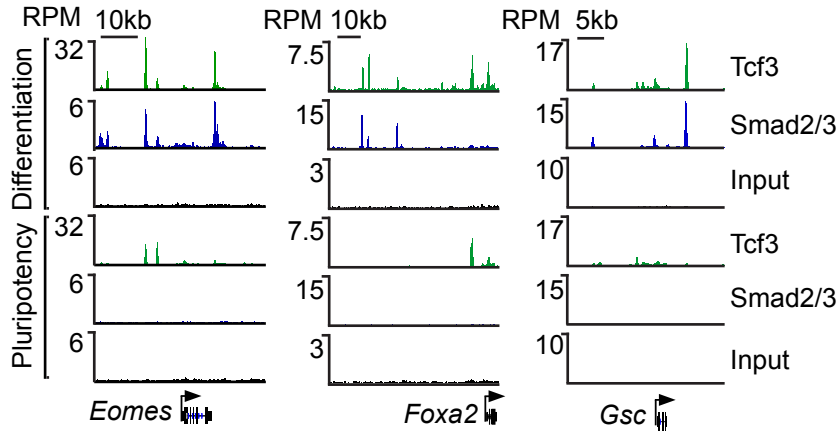
B
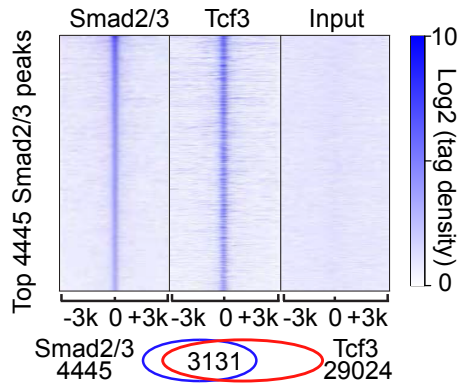
Figure 3-25. Comparison between Smad2/3 and Tcf3 genome-wide occupancies. (**A**) Gene track view of *Eomes*, *Foxa2* and *Gsc* locus. Tcf3 and Smad2/3 ChIP-Seq were performed in Activin-treated d3 EBs (Differentiation), and in d0 non-treated mouse ES cells (Pluripotency). Pre-cleared chromatin prior to primary antibody addition (Input) was used as negative control. Tag densities were normalized to reads per million reads (RPM). The gene structures from RefSeq are schematically represented at the bottom. (**B**) Heatmap of tag densities of ChIP-Seq data sets from the indicated conditions, within -3kb/+3kb genomic regions surrounding centers of 4445 high-confidence Smad2/3 binding sites. ChIP-Seq data for the indicated proteins and input control are from Activin treated D3 EBs. Color key represents log2 tag density distribution.

3.3.5 Wnt expression is rate limiting for the onset of mesendoderm differentiation

Together with Qiong Wang, we revealed that Smad2/3 and Tcf3 binding to the *Eomes*, *Foxa2* or *Gsc* enhancers in EBs did not occur until d3 (Figure 3-26), suggesting the existence of a rate-limiting factor. In search of such factor, I examined the expression of the corresponding ligands of Nodal and Wnt pathways during EB differentiation. Nodal was expressed throughout this period (Figure 3-27, 3-29), and d0 cells, as well as d3, cells responded to Activin with formation of Smad2/3-Smad4 and Smad2/3-Trim33 transcriptional complexes (Figure 3-39) (Xi et al., 2011). However, *Wnt8a* was the only Wnt family member expressed before d3 (Figure 3-27). *Wnt3,* an essential factor for mouse primitive streak formation and gastrulation (Liu et al., 1999b), was expressed starting on d3, which coincided with the onset of *Foxa2, Gsc* and *Mixl1* induction and peak *Eomes* expression (Figure 3-27). Consistent with the expression kinetics *in vitro*, Wnt3 expression in the mouse embryo is not detectable in the early epiblast, starts at the posterior visceral endoderm of E5.5 embryos and expands to the adjacent epiblast tissue a few hours later (Kemp et al., 2005; Rivera-Perez and Magnuson, 2005).

Qiong Wang used different Wnt pathway inhibitors to assess the role of endogenous Wnt signaling in mesendoderm gene expression. Addition of the Wnt receptor inhibitor Dkk1 (Niida et al., 2004), the Wnt palmitoylation

inhibitor IWP-2 (Anastas and Moon, 2013), or the Axin2 stabilizing drug XAV939 (Anastas and Moon, 2013) prevented the binding of Tcf3 to *Eomes*, *Foxa2* and *Gsc* enhancers (Figure 3-28), and the activation of these genes (Figure 3-28). Notably, Wnt inhibitors not only diminished the binding of Tcf3 to these enhancers but also that of Smad2/3 (Figure 3-28). Conversely, Nodal/Activin receptor inhibition diminished the binding of Tcf3 as well as Smad2/3 to these sites (Figure 3-28). In contrast, *Smad7* responded to Activin regardless of Wnt inhibition.

To determine whether Tcf3 chromatin binding requires canonical β-Catenin signaling in this context, we performed β-Catenin ChIP assays. Lacking suitable anti-β-catenin antibodies, I used CRISPR/Cas9 to engineer tandem FLAG and HA epitope tags into the N-terminus of *Ctnnb1* (β-Catenin) in ES cells (β-Cat[F-H] cells), and validated this modification by sequencing and western blot (Figure 3-30). Epitope tagging did not significantly alter the differentiation kinetics in β-Cat[F-H] cells (Figure 3-30). Anti-HA ChIP confirmed β-Catenin binding to *Eomes*, *Foxa2* and *Gsc* enhancers, and the binding pattern was highly concordant with Tcf3 and Smad2/3 binding (Figure 3-30). Collectively, these results suggested that the Nodal and Wnt pathways coordinately drive Smad2/3 and Tcf3 to common target enhancers in mesendoderm specification genes, and the expression of autocrine Wnt, as well as the activity of the p53 family, are rate-limiting for this process.
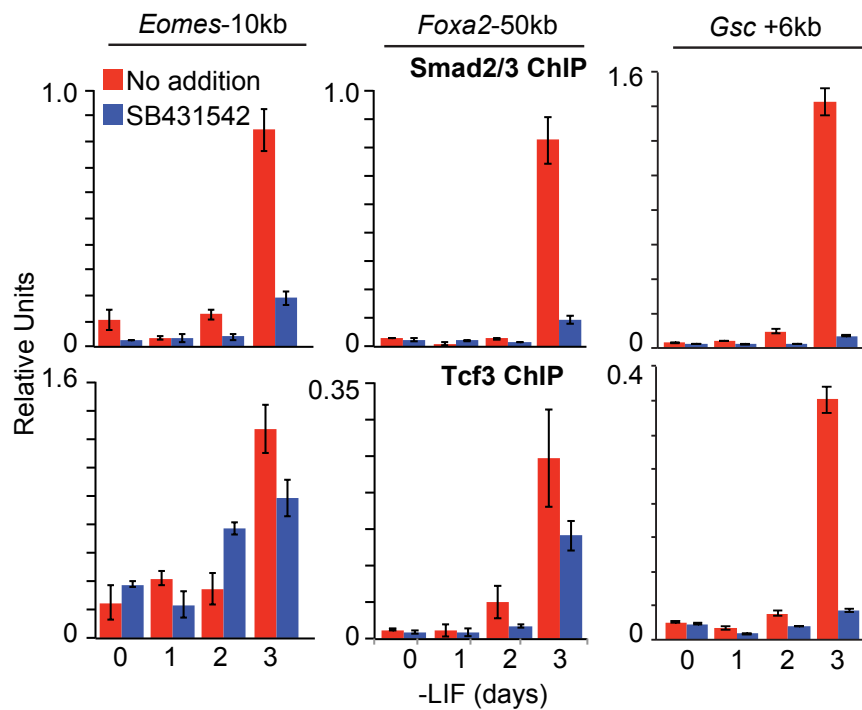
Figure 3-26. Time course ChIP-qPCR analysis of Smad2/3 and Tcf3 binding to the *Eomes* -10kb, *Foxa2* -50kb and *Gsc* +6kb enhancers of EBs at the indicated times in LIF-free culture media. SB431542 was added as indicated to inhibit signaling by endogenous Nodal. Error bars represent s.e.m.. Experiment was performed in triplicate and a representative result is presented. (Courtesy of Qiong Wang, PhD)
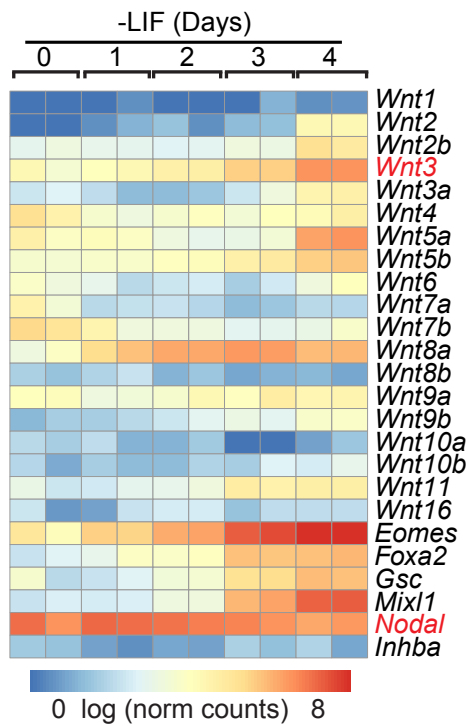
Figure 3-27. Heatmap of Wnt family mRNA expression levels (RNA-Seq) during d0 to d4 EB differentiation. Scale represents the log2 normalized read counts. *Eomes*, *Foxa2*, *Gsc*, *Mixl1*, *Nodal* and *Inhba* (encoding Activin A) expression levels are also presented for reference. Two biological replicates were analyzed for each condition.
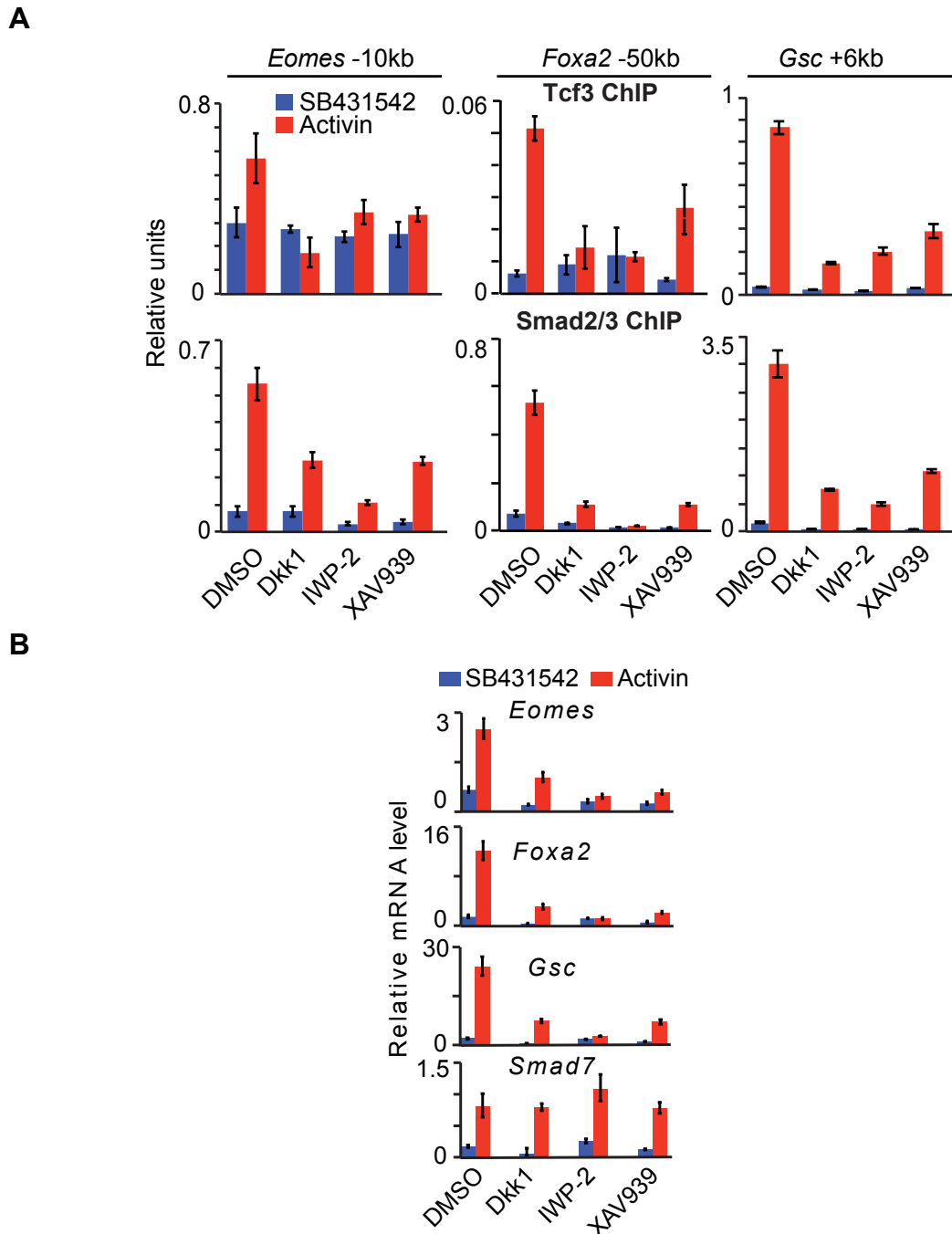
Figure 3-28. (**A**) Analysis of Tcf3 and Smad2/3 binding (ChIP-qPCR) to the *Eomes* -10kb, *Foxa2* -50kb and *Gsc* +6kb and enhancers in d3 EBs treated with SB431542 or Activin, and with DMSO or Wnt signaling inhibitors Dkk1, IWP-2 or XAV939 as indicated. (**B**) Analysis of *Eomes*, *Foxa2*, *Gsc* and *Smad7* mRNA levels (qRT-PCR) in d3 EBs that were incubated with SB431542 or Activin for 2 h. Wnt signaling inhibitors Dkk1, IWP-2 or XAV939, or DMSO vehicle were added as indicated. (Courtesy of Qiong Wang, PhD)

119

A
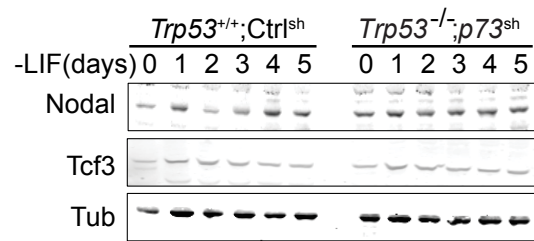
B
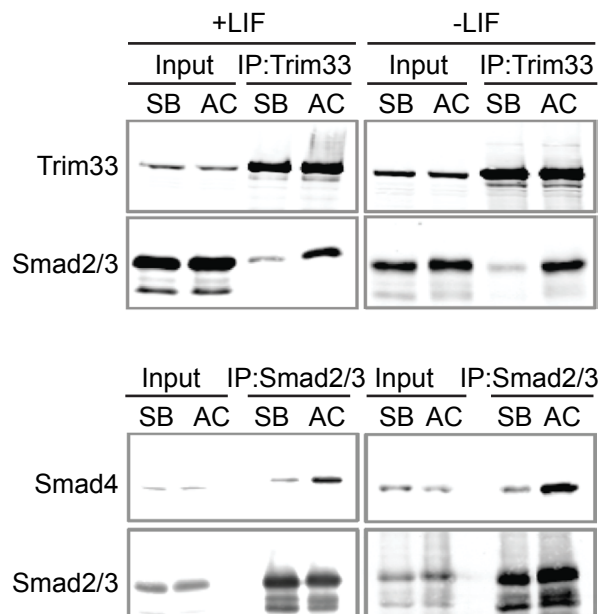
Figure 3-29. Expression and activity of Nodal in ES cells and EBs. (**A**) Western immunoblotting analysis of Nodal and Tcf3 in control (*Trp53*^{+/+};Ctrl^{sh}) and p53/p73-depleted (*Trp53*^{-/-};*Trp73*^{sh}) cells over 5 days under EB differentiation conditions. Tubulin (Tub) was used as loading control. (**B**) Western immunoblotting analysis of Activin-induced complexes for the immunoprecipitation between Smad2/3, Smad4 and Trim33 in ES (+LIF) and day-3 EB (-LIF) cells. Cells were treated with Activin A (AC) or SB431542 (SB) for 2 h. 2% input was used as control. (Courtesy of Qiong Wang, PhD)
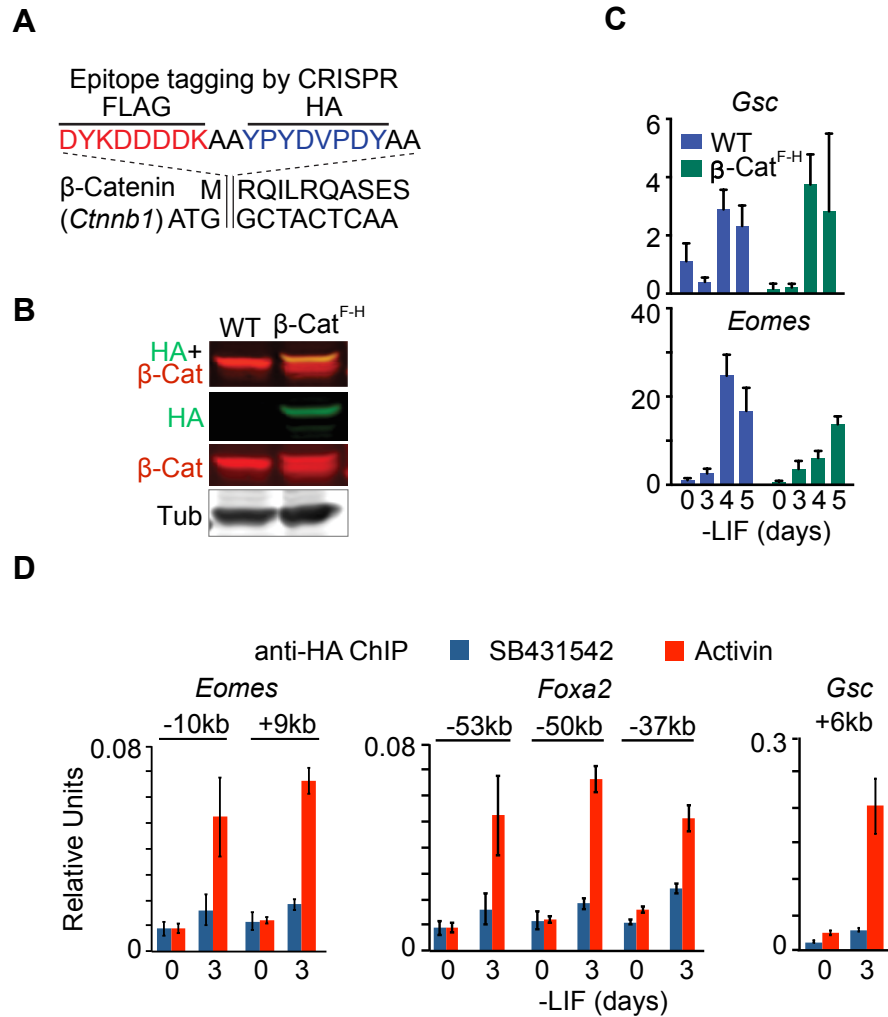
Figure 3-30. β-Catenin chromatin binding analysis in ES cells and EBs. (**A**) Scheme of CRISPR/Cas9 mediated FLAG-HA epitope tagging for *Ctnnb1*/β-Catenin N-terminus. (**B**) Western immunoblot analysis for β-Catenin[FLAG-HA] (β-Cat[F-H]) cell line that was inserted with FLAG-HA tags at *Ctnnb1* N-terminus. (**C**) *Gsc* and *Eomes* mRNA levels (qRT-PCR) in EBs derived from wild type (WT) or β-Cat[F-H] cell lines. (**D**) Anti-HA ChIP-qPCR analysis at *Eomes* -10kb and +9kb, *Foxa2* -53kb, -50kb and -37kb, and *Gsc* +6kb enhancers. Anti-HA ChIP-qPCR analysis at *Eomes* -10kb and +9kb, *Foxa2* -53kb, -50kb and -37kb, and *Gsc* +6kb enhancers.

3.3.6   The p53 family controls Wnt3 expression

Preceding results suggested that Wnt3 up-regulation is a rate-limiting event during mesendoderm specification. However, what is driving the dramatic induction of Wnt3 as ES cells exit pluripotency remains unknown. In search of a link between p53 and Wnt in an unbiased manner, I sorted the 114 genes that are differentially expressed in $Trp53^{-/-}$;$Trp73^{sh}$ EBs relative to $Trp53^{+/+}$;control^{sh} EBs (fold change >3; mean counts >100 in $Trp53^{+/+}$;control^{sh} EBs; refer to Figure 3-7). Within this group, 24 genes (Figure 3-31) showed bound p53 within 50kb from the TSS (refer to Figure 3-16). Notably, these 24 genes included *Wnt3* and its receptor *Fzd1* (Figure 3-31). Fzd1 can be activated by Wnt3 and Wnt3a, and to a less extent Wnt2 but not other Wnt factors. p53 bound near the TSS in *Wnt3* (Figure 3-32) and at +45kb in *Fzd1* (Figure 3-32). Both sites contain one or two repeats of p53 consensus binding sequence RRRCWWGYYY (Figure 3-32). ChIP-qPCR analysis of ectopically expressed FLAG-tagged human *TAp63* and HA-tagged human *TAp73* in $Trp53^{-/-}$;$Trp73^{sh}$ cells showed that both p63 and p73 can bind to this site (Figure 3-33). Importantly, *Wnt3* expression after LIF removal was inhibited by depletion of p53 and p73 in EBs (Figure 3-34). These results indicate that the p53 family directly controls Wnt3 expression.

p53 activity at this crucial juncture of signal integration is regulated by Aurka. In pluripotent stem cells, Aurora kinase A (Aurka) phosphorylates p53 at Ser212 (Serine 215 in human p53), which inhibits the activity and stability

of p53 protein. The repression of p53 by Aurka is relieved with the onset of differentiation as Aurka protein level dramatically reduces during this process (Lee et al., 2012). In addition to affecting protein stability, Ser212 phosphorylation also alters the DNA binding and transactivation capacity of p53 protein (Liu et al., 2004). In normal cells, the Aurka-p53 regulatory axis is modulated by cell cycle regulators.

Together with Qiong Wang, we investigated whether p63/73 were regulated in a similar manner as p53 (Lee et al., 2012). Expression of Aurka in ES cells depends on LIF (Lee et al., 2012). In agreement, Aurka levels in EBs dropped below detection by d3 after LIF removal (Figure 3-34). Sequence alignment revealed that the reported Aurka R(H/Q)S phosphorylation motif in p53 (Ferrari et al., 2005; Lee et al., 2012) is also conserved in p63 and p73 (Figure 3-36). We transduced $Trp53^{-/-}$;$Trp73^{sh}$ cells with vectors encoding p53 proteins with mutations that would prevent phosphorylation of this serine residue (Ser to Ala mutants p53^{S212A}, TAp63^{S285A} and TAp73^{S235A}) or mimic it (Ser to Asp mutants p53^{S212D}, TAp63^{S285D} and TAp73^{S235D}) (Lee et al., 2012). We tested the ability of these mutants to rescue Smad2/3 binding to the *Gsc*, *Foxa2* and *Eomes* enhancers in $Trp53^{-/-}$;$Trp73^{sh}$ cells. p53^{SA} but not p53^{SD} rescued the binding (Figure 3-35). Moreover, the Ala mutant p53, p63 and p73 but not the Asp mutants rescued the expression of *Wnt3* and *Gsc* (Figure 3-36). These results suggest that p53, p63 and p73 share the ability to trigger Wnt3 expression upon release from inhibition as ES cells exit pluripotency.

p53/p73-dependent, p53 bound genes

Gene	Fold Change	p-value
Phlda3	30.7	1.07E-26
Ccng1	18.0	2.21E-87
Cdkn1a	10.7	1.36E-11
Itga8	8.9	1.17E-09
Eda2r	8.1	1.17E-02
Inpp5d	6.8	6.27E-06
Ddit4l	5.9	5.49E-09
Zfp365	5.9	2.06E-07
Sp5	5.0	1.20E-04
Foxj1	4.9	9.69E-05
Wnt3	4.7	8.64E-09
Slc19a2	4.5	7.48E-08
Fst	4.4	1.17E-04
Ak1	4.1	5.96E-15
Cpt1c	3.6	7.22E-08
Fzd1	*3.6*	5.59E-05
Ass1	3.5	7.95E-08
Arap2	3.5	4.73E-03
B230120H23Rik	3.4	1.14E-13
Sulf2	3.4	5.37E-15
Dennd2c	3.2	1.08E-12
Zmat3	3.1	6.94E-11
Cox6b2	3.1	3.85E-03
Slc4a11	3.0	6.96E-06

Figure 3-31. Table representing the top 24 genes that were differentially expressed (fold change >3, p<0.05, mean normalized read counts > 100) in response to p53/p73-depletion by RNA-Seq, and exhibited significant p53-binding with 50kb from TSS. Genes are ranked by their fold change in RNA-Seq analysis.

A

RPM 10kb
1.3
p53

Input

Wnt3

B

p53 ChIP-Seq 1.3 500bp *Wnt3*
(Activin) 0
Vertebrate 1
conservation 0
Gaps 1 1
Mouse GGAACTCGCAGAAGCTGT**GACAAGCCC**AG
Human GGAACGCGCAGGAGCCAT**GACAAGCCC**AG
Dog GGAACGCGCAGGAGCCAC**GACAAGCCC**AG
Horse GGAACGCGCGAAGCCGT**GACAAGCCC**AG
p53 binding sequence

C

RPM *Fzd1* 10kb
2.2 *
p53

Input

*GCCA**GAACATGCCTGGGCAGGCCT**GTCA
p53 binding sequence
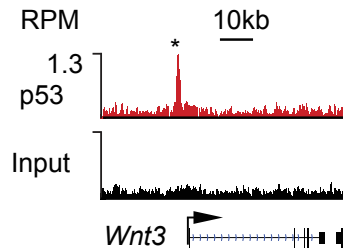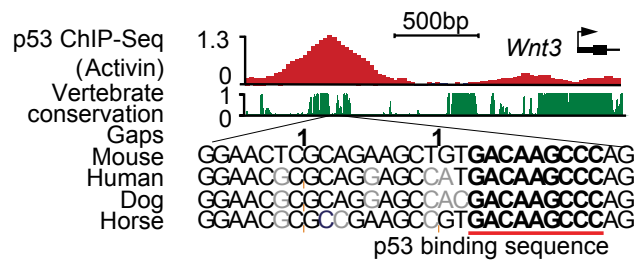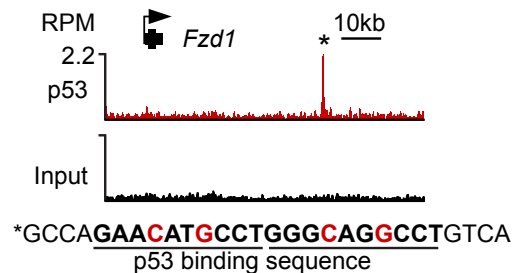
Figure 3-32. p53 ChIP-Seq analysis at the *Wnt3* and *Fzd1* gene loci.(**A**) Gene track view for p53 ChIP-Seq data or input control at the *Wnt3* locus. RefSeq gene body is schematically represented at the bottom of each track set. (**B**) Evolutionary conservation score plot for the p53-binding region near the *Wnt3* promoter, and DNA sequence alignment of the peak center in multiple species highlighting the p53 consensus binding sequence. (**C**) Gene track view for p53 ChIP-Seq data or input control at the *Fzd1* locus. RefSeq gene body for *Fzd1* is schematically represented at the top of tracks. Asterisk (*) marks the p53-binding site at the putative *Fzd1* enhancer. Genomic sequence for the center of this site is listed at the bottom and the p53 consensus binding sequence is highlighted.

125
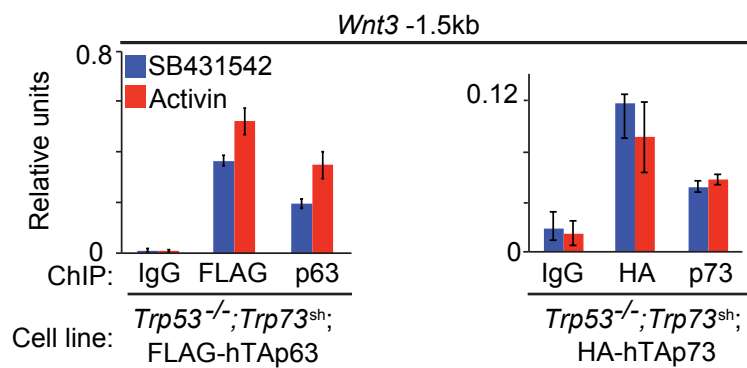
Figure 3-33. ChIP-qPCR analysis of exogenous FLAG-hTAp63 and HA-hTAp73 binding at *Wnt3* -1.5kb enhancer region in *Trp53*[-/-];*Trp73*[sh] day 4 EB treated with SB431542 or Activin for 2 h. Antibodies that recognizes FLAG, HA tags, or specifically human TAp63 and TAp73 were used for ChIP. Non-specific IgG were used as negative control. Error bars represent s.e.m.. Experiment was performed in triplicate and representative result is presented. (Courtesy of Qiong Wang, PhD)
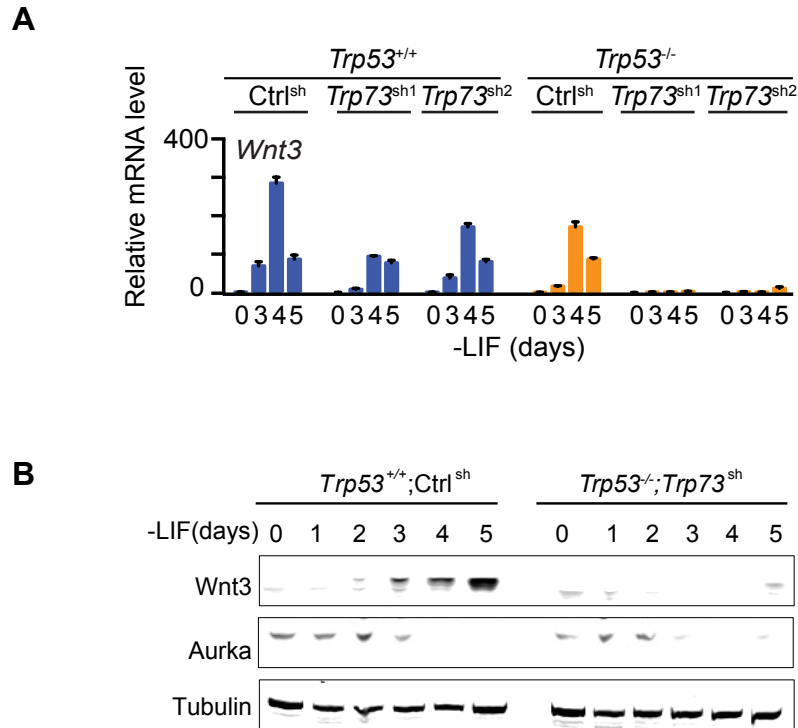
Figure 3-34. (**A**) qRT-PCR analysis of *Wnt3* mRNA expression in EBs derived from *Trp53*^{+/+};Ctrl^{sh} and *Trp53*^{-/-};*Trp73*^{sh} ES cells at indicated time points. (**B**) Western immunoblotting analysis of Wnt3 expression in EBs derived from *Trp53*^{+/+};Ctrl^{sh} and *Trp53*^{-/-};*Trp73*^{sh} ES cells at indicated time points. Experiments were performed in triplicate and a representative result is presented. Tubulin was used as loading control. (Courtesy of Qiong Wang, PhD)
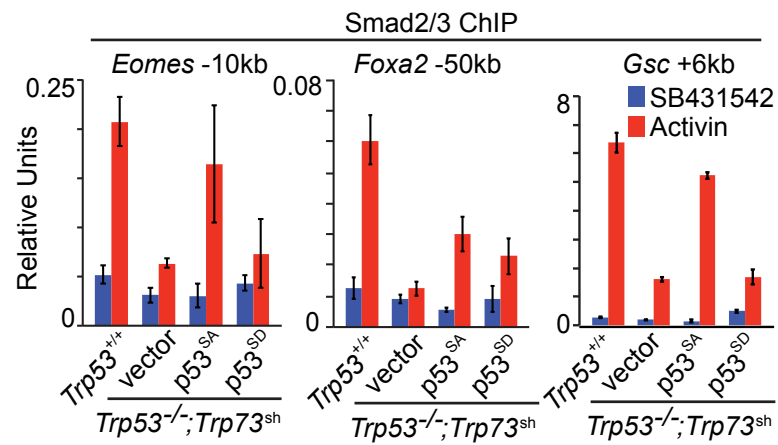
Figure 3-35. Analysis of Smad2/3 binding (ChIP-qPCR) to the *Gsc* +6kb enhancer in d3 EBs from *Trp53*[+/+];Ctrl[sh] or *Trp53*[-/-];*Trp73*[sh] ES cells transduced with p53[SA] or p53[SD], followed by 2h SB431542 or Activin treatment. Error bars represent s.e.m.. (Courtesy of Qiong Wang, PhD)

Figure 3-36. (**A**) Amino acid sequence alignment for p53/TAp63/TAp73 proteins from representative vertebrates: zebrafish, *Xenopus*, mouse and human. Red line, conserved R(H/Q)S Aurka phosphorylation motif. Ser212 in p53 and the corresponding residues in other family members are marked in red and bold. (**B**) Western immunoblotting analysis of Wnt3 and p53 family protein levels in EBs from control ES cells (*Trp53*[+/+];Ctrl[sh]), or p53/p73-depleted cells (*Trp53*[-/-];*Trp73*[sh]), or p53/p73-depleted cells expressing empty vector or Aurka-resistant (p53[SA], p63[SA], p73[SA]) or Aurka phosphorylation mimic (p53[SD], p63[SD], p73[SD]) mutant forms of p53 family members. Tubulin was used as loading control. (Figure **B**: Courtesy of Qiong Wang, PhD)

129

3.3.7 Wnt3 mediates p53 family action in mesendoderm specification

The necessity of Wnt3 in development and its direct regulation by p53 family predicted Wnt3 as an important downstream mediator of p53/p63/p73 action in lineage specification. To confirm the specificity of p53 regulation of Wnt signaling, we determined whether enforced Wnt signaling was capable of rescuing mesendoderm marker gene expression in p53 family-deficient ES cells. Wnt3a is a more stable homolog of Wnt3 that can mimic Wnt3 activity *in vitro*. Addition of recombinant mouse Wnt3a to d3 *Trp53*[-/-];*Trp73*[sh] EBs fully rescued the expression of *Eomes*, *Foxa2* and *Gsc* and the binding of Smad2/3 to the enhancers of these genes in response to Activin (Figure 3-37, 3-38). Wnt3a induction of the feedback regulator *Axin2* did not require p53 inputs (Figure 3-37). We also engineered wild type or *Trp53*[-/-];*Trp73*[sh] ES cells with inducible expression vectors encoding mouse *Wnt3* or, as a control, *Tcf3*. Doxycy-cline-mediated induction of *Wnt3,* but not that of *Tcf3,* rescued *Eomes, Foxa2*, and *Gsc* expression (Figure 3-39).

Immunofluorescence analysis of EBs showed that Eomes levels and the number of Eomes expressing cells were significantly elevated by Activin addition in *Trp53*[+/+];Ctrl[sh] EBs but not in *Trp53*[-/-];*Trp73*[sh] or in TKO triple knockout EBs. The combined addition of Activin and Wnt3a to *Trp53*[-/-];*Trp73*[sh] EBs or TKO EBs fully restored Eomes expression and the number of Eomes-expressing cells (Figure

3-40). The same pattern was observed with FoxA2 expression (Figure 3-41), confirming that Wnt3 mediates p53 family action in mesendoderm specification.

A
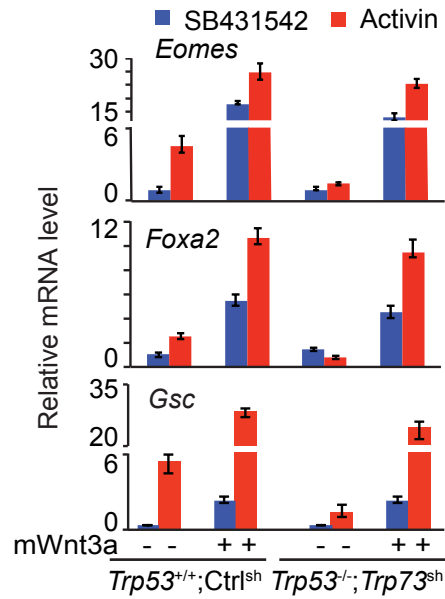
B
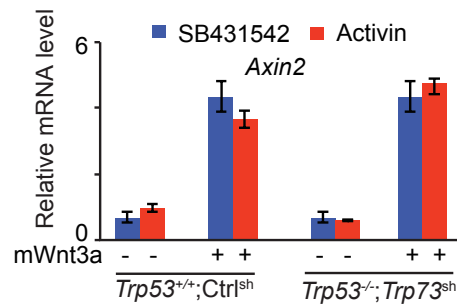
Figure 3-37. (**A**) qRT-PCR analysis of *Eomes, Foxa2, Gsc* and *Axin2* mRNA levels in control ES cells (*Trp53*+/+;Ctrl^sh), or p53/p73-depleted cells (*Trp53*-/-;*Trp73*^sh), that were treated with or without recombinant mouse Wnt3a (mWnt3a) for 24h, followed by addition of SB431542 or Activin for 2 h. Error bars represent s.e.m.. (B) qRT-PCR analysis of *Axin2.* (Courtesy of Qiong Wang, PhD)
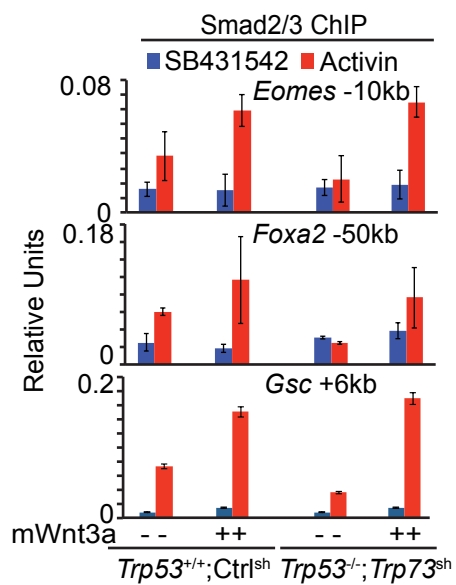
Figure 3-38. ChIP-qPCR analysis for Smad2/3 binding to the *Eomes* -10kb, *Foxa2* -50kb and *Gsc* +6kb enhancers in control ES cells (*Trp53*[+/+];Ctrl[sh]), or p53/p73-depleted cells (*Trp53*[-/-];*Trp73*[sh]) treated with PBS or recombinant mouse Wnt3a (mWnt3a), followed by addition of SB431542 or Activin for 2 h. Error bars represent s.e.m.. (Courtesy of Qiong Wang, PhD)
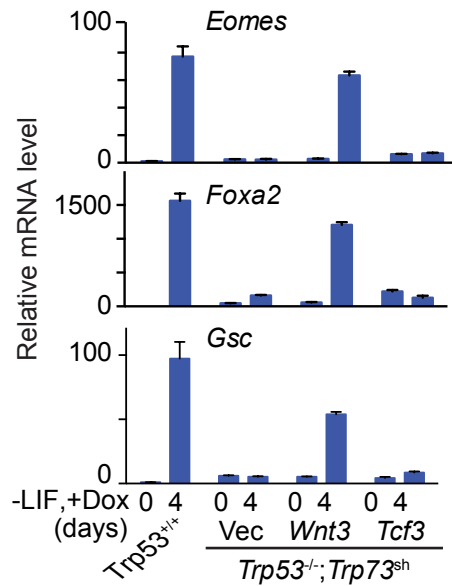
Figure 3-39. qRT-PCR analysis of *Eomes, Foxa2* and *Gsc* mRNA levels in p53/p73-depleted cells with doxycycline (Dox)-inducible expression of *Wnt3*, *Tcf3* or empty vector control. Assays were performed at d0 or day 4 of Dox treatment under differentiation conditions. Error bars represent s.e.m. (Courtesy of Qiong Wang, PhD)
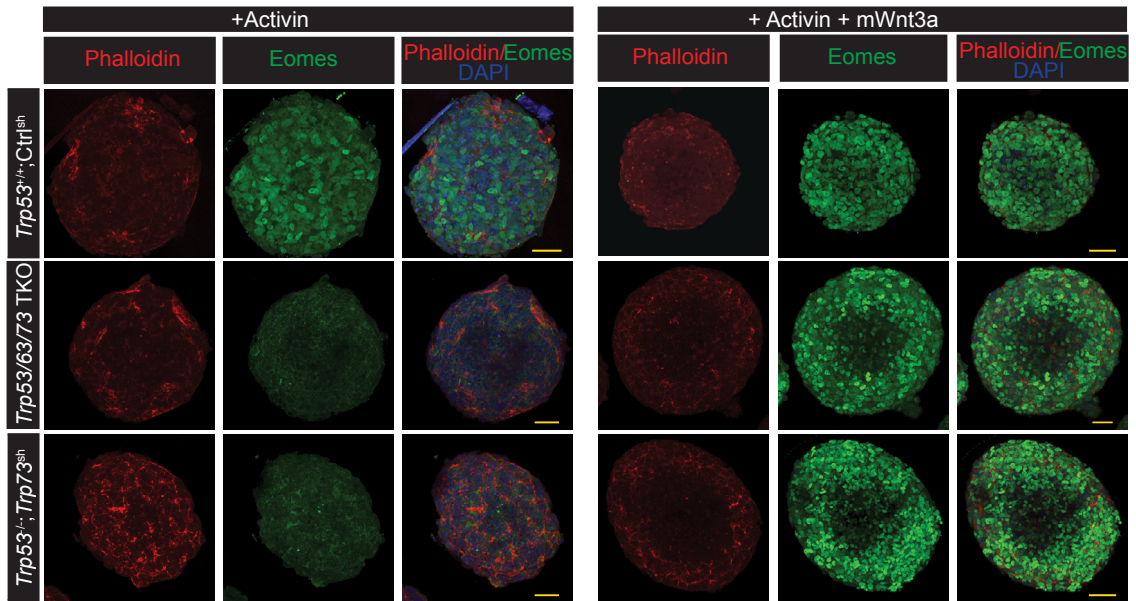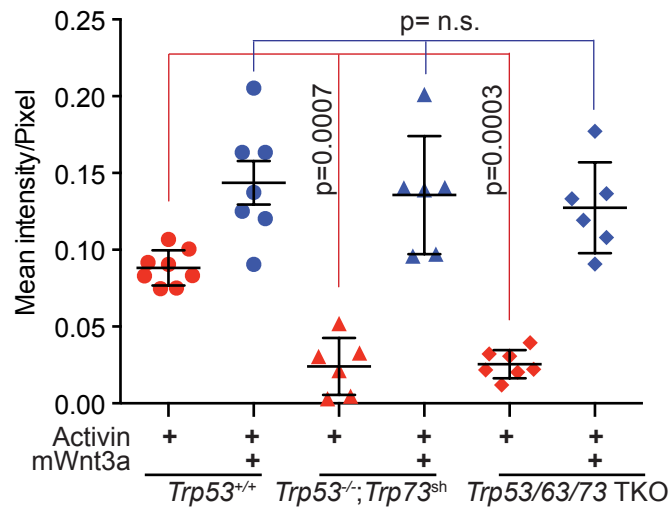
Figure 3-40. (**A**) Immunofluorescence analysis of Eomes (*green*) in d3 EBs derived from the *Trp53*[+/+];Ctrl[sh], *Trp53/63/73* TKO, or *Trp53*[-/-];*Trp73*[sh] ES cells. Cells were treated with Activin for 2h or Activin plus mWnt3a for 20h. Scale bars, 50μm. Phalloidin (*red*) was used to stain F-actin and mark the cell boundaries. (**B**) Quantitation of Eomes (*green*) signal. Each dot represents the signal measurement from one EB. p value was calculated using Mann-Whitney test.
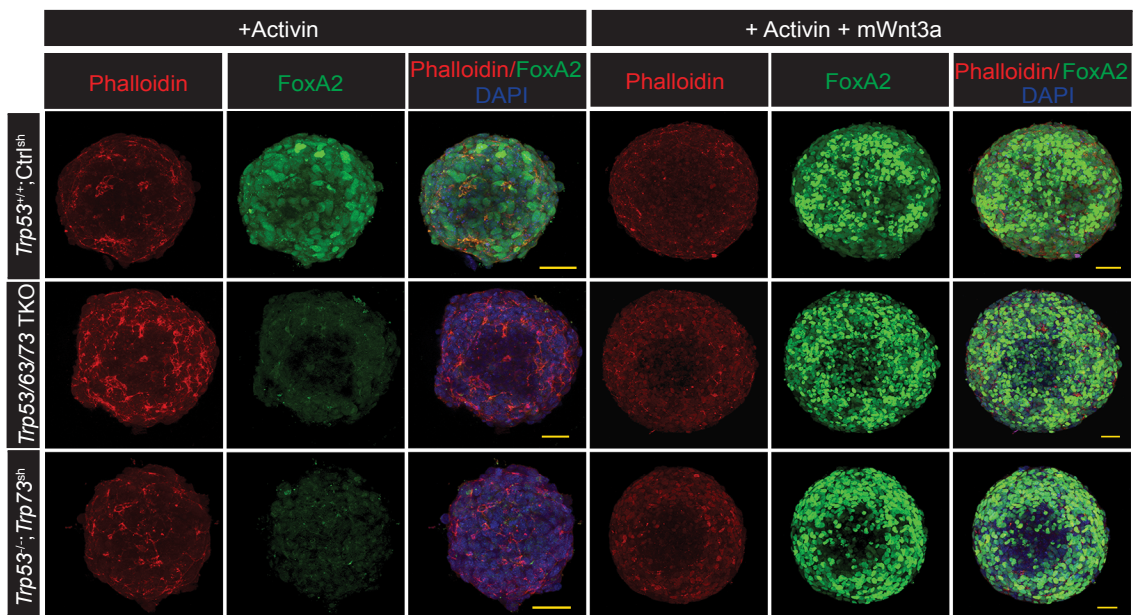
Figure 3-41. Immunofluorescence analysis of FoxA2 (*green*) expression in d3 EBs derived from the indicated ES cell lines. Cells are treated with Activin A for 20h or Activin A with mWnt3a for 20h. Scale bar of each figure represents 50µm. Phalloidin (*red*) was used to stain F-actin and mark cell boundaries.
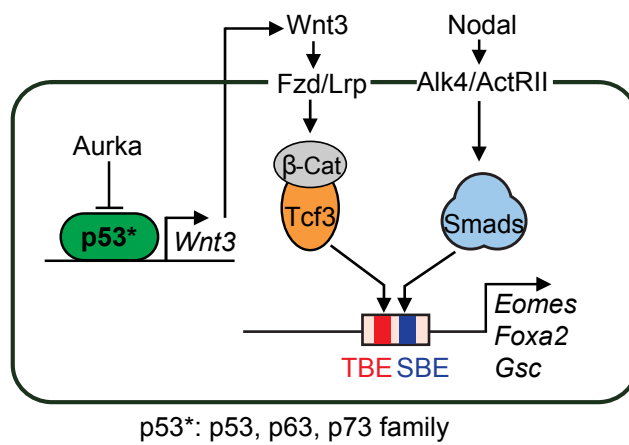
Figure 3-42. Scheme of the p53 family-Wnt-Nodal network driving mesendoderm specification.

3.4 Conclusions and Discussion

These results establish the relevance of the p53 family in the differentiation of pluripotent progenitors within the early embryo, as well as pluripotent ES cells in culture, and provide a mechanistic basis for this role. I show that the p53 family governs a regulatory network that integrates Wnt and TGF-β/Nodal inputs for mesendoderm specification. This network includes two layers of regulation, first, p53 family members that directly control Wnt3 expression, and second, Wnt activated β-Catenin/Tcf3 and Nodal-activated Smad2/3 that are mutually dependent for binding to, and activation of key mesendoderm identity genes (Figure 3-42). The activation of this network is tied to the release of the p53 family from inhibition as ES cells exit from pluripotency. These findings clarify the role of p53 during early embryogenesis, identify p53 target genes implicated in this process, and highlight the interdependent nature of Nodal and Wnt transcriptional mediators in early development.

My study could have three major implications: first, it emphasized the importance of functional redundancy among p53 family members during differentiation, which has masked the manifestation of the developmental roles of this family in context of single or double knockouts and many other species. Second, I identified Wnt3 as central mediator of p53 family action, providing an important insight in understanding the non-apoptotic functions of the p53 network. Third, I illustrated a paradigm for how TGF-β/Nodal signaling and Wnt

pathway could be coordinated by a seemingly unrelated factor, the p53 family, to act synergistically in cell fate regulation. This coordination is certainly not limited to embryonic stem cells, but also has implications for many other contexts of cell differentiation, tissue regeneration and disease development.

3.4.1   The p53 family and mesendoderm differentiation

An allelic series of knockouts of p53 family members show that p53, p63 and p73 play a critical role in mesendoderm specification of pluripotent progenitors within the embryo. All three p53 family members can fulfill this role. In line with previous reports (Bazzi and Anderson, 2014; Lin et al., 2005; Lutzker and Levine, 1996; Shigeta et al., 2013), high expression levels of p53 were observed in early embryo, embryoid bodies, and ES cells in monolayer culture. Expression of p63 and p73 in mouse embryos and EBs varies depending on the culture conditions, and mesendoderm specification can proceed as long as at least one p53 family member is present.

Mouse embryos express all p53 family members in a dynamic pattern around the time of gastrulation, providing opportunities for redundant functionality. The functional overlap of the p53 family members in mesendoderm specification would mask their role in gastrulation in single or double knockout mice. Indeed, *Trp53* mutant mice develop normally (Donehower et al., 1992; Jacks et al., 1994; Purdie et al., 1994; Tsukada et al., 1993) whereas *Trp63*

mutant and *Trp73* mutant mice have late but not early developmental defects (Mills et al., 1999; Yang et al., 1999a; Yang et al., 2000). Evidence for a role of the p53 family in the early embryo rested on transgenic overexpression of ΔNp73, a dominant negative inhibitor of all p53 family members that causes embryonic lethality at gastrulation (Erster et al., 2006; Huttinger-Kirchhof et al., 2006; Yang et al., 2000). In collaboration with the Hadjantonakis lab, we now show that strong TKO chimeras fail to execute normal gastrulation and TKO cells in chimeras do not express mesendodermal differentiation markers. The minor defects in p53 null mice in some tissues might reflect the lack of compensatory family members.

3.4.2   A p53-Wnt link

p53 activity in mouse pluripotent ES cells is kept in check by LIF through Aurka, a constraint that is relieved as Aurka levels drop upon LIF removal (Lee et al., 2012). This regulatory potential is conserved in p63 and p73. The timing of Aurka decline in differentiating EBs coincides with the onset of Wnt3 expression and the induction of mesendoderm differentiation genes. *Wnt3* and the Wnt receptor gene *Fzd1* are members of a small set of direct p53 family target genes that become activated in differentiating ES cells. p53, p63 and p73 directly bind to a common regulatory element in *Wnt3* to drive its expression. Therefore, Wnt3 is a direct p53 transcriptional target that is central for the control of mesendoderm differentiation by the p53 family. In p53

family-deficient EBs, Wnt3a addition rescues differentiation. The ability of p53 to induce the expression of Wnt components in ES cells is not unprecedented. Treatment of pluripotent ES cells with DNA damaging agents can trigger p53-dependent expression of Wnt components (Lee et al., 2010). However, this xenotoxic response was of unknown relevance to the developmental context and, moreover, it was interpreted as a mechanism for p53-dependent inhibition of differentiation. In contrast, our evidence reveals that Wnt expression is low in pluripotent cells, and is significantly elevated by p53/p63/p73 at the onset of differentiation. Wnt signaling activated under differentiation-permissive conditions cooperates with Nodal signaling and drives mesendoderm specification.

3.4.3   Integration of Nodal and Wnt inputs

The Wnt and TGF-β pathways jointly regulate progenitor identity and differentiation in many contexts (Brennan et al., 2001; Clevers and Nusse, 2012; Massague, 2012). The present identification of the p53 family as central integrators of Wnt3 and Nodal inputs in mesendoderm differentiation provides an expanded view of the synergistic properties of these two pathways, and a previously unrecognized level of selectivity and interdependency in the interaction of activated Tcf and Smad transcription factors.

Nodal target genes in differentiating ES cells fall into two classes depending on the requirement for p53. Genes of the p53-independent class contain Smad2/3 binding enhancers that are not enriched for TCF binding

141

sites or bound by Tcf3. These genes include Nodal/Smad pathway feedback regulators, among others. Genes in the p53-dependent class contain Smad2/3 binding enhancers that are co-occupied by Tcf3, and include the mesendoderm cell identity genes highlighted here. Nodal-activated Smad2/3 and Wnt-activated Tcf3 cooperate in binding to common target enhancers in p53-dependent mesendoderm cell identity genes that drive this crucial developmental transition during gastrulation. Smad2/3 binding to these enhancers requires not only Nodal signaling but also Wnt signaling. Tcf binding to these enhancers also requires signaling by both pathways. Nodal expression and Smad2/3 activation are constitutive in differentiating ES cells, whereas Wnt3 expression for β-Catenin/Tcf activation requires the upstream input of the p53 family.

The evidence suggests that the p53 family governs the integration of Wnt and Nodal signals for mesendoderm differentiation. Although the present work highlights the importance of this p53-Wnt-Nodal network in mesendoderm differentiation, this process is also regulated by other determinants including histone modifications (Bernstein et al., 2006; Whyte et al., 2012), chromatin topology (Dixon et al., 2015), chromatin remodeling (Alexander et al., 2015; Chen and Dent, 2014) and cell cycle regulators (Pauklin and Vallier, 2013). My work also demonstrated the functional importance of developmental enhancers in serving as a platform that integrates multiple transcription factors, epigenetic regulators and general transcriptional machinery. It will be of interest in the future to discern how these regulatory and epigenetic events converge and interlock to

bring about the transition from pluripotency to lineage restricted progenitors.

3.4.4   Implications of p53-Wnt-Nodal network in iPS reprogramming

Generation of induced pluripotency cells from somatic cells is a major breakthrough in regenerative medicine. By genome-wide shRNA screening, p53 was identified and characterized as a major barrier for induced pluripotency from somatic cells (Zhao et al., 2008) (Hong et al., 2009; Huang et al., 2009; Kawamura et al., 2009; Marion et al., 2009). However, the mechanisms that p53 utilizes to inhibit pluripotency is unknown. As p53-depletion is potentially hazardous for stem cells due to unprotected genome integrity, a thorough understanding about the molecular mechanisms that p53 utilizes to control differentiation is in urgent need. Our analysis with differentiation of pluripotent ES cells suggests that p53 could inhibit pluripotency through induction of differentiation signals. p53 induces expression of Wnt3 and Fzd1, and perhaps other components of Wnt signaling pathway including Wnt2, Wnt8a, Fzd6, Lef1 etc (p53 ChIP-Seq data). Activation of Wnt pathway, together with active TGF-β/ Nodal signaling, could induce differentiation and serve as a roadblock for induced pluripotency.

3.4.5   Downstream mediators of p53

More than 2000 p53 target genes were identified by p53 ChIP-Seq and

more than 290 genes exhibit differential expression in response to p53/p73 depletion in differentiating ES cells, many of which are important transcription factors or epigenetic regulators. Some of the other p53 target genes associated with ES cell differentiation (*Ccng1, Cdnk1a*; refer to Figure 3-31) are also regulated by p53 in the context of DNA damage responses in other cell types (el-Deiry et al., 1993; Kawase et al., 2009; Okamoto and Beach, 1994). Overall however, the p53 target gene set in the context of mesendoderm differentiation does not resemble a canonical p53 DNA damage response (Menendez et al., 2009; Riley et al., 2008). Typical p53-regulated pro-apoptotic genes, for example, are absent. DNA damage caused p53 to repress *Nanog* in ES cells (Lin et al., 2005), but in this study no significant change in *Nanog* expression was observed in p53/p73-depleted ES cells.

One of the top target genes of p53 is *Phlda3* (Pleckstrin Homology-Like Domain, Family A, Member 3) (Figure 3-31), a suppressor of PI3K/Akt signaling (Kawase et al., 2009). Loss of p53/p73 leads to downregulation of Phlda3 (Figure 3-31). This link provides an intriguing link between the p53 network and PI3K/Akt/ mTOR signaling. In fact, in the primordial germ cells, Akt signaling is required for pluripotency, and p53 deficiency results in similar defects as Akt hyperactivation (Kimura et al., 2008). Moreover, Akt inhibition is required for the onset of mesendoderm differentiation (Figure 3-43). These results provide a potential molecular link between the down-regulation of Akt signaling and the initiation of mesendoderm specification controlled by p53. Further experimentation is in need

to determine whether the activity of Akt pathway is controlled by p53 in ES cells and embryos.

3.4.6   The p53 family: maintaining genome-integrity and guiding differentiation

The p53 family has been established as central regulators in genome integrity maintenance, apoptosis, cell cycle arrest and senescence. Here I demonstrated another potential branch of p53 family functions: driving Wnt pathway activation and coordinating mesendoderm differentiation. The differentiation function of p53 is not limited to embryo patterning, indeed, the p53 family members are directly and redundantly involved in myogenic differentiation and rhabdomyosarcoma development (Cam et al., 2006). Further studies are required to dissect how these two branches of functions in the p53 family are coordinated.

A

B
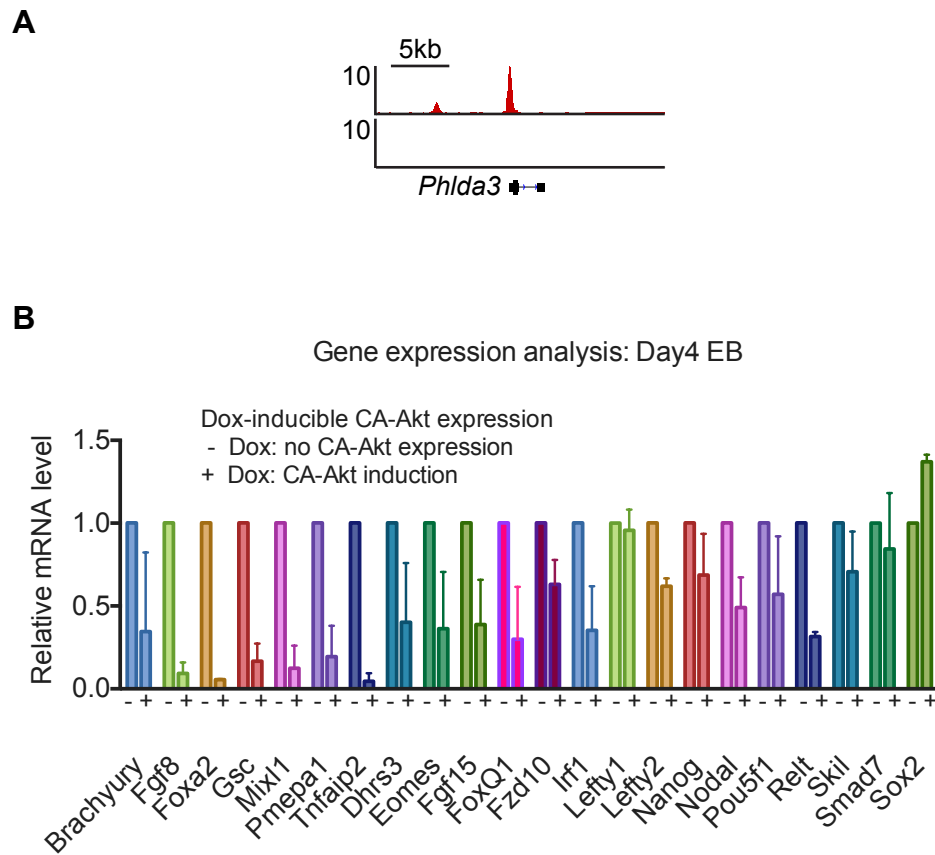
Figure 3-43. Link between p53-Phlda3 and Akt signaling in differentiation regulation. (**A**) Gene track view of p53 ChIP-Seq signal at *Phlda3* locus. (**B**) qRT-PCR analysis of Nodal response genes in the presence or absence of Doxycycline induced CA-Akt (constitutive active Akt) expression.

Chapter 4   Smad2X, a Splicing Isoform of *Smad2* Transmits Nodal Signaling and Governs Mesendoderm Specification

4.1 Summary

Smad2/3 are the major downstream effectors of Nodal/Activin/TGF-β signaling. Though the protein products from *Smad2* and *Smad3* genes are structurally similar and functionally interchangeable in various contexts, *Smad2* but not *Smad3* depletion is embryonic lethal in mice. The underlying mechanisms for these differences are unknown and the R-Smad protein that mediates Nodal-driven mesendoderm specification has not been determined. Here I describe that *Smad2X*, a splicing isoform of *Smad2* retaining the exon3, is the most potent inducer of mesendoderm differentiation among all *Smad2/3* protein products. In response to Nodal/Activin stimulation, Smad2X recruits key co-factors to assemble functional transcriptional machinery. Loss of function studies revealed that Smad4, FoxH1 and Trim33 are all required for the transcriptional activation of Smad2X-dependent genes. Besides protecting Smad2X from GSK3β/Axin mediated phosphorylation and degradation, the exon3 insert inhibits Smad2X from binding DNA directly. The reduced DNA binding affinity and specificity potentially improves the flexibility for Smad2X to

act as molecular partners of lineage-specific transcription co-factors. This study suggests a novel layer of regulation involved in controlling the target gene specificity of the Nodal/Activin/TGF-β pathway.

4.2 Introduction

In the preceding chapter, I identified a p53-Wnt-Nodal signaling network in controlling the embryonic mesendoderm specification. This work underscores the concerted action between Wnt-Tcf and Nodal-Smad2/3 signaling in directly activate mesendoderm specification genes, and establishes the p53 family as central regulators of embryo development. Thus, transcriptional activation of Nodal target genes is facilitated by cooperation between the core Smad machinery composed of Smad2/3/4 and transcriptional co-factors including FoxH1, Trim33, and p53/p73 stimulated Tcf3/β-Catenin complex (Chen et al., 1996; Chen et al., 1997; Chiu et al., 2014; Xi et al., 2011). It is unclear whether these factors are involved in regulating the same set of target genes, and how is the assembly of multiple co-factors regulated.

Notably, mesendoderm specification genes only comprise a small subset of Nodal/Activin response genes in embryos or EBs. Once activated, Nodal stimulates the expression of a wide range of genes including pluripotency-associated factor *Nanog* (Vallier et al., 2009)*,* left-right asymmetry determining factor *Lefty1/2* (Yamamoto et al., 2003; Yamamoto et al., 2004), intracellular feedback regulators *Smad7* and *SnoN/Skil* etc (Macias-Silva et al., 2002; Mizuide et al., 2003; Nakao et al., 1997; Stroschein et al., 1999). Moreover, beyond germ layer specification, the TGF-β family is involved in a plethora of biological processes, each activating radically divergent target gene sets. For instance,

TGF-β drives pancreatic progenitor differentiation through Sox4 and Klf5 (David et al., 2016), induces epithelial-mesenchymal transition through Snail1/2 and Zeb1/2 in epithelial cells (Xu et al., 2009), converts CD4+CD25- naïve T cells to CD4+CD25+ T_{reg} cells through Foxp3 (Chen et al., 2003), induces cytostasis in pre-malignant cells through p27 and p57 (Lee et al., 1995; Polyak et al., 1994a; Polyak et al., 1994b; Scandura et al., 2004). In each cell type TGF-β activates a set of cell type-specific target genes in addition to a group of universal target genes such as *Smad7* and *Skil*. In addition, the effector Smad2/3 binds to drastically different locations in the genome in various cell types (Mullen et al., 2011; Qin et al., 2016). It remains poorly understood about the structural basis for the versatility in target genes and functions downstream of Nodal signaling and how a single signaling cascade coordinates multiple target gene programs in the same cells.

Smad2/3/4 transmits the Nodal/Activin/TGF-β signaling. Each Smad contains two conserved MH1 and MH2 domains, connected by a linker region. The MH1 domain recognizes particular DNA sequences, and the MH2 domain can mediate protein-protein interactions (Shi and Massague, 2003). One exception in the R-Smad family is Smad2. The amino acid sequence between Smad2 and Smad3 is 92% identical except that Smad2 contains an extra 30 residue insert in the MH1 domain resulted from exon3 in the Smad2 gene (Figure 4-1). The retaining or splicing of exon3 results in a long isoform designated as "*Smad2X*", and a short isoform *Smad2* (Takenoshita et al., 1998). It is poorly

understood that how each of these R-Smad proteins is involved in TGF-β signaling and whether they play distinct roles in mediating the versatile functions of TGF-β.

Smad2 and Smad3 proteins are structurally and functionally similar whereas Smad2X function differently in many contexts. *Smad2* gene (which is largely transcribed and translated into Smad2X protein rather than Smad2) knockout led to embryonic lethality due to failure in gastrulation events (Nomura and Li, 1998; Tremblay et al., 2000; Waldrip et al., 1998; Weinstein et al., 1998), recapitulating phenotypes of knockout other Nodal pathway components including Nodal, Smad4, Trim33 and Foxh1 (Conlon et al., 1994; Zhou et al., 1993; Yang et al., 1999b; Sirard et al., 1998; Morsut et al., 2010; Hoodless et al., 2001; von Both et al., 2004; Yamamoto et al., 2001). In contrast, *Smad3*[-/-] mice were viable and fertile, with minor later stage defects in immunity (Ashcroft et al., 1999; Yang et al., 1999b). Smad2/2X/3 all can be effectively phosphorylated by activated TGF-β receptors (Yagi et al., 1999). Previous reports suggested that Smad2/3/4 but no Smad2X MH1 domains bind to a palindromic CAGATCTG sequence (Zawel et al., 1998; Dennler et al., 1999). Nonetheless, the strict conservation in the exon3 sequence during evolution and why most vertebrates maintain a diversified pool of Smad2/2X/3 isoforms remain puzzling.

Xenopus embryo explant experiments suggested that Smad2X MH2 domain is sufficient for activating downstream components and the MH1 domain

is functionally repressive and anchors the full-length protein in the cytoplasm (Baker and Harland, 1996; Hata et al., 1997). Transgenic mouse experiments further supported that Smad2X C-terminal (241-467) containing the linker and the MH2 domain is sufficient in transiently activating mesendoderm marker genes and inducing mesoderm (Brennan et al., 2001; Das et al., 2009; Tremblay et al., 2000; Waldrip et al., 1998). These observations suggest that the MH1 domain of Smad2X is not directly required for its transactivation activity.

Here using mouse ES cells and EBs as experimental differentiation models, I systematically characterized the transcriptome and cistrome downstream of Nodal/Activin signaling. A panel of epigenetically regulated, mesendoderm-specific enhancers was identified as regulatory platforms that accommodate Nodal effectors. I demonstrate that *Smad2X* is the most potent inducer of mesendoderm differentiation among all *Smad2/3* protein products. In response to Nodal/Activin stimulation, Smad2X recruits key transcriptional co-factors to assemble functional transcriptional machinery. Loss of function studies revealed that Smad2X, Smad4, FoxH1 and Trim33, in addition to p53/ p73-stimulated Wnt-Tcf3/β-Catenin signaling axis, cooperatively regulate an overlapped set of mesendoderm differentiation genes. The unique requirement of Smad2X in serving as a nucleating factor is potentially assisted by its suppressed DNA binding specificity and affinity. This study discerns a new regulatory mechanism in controlling the target gene specificity of the TGF-β pathway.

4.3 Results

4.3.1 Smad2X is the most abundantly expressed isoform among Smad2/2X/3 proteins

Nodal/Activin/TGF-β signaling is transmitted by the receptor-regulated Smads, Smad2/3 and the co-Smad, Smad4 (Figure 4-1). The exon3 in the *Smad2* pre-mRNA can be alternatively spliced, resulting in two mRNA and protein products: exon3 containing isoform *Smad2X* (Smad2X) and exon3 depleted isoform *Smad2* (Smad2) (Figure 4-1, 4-2). Both isoforms contain highly conserved MH1 and MH2 domains, and a linker region (Figure 4-1, 4-2). The short Smad2 isoform structurally and functionally resembles Smad3. Though the biochemical function of exon3 in Smad2X protein is unclear, the amino acid sequence encoded by the exon3 is highly conserved throughout vertebrate evolution (Figure 4-2). In addition, the DNA sequences surrounding exon3 in intron 2 and intron3, which contain potential regulatory elements for alternative splicing, are highly conserved across mammals (Figure 4-2). These analyses suggest that exon3 splicing is highly regulated and exon3 might possess fundamentally important functions.

To determine the relative expression of *Smad2/2X* transcripts in mouse ES cells, I examined the read distribution from paired-end RNA-Seq analysis in ES cells and EBs. Sequencing read pairs derived from *Smad2X* transcripts

will be at least partially mapped to the exon3 region, whereas read pairs from *Smad2* transcripts will directly connect exon2 and exon4, skipping the exon3 locus (Figure 4-3). Read pair quantitation showed that *Smad2X* transcripts are 10-20 fold more abundant than *Smad2* transcripts in the ES cells and EBs. To determine the relative abundance of *Smad2/2X* in mature tissues, I applied similar analysis methods to interrogate the ENCODE mouse tissue specific RNA-Seq databases. In majority of the examined mouse tissues including the kidney, liver, heart and lung, *Smad2X* is expressed 10-40 fold higher than *Smad2* (Figure 4-3). One exception is in the brain tissues, *Smad2* transcripts accounts for 20-40% of total *Smad2/2X* transcripts (Figure 4-3). Both analyses suggested that *Smad2X* is the major transcript spliced from the *Smad2* gene in ES cells and most adult tissues.

To determine the expression of Smad2/2X/3 at protein level, I identified one antibody that recognizes all forms of Smad2, Smad2X and Smad3 proteins (designated as "Smad2/2X/3 antibody"), another antibody that recognizes only the two Smad2/2X isoforms without cross-reacting with Smad3 ("Smad2/2X antibody"), and a third antibody that specifically recognizes Smad2X ("Smad2X antibody") (Figure 4-3). These antibodies were validated with the KT15 *Smad2*[-/-] cell line (disrupting the *Smad2* gene locus and lacking both *Smad2/2X*) transduced with HA tagged human SMAD2/2X transgene expression (Tremblay et al., 2000; Waldrip et al., 1998). Consistent with the mRNA analyses, western blot analysis in ES cells and EBs showed that Smad2X protein is the most

abundant protein form among Smad2/2X/3, while Smad3 is medially expressed and Smad2 is below the detection limit in these conditions (Figure 4-3). Collectively, at both mRNA and protein levels, Smad2X is the highest expressed isoform from *Smad2/3* gene locus in ES cells, EBs, and majority of mature tissues. The dominance of Smad2X in protein expression and high conservation in exon3 sequence support that this isoform plays an indispensable role in mediating TGF-β family responses.
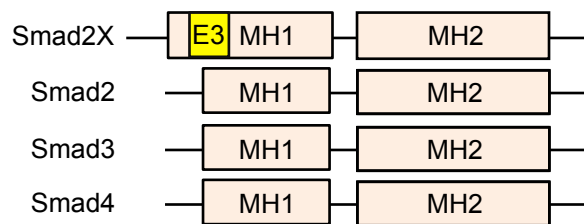
Figure 4-1.   Scheme of domain structures of Smads downstream of Nodal/Activin
signaling. Each Smad protein contain a conserved MH1 (MAD homology 1) and a
conserved MH2 (MAD homology 2) domain, connected by a less conserved linker
region. An additional exon3 (E3) region in Smad2X is retained through alternative
splicing.

A

Smad2X

	Human	T K C V T I P S T C S E I W G L S T P N T I D Q W D T T G L Y S F S E Q T R S L D
	Chimpanzee	T K C V T I P S T C S E I W G L S T P N T I D Q W D T T G L Y S F S E Q T R S L D
	Rat	T K C V T I P S T C S E I W G L S T A N T V D Q W D T T G L Y S F S E Q T R S L D
Smad2X	Mouse	T K C V T I P S T C S E I W G L S T A N T V D Q W D T T G L Y S F S E Q T R S L D
	Pigeon	T K C V T I P S T C S E I W G L S T P N T I D Q W D T T G L Y S F S E Q T R S L D
	Xenopus	T K C V T I P S T C S E I W G L S T P N T I D Q W D T T G L Y S F S E Q T R S L D
	Zebrafish	T K C V T I P S N C S E I W G L S T P N T I E Q W D T S G L Y S Y P D Q T R S L D
Smad3	Human	T K C I T I P - R S L D
	Mouse	T K C I T I P - R S L D

B

Figure 4-2. Conserved exon3 sequence and the related regulatory DNA regions. (**A**) Amino acid sequence alignment of Smad2X and Smad3 from representative vertebrate species. The region surrounding exon3 locus is highlighted. (**B**) Conservation of Smad2 exon3 genomic sequence and the surrounding DNA sequences in intron2 and intron3. Regions surrounding exon7 is shown as control.
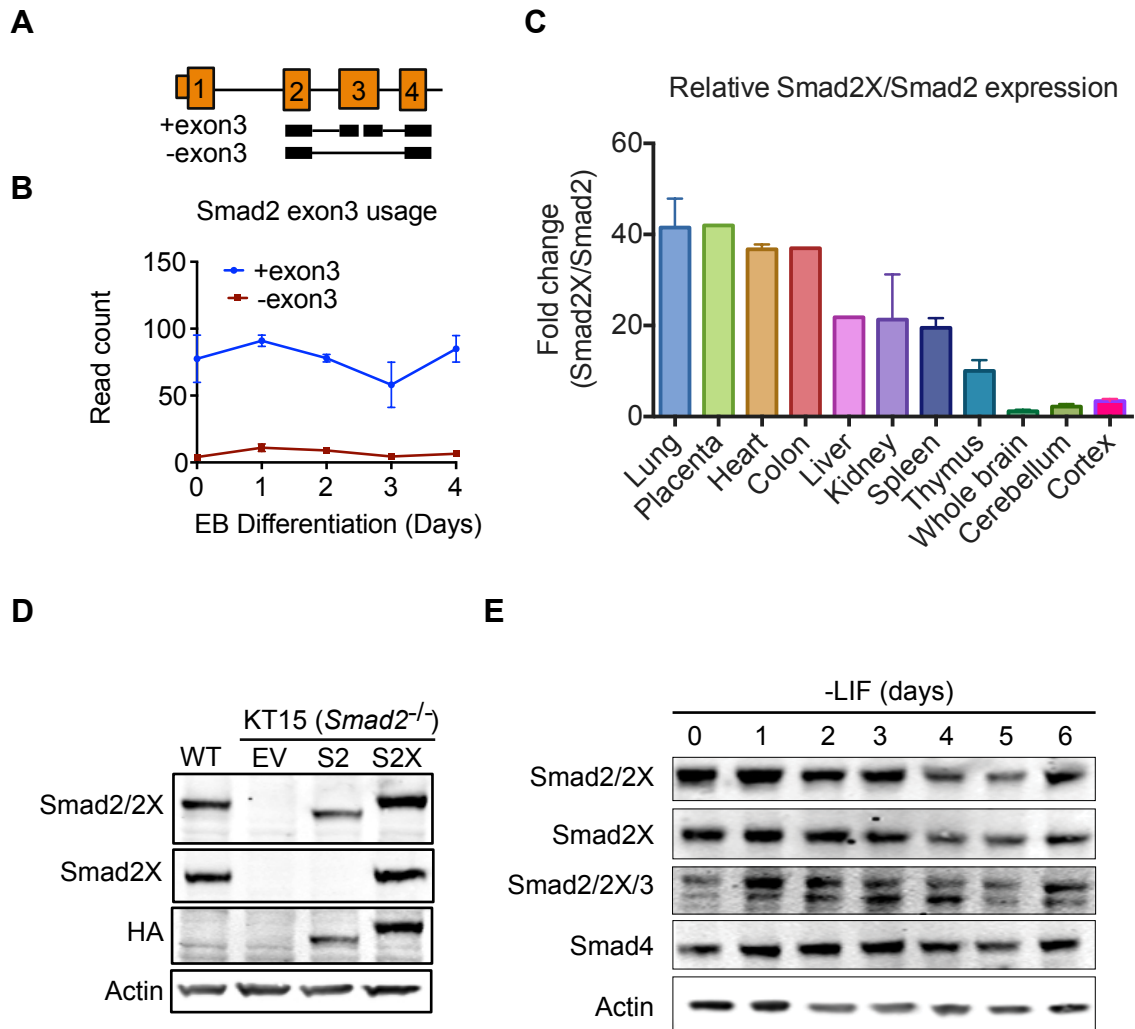
Figure 4-3. Relative mRNA and protein expression levels of Smad2 and Smad2X isoforms. (**A**) Scheme of RNA-Seq distribution in -exon3 (Smad2) or +exon3 (Smad2X) transcripts. (**B**) RNA-Seq count number in d0-4 EBs. (**C**) Ratio of Smad2X/Smad2 relative mRNA level determined by analyzing RNA-Seq read pair distribution. Datasets retrieved from the ENCODE project (www.genome.gov/encode). (**D**) Western immunoblotting analysis of Smad2/2X expression in wild type ES cells or KT15 (*Smad2*[-/-]) cells transduced with HA-hSMAD2/2X expressing vectors. The first panel is blotted with an anti-Smad2/2X antibody. The second panel is blotted with an antibody that specifically recognizes Smad2X isoform. (**E**) Western immunoblotting analysis of Smad2/2X/3, Smad4 expression in d0-6 EBs. Actin was used as loading controls.

4.3.2 Smad2X mediates mesendoderm differentiation in ES cells

Smad2 null mice are embryonic lethal whereas Smad3 knockout leads to no obvious developmental defects (Nomura and Li, 1998; Weinstein et al., 1998; Yang et al., 1999b). My analysis emphasized that majority of the transcripts from Smad2 gene is translated into the Smad2X protein. These connections lead to the hypothesis that the developmental functions of Smad2 gene are executed by Smad2X protein. To investigate the requirement of Smad2X in transducing Nodal stimulated mesendoderm differentiation, I utilized the CRISPR/Cas9 technology to delete the exon3 from both Smad2 alleles in ES cells (Figure 4-4). To reduce the off-target effects with wild type Cas9 protein, I followed a modified CRISPR sgRNA design protocol by reducing the length of base-pairing sequence from 20bp to 18bp (Fu et al., 2014) (Figure 4-4). Truncated sgRNAs (Tru-sgRNAs) were designed flanking the exon3 of Smad2 and transiently transfected into ES cells (Figure 4-4). Exon3 deleted ES cell clones were selected by PCR product size shift using primers spanning the exon3 region (Figure 4-4). Notably, exon3 deletion leads to up-regulation of short Smad2 transcript (Figure 4-4). Western blot analysis confirmed two clones that contain homozygous deletions of exon3 (Smad2-Ex3$^{-/-}$) (Figure 4-4).

Knockout Smad2 exon3 did not alter the growth and morphology of mouse ES cells (Figure 4-5). To determine whether deletion of Smad2 exon3 leads to alterations in Activin dependent lineage commitment, I measured the expression

of mesendoderm marker genes *Gsc, Mixl1* and Nodal feedback gene *Smad7* in wild type (Ex3$^{+/+}$) and *Smad2*-Ex3$^{-/-}$ cells during EB differentiation. Expression of *Gsc* and *Mixl1*, but not *Smad7* was significantly reduced in day 4 EBs with short-term Activin A treatment (Figure 4-6). To confirm the defect in *Gsc* induction was specifically due to loss of Smad2X, I engineered the *Smad2*-Ex3$^{-/-}$ cells with constitutive expression of HA-epitope tagged human *SMAD2/2X* cDNA or empty vector. qRT-PCR analysis confirmed that only SMAD2X could effectively rescue *Gsc* and *Mixl1* expression without altering Activin-induced *Smad7* expression (Figure 4-6). These results suggest that Smad2X is required for the expression of mesendoderm marker genes.

To systematically assess the function of Smad2X in mediating Nodal/ Activin response, I determined the Activin pathway target genes induced in pluripotency versus differentiation conditions by RNA-Seq. Differential gene expression analysis identified 72 Activin-dependent genes in the ES cells and 41 genes in the EBs (Figure 4-7). Surprisingly, only 16 genes overlapped between these two conditions. Moreover, a number of these 16 genes such as *Nanog*, *Smad7*, *Lefty1*, *Lefty2* and *Pmepa1* levels were significantly down-regulated during differentiation (Figure 4-7; refer to Chapter 3, Figure 3-1, 3-2). 56 genes were only responding to Activin in ES condition but not in EB condition, which is consistent with the engagement of Activin signaling in regulating pluripotency (James et al., 2005). In contrast, all major lineage specification factors such as *Eomes*, *Gsc*, *Mixl1*, *Foxa2* and *Fgf8* were induced merely in the EB condition

(Figure 4-7). These results underscore that Nodal/Activin signaling is necessary but not sufficient for inducing mesendoderm differentiation, and a licensing event is required for the transcriptional activation of developmental genes. As revealed in Chapter 3, the p53-Wnt-Tcf axis drives the activation of mesendoderm specification genes through cooperation with Nodal/Smad signaling.

I determined by qRT-PCR the Activin response genes that specifically require Smad2X. Loss of exon3 specifically reduced the Activin dependent induction of *Gsc*, *Fgf8*, *Eomes*, *T/Brachyury*, *Mixl1, Foxa2* and *Wnt3*, whereas other genes including *Lefty1*, *Hes1*, *Smad7*, *Chst15*, *Foxq1, Nodal, Nanog*, *Pitx2*, *Fzd10*, *Pmepa1*, *Rgs16*, *Skil*, *Irf1* and *Lefty2* were independent of exon3 status (Figure 4-8)*.* These dependencies were confirmed in KT15 (*Smad2$^{-/-}$*) ES cells (Waldrip et al., 1998) transduced with HA-*hSMAD2/2X* cDNA or empty vector (Figure 4-9, 4-10). The Smad2X-independent genes were presumably mediated by Smad2/3 proteins in the *Smad2*-Ex3$^{-/-}$ cells. These results suggest that Smad2X is required for the transcriptional activation of a specific subset of Activin response genes, and imply that Smad2X and Smad2/3 play distinct functions during Nodal/Activin-driven cellular differentiation.
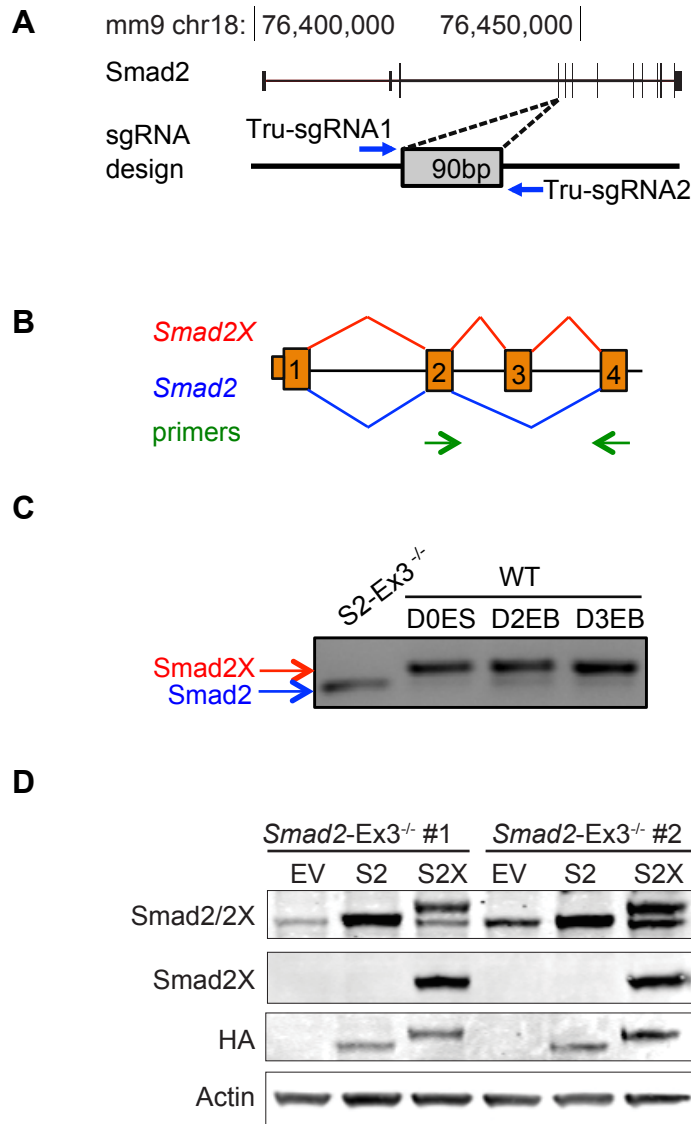
Figure 4-4. CRISPR/Cas9-mediated *Smad2* exon3 knockout and validation. (**A**) Tru-sgRNA design for CRISPR/Cas9-mediated deletion of *Smad2* Exon3. (**B**) PCR primer spanning exon3 region for genotyping analysis. (**C**) PCR validation of exon3 deletion in WT ES cells or EBs, and *Smad2*-Ex3$^{-/-}$ (S2-Ex3$^{-/-}$) ES cells. The PCR products for Smad2X (*red*) and Smad2 (*blue*) are highlighted. (**D**) Two *Smad2*-Ex3$^{-/-}$ clones #1 and #2 were derived from CRISPR experiment. Both #1 and #2 cells were transduced with HA-hSMAD2/2X expressing vectors or empty vector.
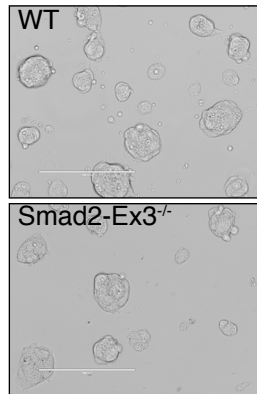
Figure 4-5.   Brightfield imaging showing the morphologies of wild type (WT) or *Smad2*-Ex3[-/-] ES cell colonies maintained under LIF+serum condition.
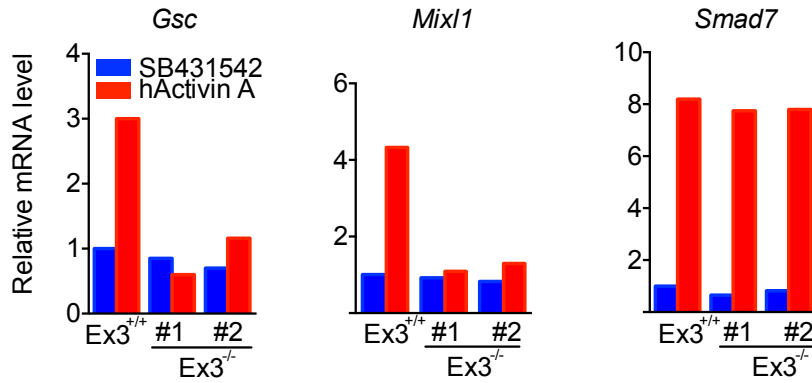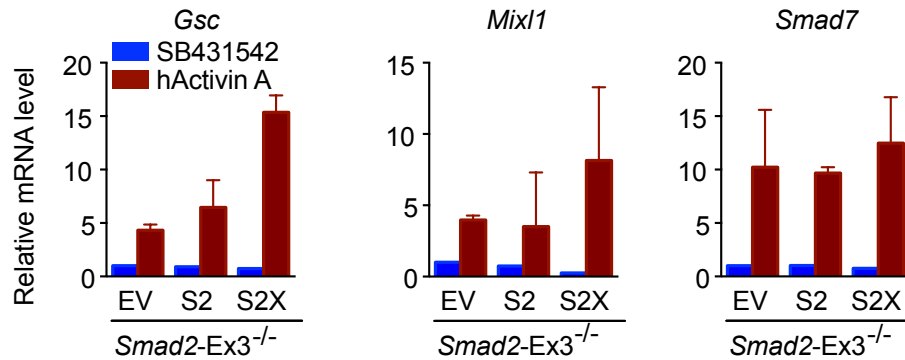
Figure 4-6. (**A**) qRT-PCR analysis of *Gsc*, *Mixl1* and *Smad7* mRNA levels in Wild type (Ex3$^{+/+}$) or *Smad2*-Ex3$^{-/-}$ (Ex3$^{-/-}$) day 4 EBs treated with SB431542 or human Activin A. (**B**) qRT-PCR analysis of *Gsc*, *Mixl1* and *Smad7* mRNA levels in *Smad2*-Ex3$^{-/-}$ ES cells transduced with HA tagged human *SMAD2* (S2), *SMAD2X* (S2X) genes or empty vector (EV). day 4 EBs from the indicated cells were treated with SB431542 or human Activin A and collected for gene expression analysis.

A

B

RNA-Seq analysis:
Activin vs SB431542
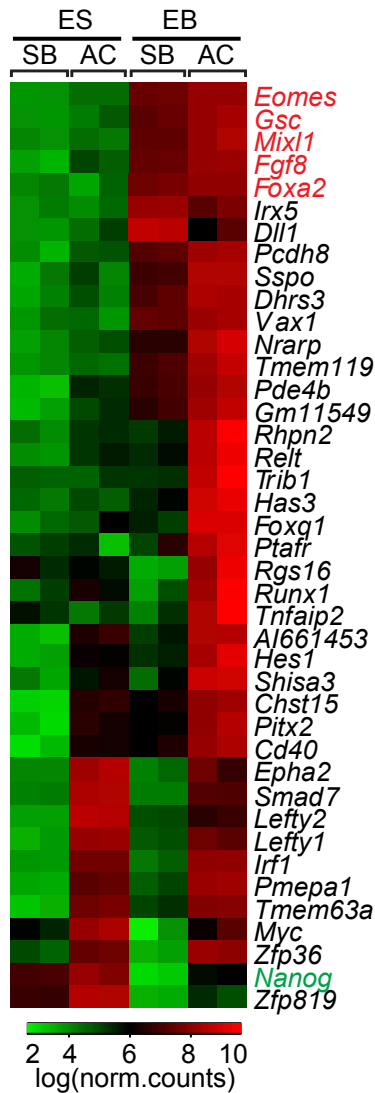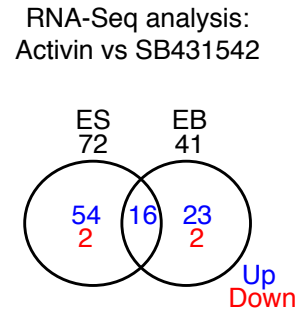
Figure 4-7. RNA-Seq analysis of Activin response genes in ES cells and EBs. (**A**) Heatmap of the expression level of 41 genes that respond to Activin A under EB condition. Normalized read counts from RNA-Seq are used in this plot. Mesendoderm specification genes (*red*) and pluripotency-assicated genes (*green*) are highlighted. (**B**) Venn Diagram showing RNA-seq analysis results of Activin response genes in embryonic stem cells (ES) or day 3 EBs. ES cells or EB cells are either treated with 50ng/ml of recombinant Activin A for 90 min or 10μM SB431542 for 2h. Threshold for differential expression analysis are: adjusted p value < 0.05 and fold change >2 (up, show in blue) or <0.5 (down, show in red).
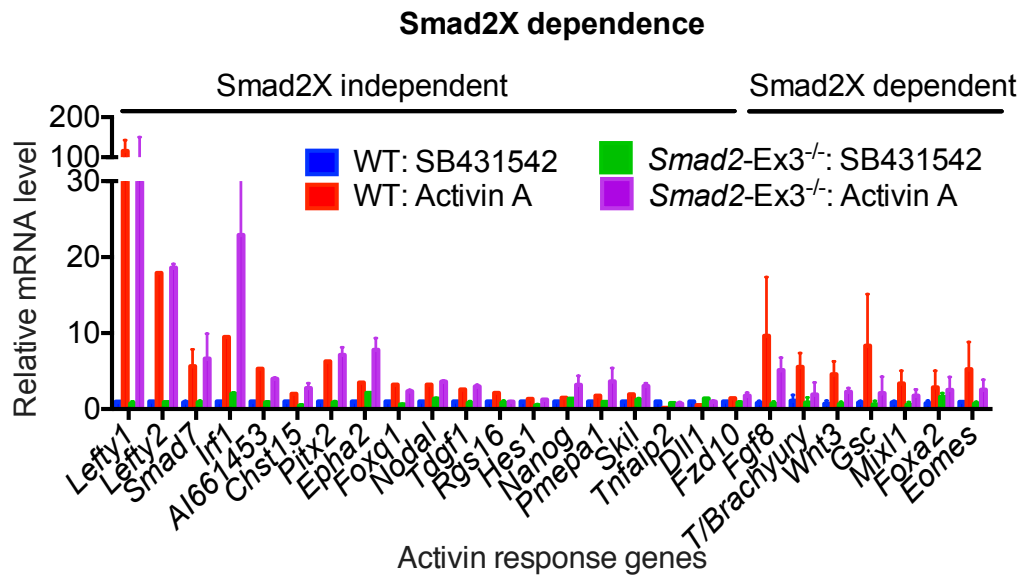
Figure 4-8. qRT-PCR analysis of Smad2X-dependent Activin response genes. mRNA expression levels of Activin response gens were measured in wild type (WT) and *Smad2*-Ex3[-/-] day 4 EBs treated with SB431542 or Activin.
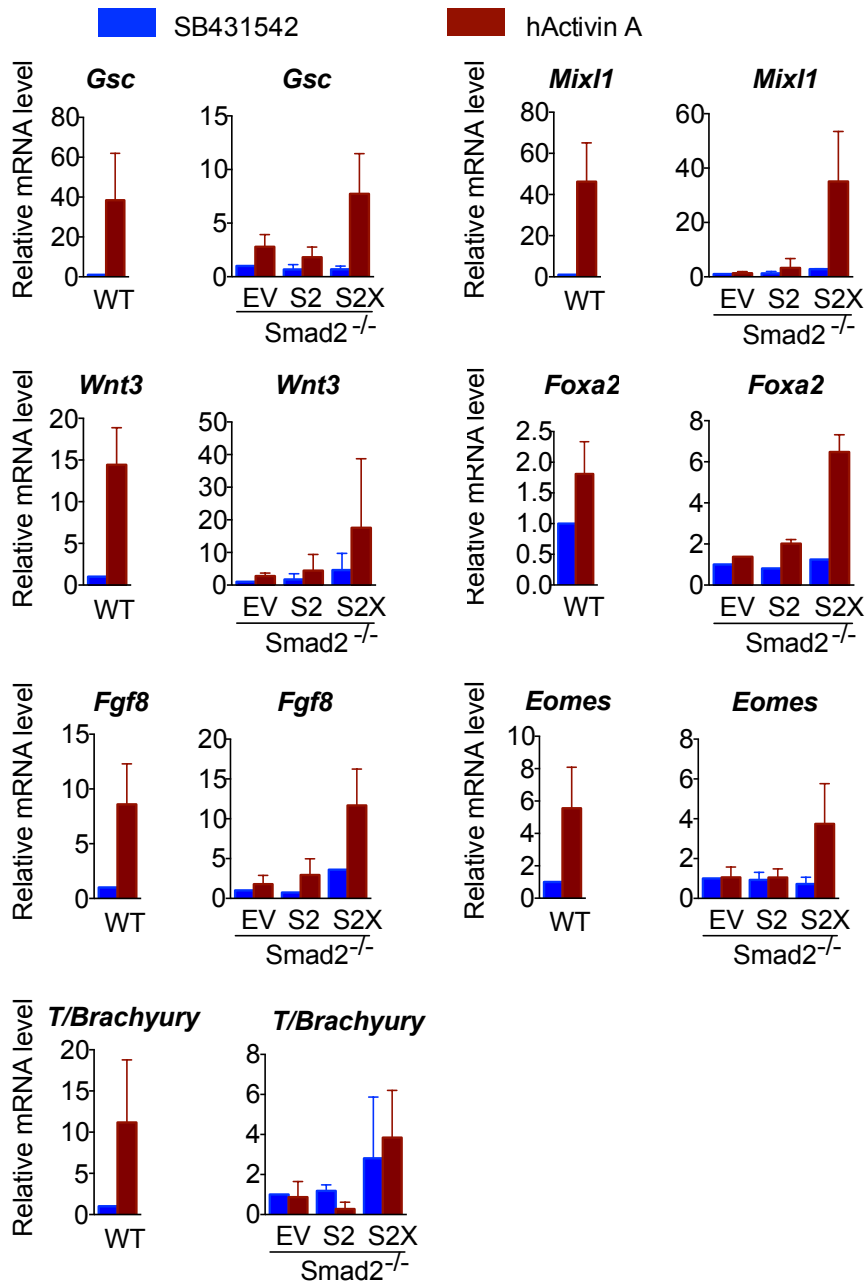
Figure 4-9.   qRT-PCR analysis of Smad2X-dependent genes in wild type (WT) or KT15 (*Smad2*[-/-]) ES cells transduced with HA-hSMAD2/2X or empty vector. Day 4 EBs were treated with SB431542 or hActivin A before harvest for mRNA analysis. The WT ES cells were derived from a different genetic background and are plotted in separate graphs to the results from the KT15 cells.
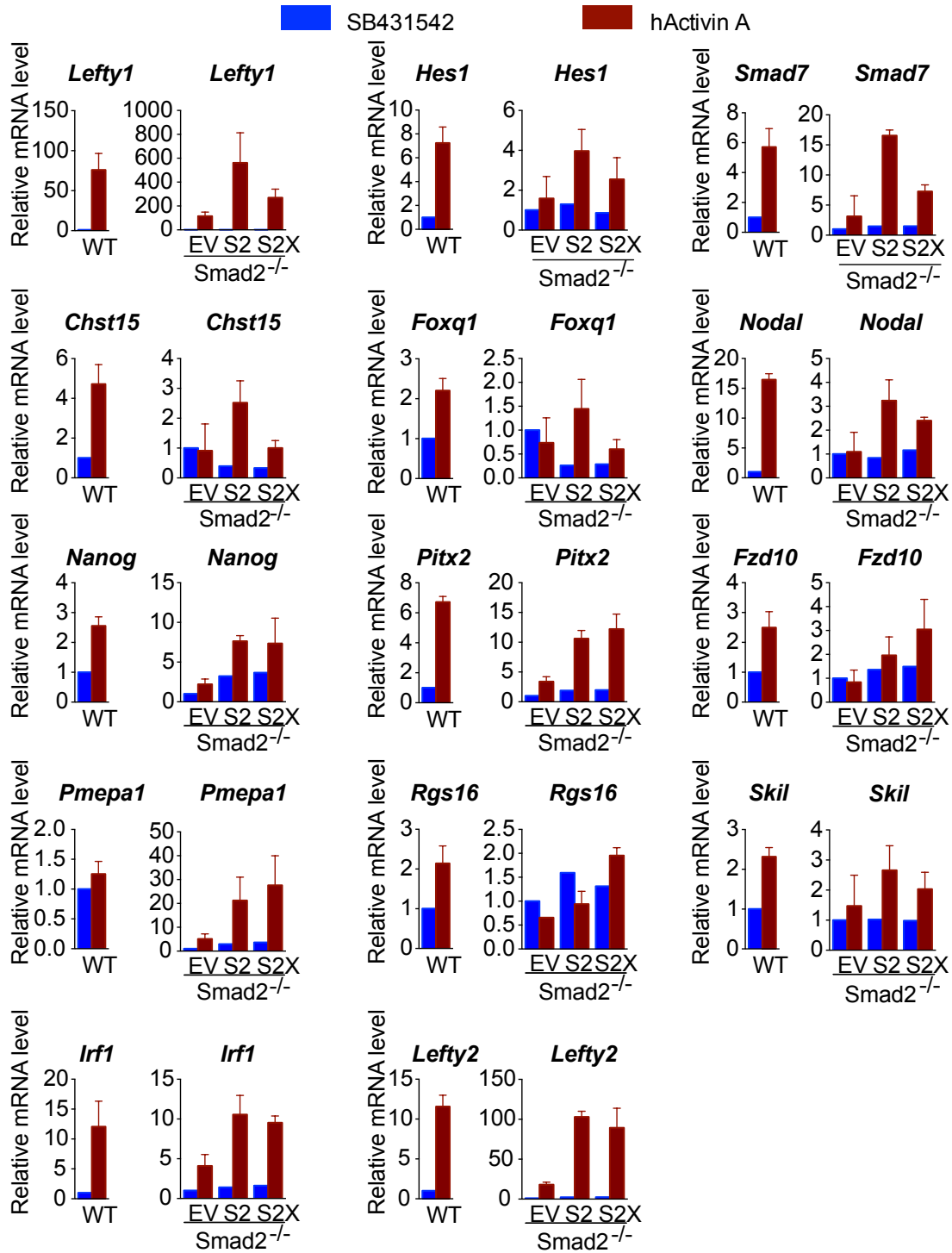
Figure 4-10. qRT-PCR analysis of Smad2X-independent genes in wild type (WT) or KT15 (*Smad2*[-/-]) ES cells transduced with HA-hSMAD2/2X or empty vector. Day 4 EBs were treated with SB431542 or hActivin A before harvest for mRNA analysis. The WT ES cells were derived from a different genetic background and are plotted in separate graphs to the results from the KT15 cells.

4.3.3   Smad2X cooperates with Nodal co-activators

To determine the components of the transcriptional complexes responsible for Smad2X-dependent genes, I determined the requirements of known co-factors of Nodal/Activin signaling. In response to Nodal/Activin stimulation and receptor activation, Smad2/2X/3 cooperate with co-Smad Smad4, lineage-restricted transcription factor FoxH1 and histone H3K9me3-H3K18ac reader Trim33 to regulate mesendoderm differentiation (Chen et al., 1996; Chen et al., 1997; Xi et al., 2011). By RNA-Seq analysis, type I Activin receptors *Acvr1* and *Acvr1b*, type II receptors *Acvr2a*, *Acvr2b*, Smad effectors *Smad2/2X*, *Smad3*, *Smad4*, and co-factors *Foxh1* and *Trim33* are constantly and highly expressed in ES cells and EBs (Figure 4-11).

To determine what factors are involved in Activin stimulated mesendoderm differentiation, I performed qRT-PCR analysis (and RNA-Seq analysis) of Activin or SB431542 treated, day 3 EBs derived from a series of genetically modified cell lines including *Smad4*[-/-], *Foxh1*[-/-] and shRNA-mediated *Trim33* knockdown (E14-pLKO-sh*Trim33*). In the absence of Smad4, expression of the majority of Activin response genes including *Gsc*, *Smad7*, *Pmepa1*, *Wnt3*, *Mixl1*, *Fgf8*, *Chst15*, *Foxa2*, *Hes1*, *Eomes* and *Nanog* were compromised (Figure 4-12), while a few genes remain responsive to Activin treatment, including *Lefty1*, *Lefty2*, *Pitx2* and *Fzd10* (Figure 4-12). The expression levels of these Smad4-dependent genes were highest in the pluripotent stage, and steadily declined during EB

differentiation (refer to Chapter 3: Figure 3-1, 3-2). In *Trim33* knockdown cells, Activin stimulated induction of *Lefty1*, *Foxa2*, *Gsc*, *Pmepa1*, *Nanog*, *Chst15*, *Mixl1*, *Fgf8* and *Eomes* were at least partially compromised (Figure 4-13). Genes including *Lefty2*, *Smad7*, *Pitx2*, *Hes1* and *Fzd10* were insensitive to Trim33-depletion (Figure 4-13).

In the absence of FoxH1, a large group of Activin response genes lost their expression (Figure 4-14). These genes include *Lefty1*, *Lefty2*, *Gsc*, *Mixl1*, *Pitx2*, *Fgf8*, *Foxq1*, *Eomes*, *Foxa2*, *Chst15*, *Wnt3*, *Nodal*, *Fzd10*, *T/Brachyury* (Figure 4-14). In contrast, another subset of Activin response genes including *Smad7*, *Irf1*, *Pmepa1*, *Hes1*, *Epha2*, *Nanog* and *Skil* were induced by Activin regardless of *Foxh1* status (Figure 4-14). Though FoxH1 is specifically required for mesendoderm specification, it is expressed in the pluripotent ES cell condition albeit at lower levels than that in the EBs (Figure 4-11). I investigated whether FoxH1 is required for Nodal-stimulated gene responses in the ES cells by qRT-PCR. Results showed that expression of *Lefty1*, *Lefty2*, *Pitx2* and *Fzd10* in the pluripotent ES cells indeed require FoxH1 function (Figure 4-14). This data is consistent with the requirement of FoxH1 in guiding the establishment of left-right asymmetry. Intriguingly, this group of FoxH1-dependent genes responded to Activin stimulation regardless of *Smad4* status (Figure 4-12).

As determined by RNA-Seq analysis (refer to Chapter3, Figure 3-7; and ChIP-Seq analysis of HA-FLAG-β-Catenin in day 3 EBs), expression of a subset

of Activin response genes including *Gsc*, *Mixl1*, *Fgf8*, *Foxa2*, *Wnt3*, *Eomes* and *T/Brachyury* is dependent on p53/p73 proteins. These genes require activated Wnt-Tcf/β-Catenin downstream of p53/p73 for transcriptional activation.

In summary (Figure 4-20), genes including *Gsc*, *Mixl1*, *T/Brachyury*, *Foxa2*, *Eomes* and *Fgf8* are dependent on Smad2X, Smad4, Trim33, FoxH1 and p53. Genes including *Lefty1*, *Lefty2*, *Pitx2* and *Fzd10* are dependent on Smad2/2X/3, FoxH1, at least partially independent of Smad4, Trim33 and p53, and not exclusively dependent on Smad2X. Genes including *Smad7*, *Skil*, *Pmepa1* and *Irf1* are dependent on Smad4, Smad2/2X/3, not exclusively dependent on Smad2X (Figure 4-20). The significant overlap between genes that require Smad2X and other co-factors suggest that Smad2X mediates mesendoderm specification together with transcription factors FoxH1 and Smad4, histone modification H3K9me3-H3K18ac reader protein Trim33 as well as Tcf3/β-Catenin effector complex downstream of p53/p73. The cooperative interactions between Smad2X and co-factors imply that Smad2X serve as a recruitment factor for other molecular players in mediating Nodal transcriptional response.
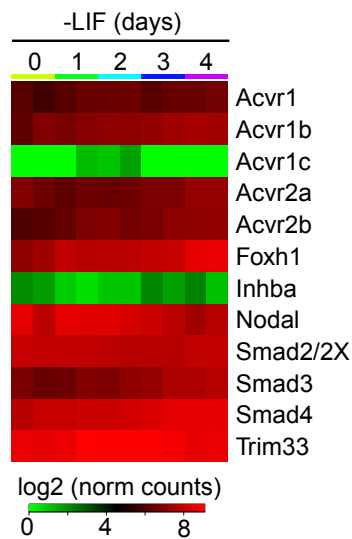
Figure 4-11. Heatmap representing the mRNA expression dynamics of Nodal/ Activin pathway components Nodal, Activin (Inhba), Activin receptor family (Acvr1/1b/1c/2a/2b), Smad2/3/4, Trim33 and FoxH1 during four day EB differentiation. Acvr1, encoding Activin Receptor-Like Kinase 2 or Alk2, transduces BMP signaling. Acvr1b, encoding Alk4 (or Activin Receptor type IB), and Acvr1c, encoding Alk7 (or Activin Receptor type IC) function as the type I transmembrane receptors for Nodal and Activin. Acvr2a, encoding ActrIIA (Activin Receptor Type IIA) and Acvr2b, encoding ActrIIB (Activin Receptor Type IIB) function as the type II transmember receptors for Nodal and Activin.

Smad4 dependency: day 3 EB
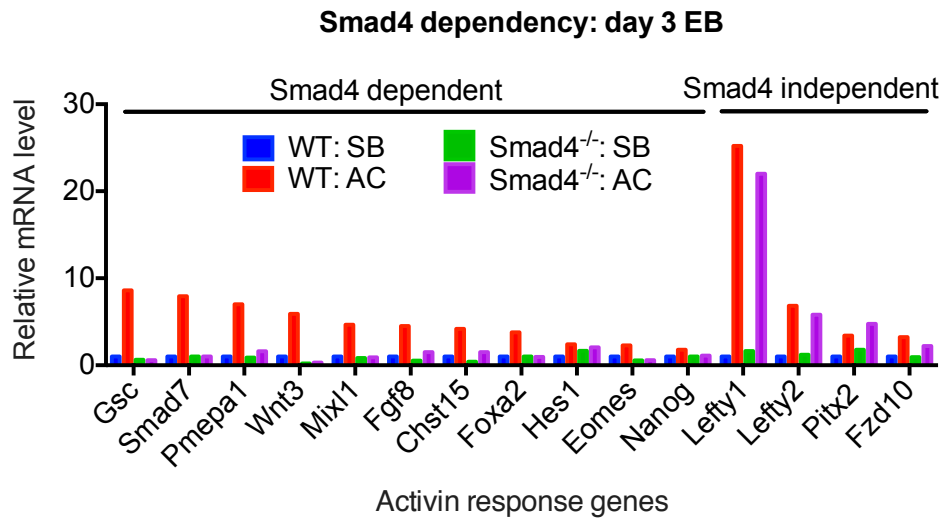
Figure 4-12. qRT-PCR analysis of Activin response genes in wild type (WT) or *Smad4*[-/-] cells treated with SB431542 (SB) or Activin A (AC).
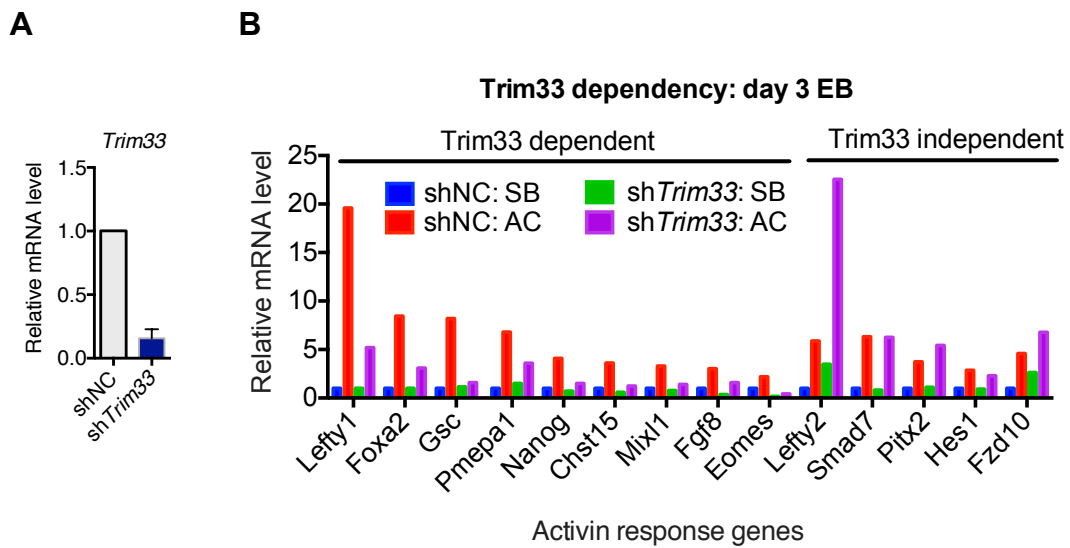
Figure 4-13. (**A**) qRT-PCR validation of *Trim33* mRNA knockout down by shRNA. (**B**) qRT-PCR analysis of Activin response genes in wild type (WT) or shTrim33 cells treated with SB431542 (SB) or Activin A (AC).
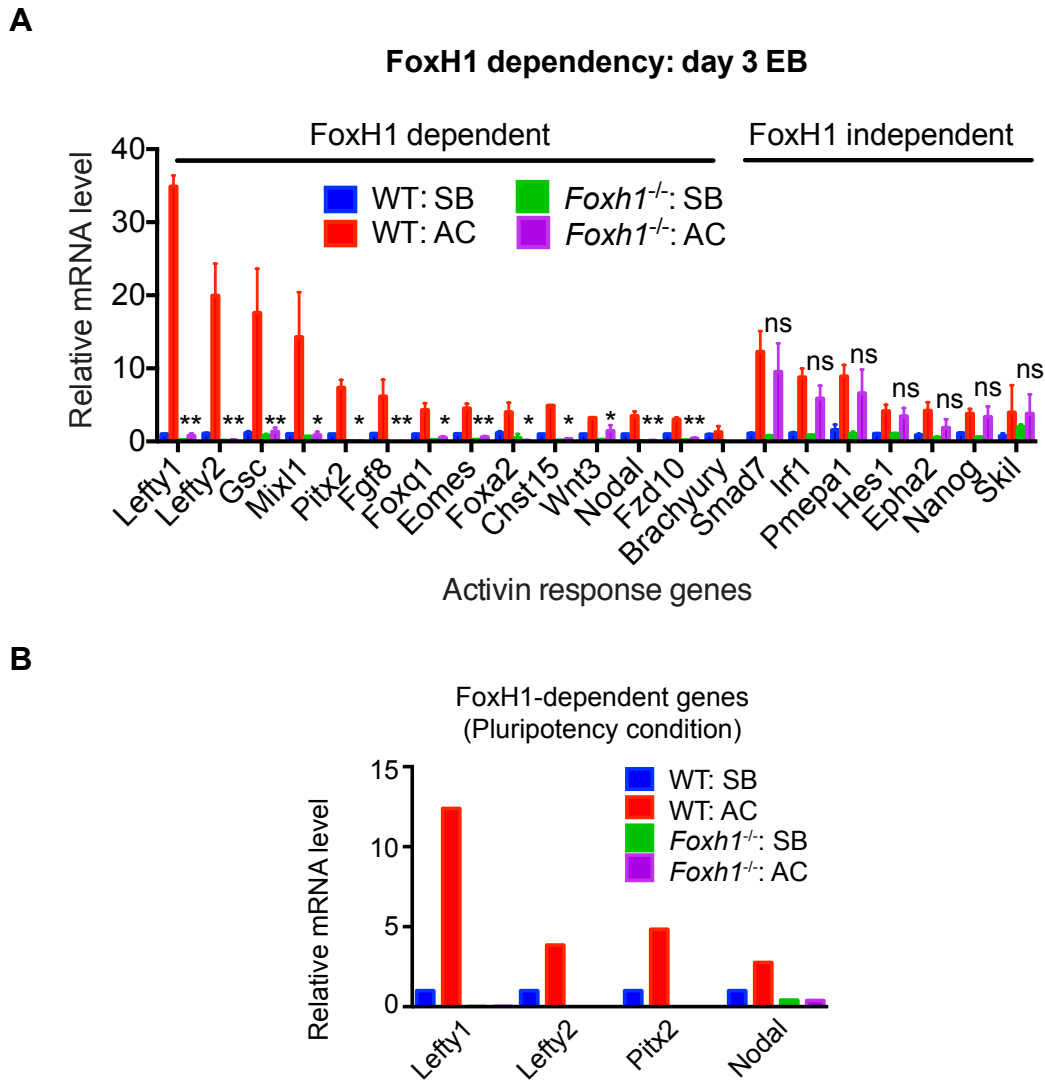
Figure 4-14. (**A**) qRT-PCR analysis of Activin response genes in wild type (WT) or *Foxh1*[-/-] day 3 EBs treated with SB431542 (SB) or Activin A (AC). (**B**) qRT-PCR analysis for the mRNA expression levels of *Lefty1*, *Lefty2*, *Pitx2* and *Nodal* in the wild type (WT) and *Foxh1*[-/-] ES cells.

4.3.4 Smad2X recruits Nodal response co-regulators to mesendoderm enhancers

To study the role of Smad2X in regulating transcription and the interaction between Smad2X and other co-activators on the chromatin, I performed ChIP-Seq analysis of Smad2X using the Smad2X-specific antibody. For comparison, ChIP-Seq was performed for total Smad2/2X/3, Smad4 and Trim33 in ES cells and day 2.5 EBs treated with Activin A or SB431542. FoxH1 antibodies gave unsatisfying results in ChIP-Seq.

The number of binding sites in each condition is summarized (Figure 4-15). 6267 Smad2/2X/3 binding sites were identified genome-wide, with > 98% of which are Activin signaling dependent (Figure 4-15) (See also Chapter 3). In the 41 Nodal response genes identified by RNA-Seq, 39 genes exhibit Smad2/2X/3 binding at nearby loci. The Smad2/2X/3 unbound genes are likely to be indirect target genes of Nodal signaling. In Smad2X ChIP-Seq, 1473 peaks were identified, majority of which overlap with Smad2/2X/3 binding sites and are Activin-dependent (Figure 4-15, 4-16). In Smad4 ChIP-Seq, 523 peaks in Ac condition and 136 peaks in SB condition were identified, with 135 overlapped peaks (Figure 4-15). In Trim33 ChIP-Seq, 4236 peaks in the AC condition and 6506 peaks in the SB condition were identified, with 3581 overlapping peaks (Figure 4-15).

I noticed that majority of the Smad2/2X/3/4 binding events under Activin treatment were abolished or significantly reduced by SB431542 treatment, suggesting that the ChIP antibody is of high quality and the major chromatin-binding activity of Smad2/2X/3/4 is driven by Activin/Nodal signaling. The binding sites of Trim33 and Smad4 exhibited higher levels of retention rates after Activin signaling inhibition, suggesting that these two factors might possess Smad2/2X/3 independent functions (Figure 4-15, 4-16). Consistent with the active participation of Smad2X in recruiting all Nodal response co-factors, Smad2X, Smad2/2X/3, Smad4 and Trim33 ChIP-Seq significantly overlap with each other (Figure 4-15, 4-16). These results suggest that Smad2X serves as a nucleating factor for Nodal stimulated transcriptional responses.

To determine the regulatory functions of Smad2/2X/3/4-Trim33 binding sites in transcription, I annotated each peak with the nearest gene and determined the relative position by categorizing the genomic regions into 7 groups: Promoter-TSS (-1kb to +100bp from transcription start site (TSS)), Exon, Intron, Intergenic, 3' UTR, TTS (-100bp to +1kb from transcription termination site (TTS)), 5' UTR. Compared with random genomic distribution, all factors show significant enrichment at the Pomoter-TSS regions. In addition, majority of Smad2/2X/3/4-Trim33 peaks (> 85%) are located in exons/introns or intergenic regions, many of which are presumably enhancer sites (Figure 4-15).

To understand the mechanisms for selective activation of target genes,

I analyzed the Smad complex occupancy in genomic regions surrounding lineage-specification factors such as *Gsc, Eomes, Fgf8* and *Lefty1* (Figure 4-16). Besides the promoter binding, most of these target genes are bound at proximal or distal enhancers with distance to TSS up to 50kb. For instance, *Gsc* is bound at the proximal promoter site (~0.4kb upstream from TSS) and an enhancer site (~6kb downstream from TSS) (Figure 4-16; refer to Chapter 3 Figure 3-19). DNA sequences within these *cis* elements are conserved during vertebrate evolution (refer to Chapter 3 Figure 3-21, 3-22, 3-23) and the binding pattern is largely maintained in Smad2/2X/3/4 ChIP-Seq analysis of definitive endodermal cells differentiated from human ES cells (Brown et al., 2011; Kim et al., 2011; Yoon et al., 2011).

To further determine the activity of these enhancers during the transition from pluripotency to differentiation, I performed ATAC-Seq analysis to assess the chromatin accessibility in ES cells and EBs. Enhancers adjacent to pluripotency genes such as *Nanog* and *Pou5f1* are in the open state in ES cells and turn closed during EB differentiation (Figure 4-17). In contrast, majority of the mesendoderm gene enhancers are at an inaccessible state during pluripotency, and they acquire accessibility in EBs (Figure 4-17). Collectively, the collaborative binding by co-factors at these enhancer sites implied that enhancer binding is a critical step in Smad mediated transcription activation. Indeed, CRISPR mediated mutagenesis at these enhancer sites significantly compromised adjacent mesendoderm gene expression (see Chapter 3 Figure 3-21, 2-22).

A
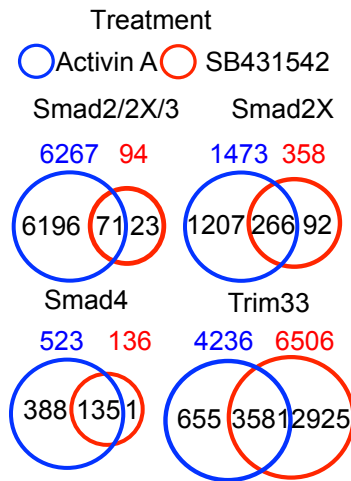
B
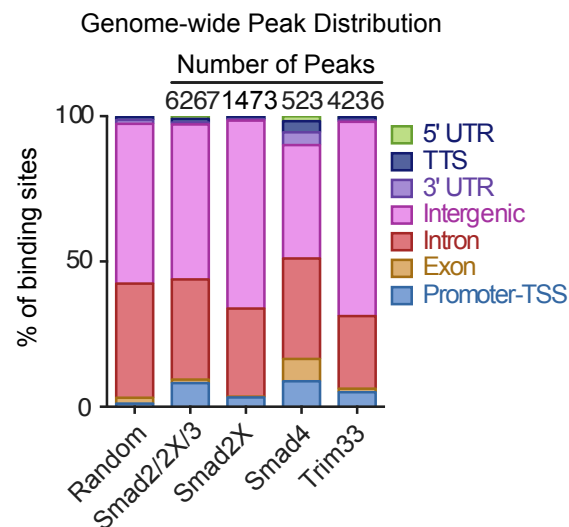
Figure 4-15. ChIP-Seq analysis of Smad2/2X/3, Smad2X, Smad4 and Trim33 in EBs. (**A**) Venn diagram summarizing the numbers of ChIP-Seq peaks identified in indicated conditions. (**B**) Bar graph representing the Genome-wide distribution of peaks. Binding sites are categorized into 7 groups according to there relative position to the nearest gene: Promoter-TSS (-1kb to +100bp from transcription start site (TSS)), Exon, Intron, Intergenic, 3' UTR, TTS (-100bp to +1kb from transcription termination site (TTS)) and 5' UTR. The percentages of peaks in each category are plotted. Random genome distribution is used as a control.
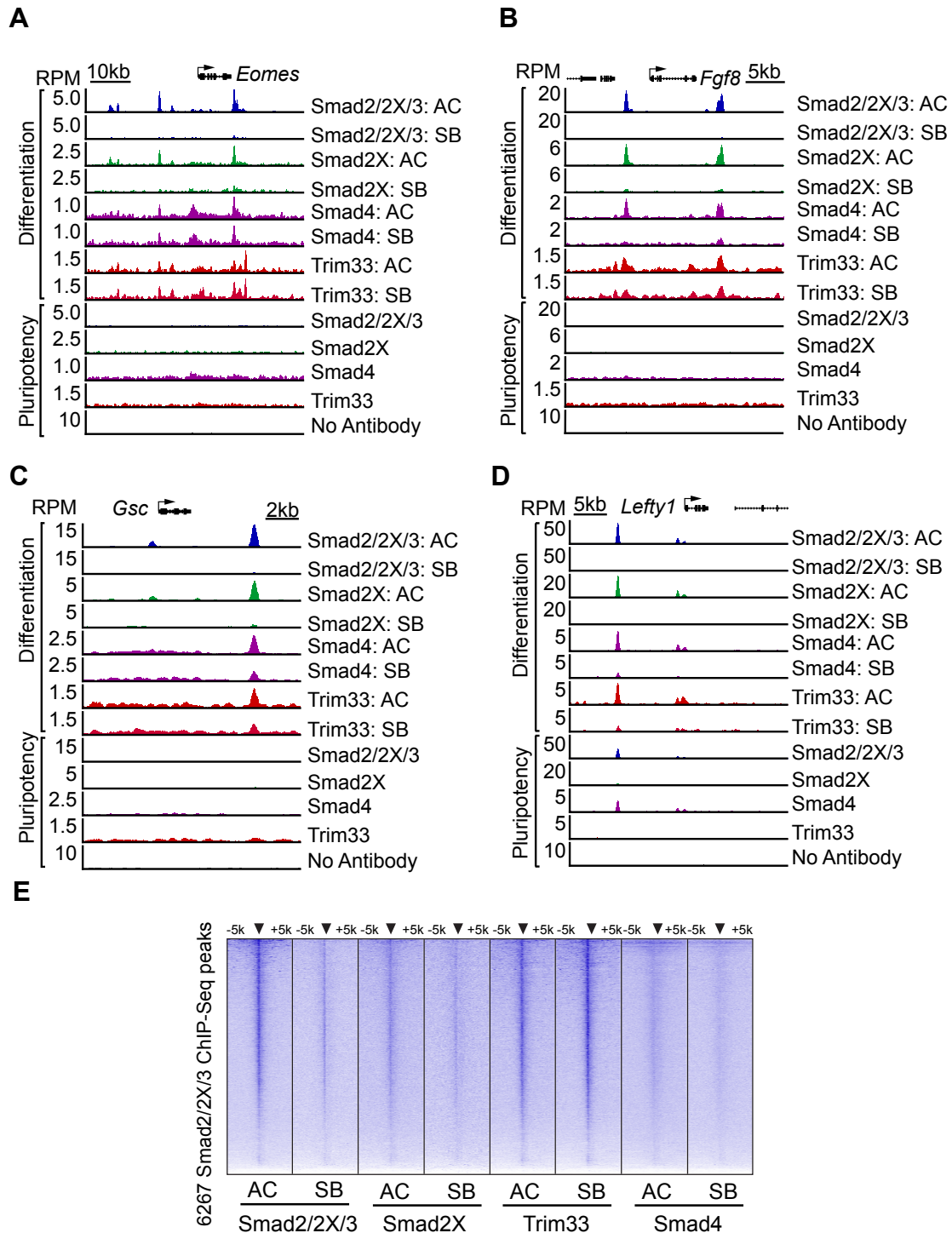
Figure 4-16. ChIP-Seq gene track view for Smad2/2X/3 (blue), Smad2X (green), Smad4 (purple), Trim33 (red) binding to *Eomes* (**A**), *Fgf8* (**B**), *Gsc* (**C**) and *Lefty1* (**D**) loci in ES cells (pluripotency) or EBs (Differentiation) treated with SB431542 (SB) or Activin A (AC) for 2h. (**E**) Heatmap for Smad2X, Smad2/2X/3, Trim33 and Smad4 ChIP-Seq tag density from +/- 5kb regions centered on 6267 Smad2/2X/3 binding sites.
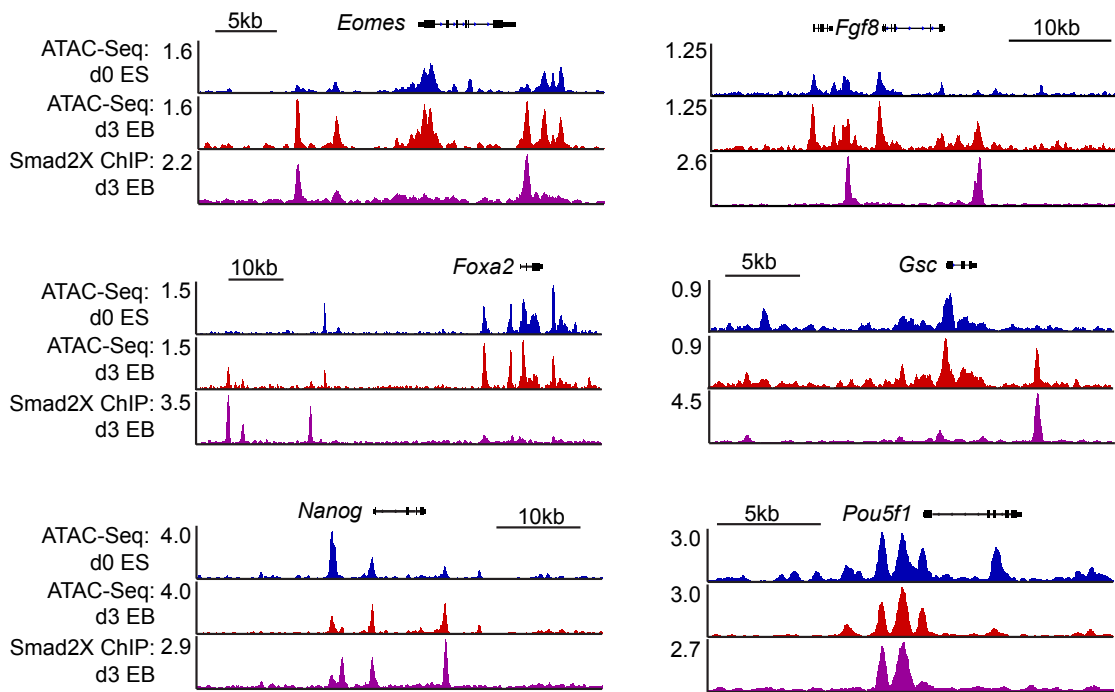
Figure 4-17. ATAC-Seq analysis for the chromatin accessibility of representative mesendoderm differentiation genes and pluripotency-associated genes in d0 ES cells or d3 EBs. Smad2X ChIP-Seq in d3 Activin treated EBs is included as reference to mark the Nodal-responsive enhancer sites.

4.3.5 FoxH1 is required for Smad2X recruitment to mesendoderm genes

Previous studies suggested that Smad2/3 directly binds to DNA sequences designated as "Smad binding element" (SBE, CAGAC/T or its reverse complement sequence) whereas Smad2X did not appear to be capable of DNA binding. To confirm this observation, I performed electrophoretic mobility shift assays (EMSA) with purified recombinant Smad2/2X-MH1 domain proteins and palindromic SBE or GC-rich DNA sequences (refer to Conclusion and Discussion 4.4.2 for rationale to test GC-rich sequences). Unlike Smad2-MH1 domain that binds to GC-rich and palindromic SBE sequences with 5-20nM affinities, Smad2X-MH1 exhibited only moderate binding at 2-5 μM affinities (Figure 4-18). These results are consistent with previous reports and establish that Smad2X-MH1 domain has 1-2 order of magnitude lower DNA binding capacity than Smad2/3 MH1 domains. Of note, the ability of Smad2-MH1 to bind GC-rich sequences effectively supports the similarity between the target DNA preferences of Smad4 and Smad2/3 (refer to Discussion 4.4.2). It is unclear whether dimeric Smad2/Smad4, Smad3/Smad4 complexes and trimeric Smad2/3/4 complexes are structurally and functionally symmetric when bound to target DNA sequences.

The lack of intrinsic DNA binding capacity suggests that Smad2X recruitment to target genes is mediated by other DNA-binding transcription factors. In *in vitro* reporter assays, Smad2X, Smad4 and FoxH1 assemble into

the Activin-responsive factor (ARF) (Chen et al., 1996; Chen et al., 1997). ARF is necessary and sufficient to drive *Xenopus Mix.2* and *Gsc* promoter reporter activation. Smad2X MH2 domain is required for interactions with both Smad4 and FoxH1, whereas there is no directly contact between Smad4 and FoxH1. FoxH1 overexpression forces Smad2X to enter the nucleus (Xu et al., 2002). I determined the Smad2X binding pattern in the absence of FoxH1 by ChIP-Seq. Results suggest that FoxH1-depletion blocked the recruitment of Smad2X to the majority of Nodal response genes including mesendoderm genes *Eomes*, *Gsc*, *Fgf8*, etc, and left-right axis specification genes such as *Lefty1*, *Lefty2* and *Pitx2* (Figure 4-19). The genes that exhibit alterations in Smad2X binding are fully consistent with the genes that require FoxH1 for transcriptional activation (Figure 4-19, 4-14). Smad2X bound to FoxH1-independent, Nodal/Activin feedback genes such as *Smad7*, *Skil* and *Pmepa1* regardless of FoxH1 status.

Interestingly, none of *Smad7*, *Skil* and *Pmepa1* is exclusively dependent on Smad2X for expression, suggesting that the binding by Smad2/3 rather than Smad2X is key for the transcriptional activation of these target genes. Smad2X at these gene loci is potentially recruited by Smad2/3/4 through oligomerization. Of note, in FoxH1-dependent but Smad4-independent genes *Lefty1*, *Lefty2* and *Pitx2*, Smad2X chromatin binding is exclusively dependent on FoxH1. Together these results suggest that chromatin binding of Smad2X is relying on either FoxH1 or Smad4 in ES cells and EBs. Future investigation will determine whether depleting both FoxH1 and Smad4 could completely block Smad2X

binding to chromatin in the context of ES cell differentiation. Collectively, a comprehensive view about the requirements of Nodal response genes on Smad2X and co-factors is summarized (Figure 4-20).

GC1 (40bp)
5'-GAAAAGGTAAGCCGTCGGGGCCGCGCCGGGGCCGGAACGA-3'

Palindromic SBE (16bp)
5'-AGTATGTCTAGACTGA-3'

Smad2-MH1 (Nano-Molar range)

nM 0 10 20 40 80 160 320 640 1280 2560   nM 0 10 20 40 80 160 320 640 1280 2560

Smad2X-MH1 (Nano-Molar range)

nM 0 10 20 40 80 160 320 640 1280 2560   nM 0 10 20 40 80 160 320 640 1280 2560

Smad2X-MH1 (Micro-Molar range)

uM 0 .08 .16 .32 .64 1.3 2.5 5 10 20   uM 0 .08 .16 .32 .64 1.3 2.5 5 10 20

Figure 4-18. EMSA analysis of purified recombinant Smad2-MH1 and Smad2X-MH1 proteins binding to synthesized GC1 or palindromic SBE DNA oligos. The protein concentrations (in nM or μM) in each lane are indicated.

185
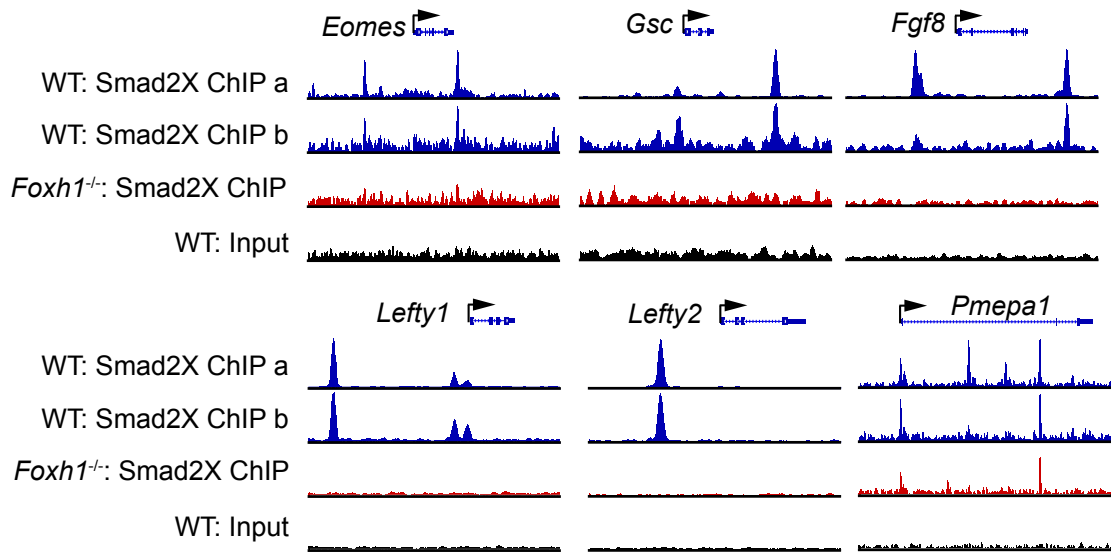
Figure 4-19. ChIP-Seq gene track view of Smad2X binding to *Eomes*, *Gsc*, *Fgf8*, *Lefty1*, *Lefty2* and *Pmepa1* loci in wild type (WT) or *Foxh1*[-/-] day 3 EBs.

Target genes	Smad2/2X/3-bound	Smad4-bound	Smad4-dependent	Smad2X-bound	Smad2X-dependent	FoxH1-dept Smad2X binding	FoxH1-dependent	Trim33-bound	Trim33-dependent	p53-bound	p53/p73-dependent	Tcf3-bound	β-catenin-bound
Gsc	Y	Y	Y	Y	Y	Y	Y	Y	Y	N	Y	Y	Y
Wnt3	Y	Y	Y	Y	Y	Y	Y	Y	na	Y	Y	Y	Y
Eomes	Y	Y	Y	Y	Y	Y	Y	Y	Y	N	Y	Y	Y
Foxa2	Y	Y	Y	Y	Y	Y	Y	Y	Y	N	Y	Y	Y
Mixl1	Y	Y	Y	Y	Y	Y	Y	Y	Y	N	Y	Y	na
Fgf8	Y	Y	N	Y	Y	Y	Y	Y	Y	N	Y	Y	Y
T/Brachyury	Y	Y	na	Y	Y	Y	Y	Y	na	N	Y	Y	Y
Lefty1	Y	Y	N	Y	N	Y	Y	Y	N	N	N	Y	Y
Lefty2	Y	Y	N	Y	N	Y	Y	Y	N	N	N	Y	Y
Pitx2	Y	Y	N	Y	N	Y	Y	Y	N	N	N	Y	Y
Fzd10	Y	Y	N	Y	N	Y	Y	na	N	N	N	Y	Y
Tdgf1	Y	Y	Y	Y	N	Y	Y	Y	na	N	Y	Y	Y
Chst15	Y	Y	Y	Y	N	Y	Y	Y	N	Y	N	Y	Y
Smad7	Y	Y	Y	Y	N	N	N	Y	N	N	N	Y	na
Ctgf	Y	Y	Y	Y	na	N	N	Y	na	N	N	Y	Y
Skil	Y	Y	Y	Y	N	N	N	Y	N	N	N	Y	N
Pmepa1	Y	Y	Y	Y	N	N	N	Y	N	N	N	Y	N
Relt	Y	Y	Y	Y	na	N	N	Y	N	N	N	Y	N
Hes1	Y	Y	Y	Y	N	N	N	na	N	N	N	Y	N
Nanog	Y	Y	Y	Y	N	N	N	Y	Y	Y	N	Y	Y

Figure 4-20. Summarized view about the requirement of Nodal response genes on Smad2X and co-factors. Protein factor binding data are extracted from Smad2/2X/3, Smad2X, Trim33, p53, Tcf3 and β-Catenin ChIP-Seq analysis. Gene expression dependencies were determined by qRT-PCR and RNA-Seq. Selected genes are highlighted in color: mesendoderm specification genes (orange shade), left-right axis specification genes (cyan shade), Nodal/Activin feedback regulatory genes (yellow shade), and pluripotency associated genes (blue shade). Genes marked in red did not appear in Figure 4-7 RNA-Seq analysis but however they were repetitively detected as Activin induced genes in other experiments. Y (Yes): exhibiting binding or dependency; N (No), no binding or dependency; na: data not available or not applicable.

4.4 Conclusions and Discussion

In this study, I investigated the transcriptome and cistrome of Activin/ Nodal signaling driven differentiation and the mechanisms regulating the tightly controlled Nodal responses. I demonstrated that Smad2X, a splicing isoform transcribed from the *Smad2* gene, mediates Nodal/Activin-driven mesendoderm differentiation in ES cells. Smad2X and Smad3 are co-expressed in ES cells and EBs while Smad2X is the most abundant isoform among Smad2/2X/3 proteins. Smad2X responds to Nodal signaling and recruits co-activators including Smad4, FoxH1 and Trim33 to assemble a transcription activation complex (Figure 4-20). The DNA binding capacity of Smad2X is suppressed by the exon3, rendering it a flexible nucleating factor for cooperating with DNA-binding transcription co-factors. Thus diversification of R-Smads through alternative splicing is potentially a novel mechanism for modulating the Nodal/Activin target gene specificity.

4.4.1   Smad2X in development

Transgenic mouse studies recognized that *Smad2* gene but not *Smad3* is required for the gastrulation process of mouse development (Ashcroft et al., 1999; Nomura and Li, 1998; Tremblay et al., 2000; Waldrip et al., 1998; Weinstein et al., 1998; Yang et al., 1999b). My analysis showed that Smad2X accounts for majority of the protein products from *Smad2* gene, and it is plausible

that Smad2X directly mediates Nodal signaling during mouse gastrulation. My investigation with the role of Smad2X in ES cells and EBs reinforced this concept. Homozygous knockout of exon3 in the mouse should be performed to further support the role of Smad2X in development. Indeed, it was shown previously that mice exclusively expressing Smad2 short isoform (with targeted disruption of exon3) are viable and fertile (Dunn et al., 2005). However, this experiment is perplexing in several aspects: (1) In the Dunn et al. study, $Smad2^{-/-}$ mice could be rescued by knock-in alleles of neither human FLAG-$SMAD2$ nor FLAG-$SMAD2X$, albeit rescuing was demonstrated with FLAG-$SMAD2$ overexpression driven by constitutive active CAG promoter. (2) $Smad2^{-/-}$ mice, which retain wild type Smad3, is embryonic lethal; whereas the human $SMAD3$ knockin to mouse $Smad2$ locus, which maintains wild type mouse Smad3 alleles also, partially rescue the developmental deficiency. Why does Smad3 protein in the $Smad2^{-/-}$ background not rescue the mice? (3) Exon3 deletion by genome editing spontaneously leads to dramatically elevated levels of $Smad2$ short isoform. It is unclear what the consequences of this alteration are. With these concerns, characterizing the ability of $Smad2$-Ex3$^{-/-}$ in chimeric embryo contribution is required to clarify the requirement of Smad2X in development.

4.4.2   Involvement of novel regulatory DNA elements and co-factors

Transcription factor motif analysis in Smad2/2X/3/4 binding sites revealed significant enrichment of Smads, FoxH1 and HEB(Tcf12)-a bHLH family protein

been implicated in endoderm specification from human ES cells (Figure 4-21) (Yoon et al., 2011). Interestingly, the Smad4 binding motif is closely associated with a T-box binding motif (Figure 4-21, 4-22). The most highly expressed T-box family member Eomes is a known player in gastrulation (Arnold et al., 2008; Russ et al., 2000; Ryan et al., 1996; Nowotschin et al., 2013). It directly interacts with Smad2/2X/3/4 and drives transcription of mesendoderm differentiation genes (Slagle et al., 2011; Teo et al., 2011). Further studies with transcriptional profiling and genome-wide occupancy analysis are required to determine the precise role of *Eomes* in mediating Nodal signaling and mesendoderm differentiation.

Unlike the promoters and enhancers of *Eomes* and *Mixl1* that are centered by classical Smad binding sequences CAGAC/T, the Smad2/2X/3/4 bound region of *Gsc* promoter is centered by two consecutive GC-rich boxes between -390bp to -440bp from TSS despite the presence of multiple CAGA sequence cassettes adjacent to FoxH1 binding motif (Figure 4-23). DNA sequences at both the GC-rich region and FoxH1 binding sites are conserved between mouse and human (Figure 4-23). These two GC-rich boxes, together with another upstream sequences were previously shown as protected by Smad4 MH1 domain independent of FoxH1 in *in vitro* DNase footprint assays (Labbe et al., 1998). To test whether these GC-rich boxes are functionally important, I utilized CRISPR/Cas9 to introduce mutations in these sites. CRISPR mutagenesis resulted in one clone that contains a homozygous

deletion from GC1 to FoxH1 binding motif, and another clone that contains a homozygous deletion from part of GC2, GC1 to FoxH1 binding motif (Figure 4-23). EB differentiation assay showed that both clones had compromised *Gsc* induction while control gene *Smad7* was unaffected. This result suggests that the GC2-GC1-FoxH1 binding motif cassette is required for *Gsc* induction during differentiation and in particular, DNA sequences in GC1 region are required for *Gsc* transcriptional activation.

4.4.3 Exon3 and protein stability

Previous studies and my current analyses suggest that exon3 plays a regulatory role in Smad2X function. The function of exon3 is likely related to its close proximity to a few post-translational modification sites and the DNA-contacting β-hairpin. In the absence of TGF-β stimulation, Smad3 is phosphorylated by GSK3β at Thr66, which leads to Smad3 ubiquitination and degradation (Guo et al., 2008a). Smad3 phosphorylation by GSK3β requires a priming phosphorylation by PKC at 4 residues downstream, Ser70. In Smad2X, the 30 amino acid insert from exon3 separates the corresponding residues Thr76 and Ser110. Therefore Smad2X could not be phosphorylated by GSK3β (Guo et al., 2008a), and Smad2X is longer-lived than Smad2/3 (Guo et al., 2008a). The selective interaction between GSK3β and Smad3 but not Smad2X has been recapitulated in other contexts such as cardiomyocyte hypertrophy (Hua, et al., 2010). In addition, Smad3 but not Smad2X interacts with CKIγ2, a regulator of

protein ubiquitination and degradation (Guo et al., 2008b). The distinct protein degradation rates between Smad2X and Smad2/3 is reproducible in ES cells and EBs (data not shown). Nevertheless, how these longevity differences affect functions of Smad2X versus Smad2/3 remains to be explored.

4.4.4 Exon3 and the conformation of Smad2X

Without TGF-β stimulation, exon3 causes Smad2X and Smad2/3 to be present in distinct oligomeric states: Smad2X is a monomer without associating with protein factors that could alter the gel filtration behavior while Smad2/3 exists in a number of high-molecular weight complexes and Smad4 is detected as homo-oligomers (Jayaraman and Massague, 2000). It is unclear yet whether the monomeric Smad2X represents the pool of "poised" or "free" Smad2X ready for receptor interaction and phosphorylation. Upon receptor-mediated phosphorylation, both Smad2X and Smad2/3 are incorporated into high-molecular weight complexes (Jayaraman and Massague, 2000). Since the MH2 domains mediate most homophilic and heterotypic interactions, the difference in the oligomeric states is likely due to conformational changes occurred to Smad2X-Smad2/3 MH2 domains.

Interestingly, GSK3β/Axin and CKIγ2 interact with Smad3 at its MH2 domain, yet these interactions could be blocked by the presence of exon3 in the MH1 domain (Guo et al., 2008b). This data suggests that exon3 in the Smad2X

MH1 domain could directly impact the function of MH2 domain. Indeed, in the un-stimulated state, the MH1 domain of Smad2X interacts with the MH2 domain intramolecularly (Hata et al., 1997). This interaction represses the activities of Smad2X MH2 domain, presumably through excluding it from contacting other protein partners (Hata et al., 1997). Receptor mediated phosphorylation releases the MH1-MH2 interaction (Hata et al., 1997). Smad2X R133C mutation, a tumor-associated mutation, enhances the MH1-MH2 interaction, increases Smad2X degradation rate, and diminishes Smad2X mediated TGF-β tumor suppressive responses (Hata et al., 1997). Intramolecular interaction between Smad3 MH1-MH2 has not been reported. Further experimentation is required to determine whether exon3 is involved in regulating the functionally repressive MH1-MH2 interactions.

4.4.5   DNA-binding properties of R-Smads

The TGF-β family growth factors are involved in regulating a plethora of biological processes, and the R-Smad effectors exhibit highly divergent genome-wide occupancies across multiple cell types (Mullen et al., 2011; Qin et al., 2016). While Smad2/2X/3 share functional redundancies and similar expression patterns, Smad4 is specifically required for a large proportion of TGF-β response genes. Previous reports suggested that Smad2/3 bind to identical DNA sequences as Smad4 (Dennler et al., 1999; Dennler et al., 1998; Zawel et al., 1998), yet Smad2/3 could not replace Smad4 for target

gene activation. The precise role of Smad4 in mediating transcriptional activation remains elusive. Labbe et al reported that excess Smad3 inhibits the transcriptional activity of Smad2X/Smad4/FoxH1 complex through competitively occupying Smad4 binding sites in an ARE reporter (Labbe et al., 1998). This observation supports the utilization of Smad2X, a non-DNA binding R-Smad, as a nucleating factor for assembling productive transcriptional complexes with Smad4 and other lineage-specific transcription factors. Further investigation on whether the competitive inhibition of Smad4 by Smad2/3 occurs in ES cells and mesendoderm differentiation will provide important insights on the functional coordination between the Smad effectors. Plausibly, the functional diversification of Smads is a prerequisite for the highly versatile and cell-type specific functions of TGF-β signaling.

The conformation of Smads varies from steady states to ligand-stimulated states. It is possible that in the absence of TGF-β stimulation, the 30 amino acid insert encoded by exon3 enhances the direct interaction between Smad2X MH1 and MH2 domains, which forms a "poised" conformation that restricts Smad2X from other molecular interactions, and protects it from GSK3β/Axin mediated phosphorylation and degradation. Whereas in the receptor-phosphorylated and activated Smad2X, MH1-MH2 interaction is released and exon3 suppresses the DNA binding ability of MH1 domain, which makes Smad2X a favorable molecular partner for Smad4 and DNA-binding co-factors. Further investigation is required to delineate the physiological role of each conformational states.

Smad2/2X/3/4 ChIP-Seq enriched motif

Motif	Factor	logP-value
	HEB	-2.30e3
	Foxh1	-1.72e3
	Tbox:Smad	-1.69e3
	OSN	-1.56e3
	Sox3	-1.17e3
	Oct4	-7.87e3
	Tcf3/4	-7.18e3
	Foxa2	-6.59e2

*OSN: Oct4-Sox2-Nanog

Figure 4-21. The most significantly enriched known transcription factor binding motifs predicted from the top 200 Smad2/2X/3/4 binding regions under hActivin A treatment and in EBs.

A

T-box: Smad binding motif

T-box/Eomes binding   Smad binding
        sequence   sequence

B

Center of *Gsc* +6kb enhancer
GGGATAATGAACCGATGTGTCAGACGCTCAGCTGGGTCCCTTCCTCTCTC

Center of Mixl1 -0.5kb promoter
GGGATAATGAACCGATGTGTCAGACGCTCAGCTGGGTCCCTTCCTCTCTC

Center of *Eomes* -10kb enhancer
TGGGGGGTGCTGTCTAGCAGATAGCTCCTCCTCTACCAGCGTTGGGA

Center of *Eomes* +9kb enhancer
TTCCTAAGATGGAGTGTCAGACACCAAATTGTGAAGAGAT

Center of Foxa2 -35kb enhancer
ATACAGCTGTTAATGGTGCCAGACTGGAAGGATTTCCTTT

Figure 4-22. (**A**) Scheme of T-box: Smad binding motif identified in Smad2/2X/3/4 ChIP-Seq binding sites. (**B**) DNA sequences in mesendoderm gene enhancers and promoters, highlighting the presence of T-box: Smad binding motif. Sequence in red: T-box binding site; sequence in blue: Smad binding site.

Figure 4-23. Requirement of GC-rich sequences in *Gsc* promoter. (**A**) Conserved motifs in *Gsc* promoter and enhancer regions. The conserved DNA motifs in the proximal promoter and major enhancer site of *Gsc* are highlighted. The distance between each motif is annotated below the axes. Two previously unannotated GC-rich motifs are identified in the proximal promoter region. (**B**) CRISPR/Cas9-mediated targeted mutagenesis and genotyping results of *Gsc* promoter knockout clones. Dashed blue line indicates the CRISPR deleted region in the mutant cells. (**C**) qRT-PCR determined relative *Gsc* expression in proximal promoter knockout cells. *Smad7* is used as a non-specific control.*, p<0.05.

Chapter 5 Interrogating the Transcriptional Programs of Metastatic Stem Cells by TRAP-Seq

5.1 Summary

In complex tissues such as developing embryos and tumors, cell-cell interactions play vital roles in determining the tissue structure and function. However, there are very limited approaches that allow gene expression characterization for cells of interest without dissociating each stromal component. Here I describe the adaptation of a new approach for analyzing the gene expression profile of specific cell populations *in situ* within heterogeneous tissues: Translating Ribosome Affinity Purification and Sequencing (TRAP-Seq). I characterized and optimized the TRAP-Seq protocol for profiling the molecular events in multi-organ metastasis. In collaboration with colleagues from the Massagué lab, I applied this technology in multiple contexts including breast cancer microenvironment, melanoma drug resistance, and brain metastasis development. These studies have assisted our understanding about cancer development, metastasis and therapy responses.

5.2 Introduction

Metastasis accounts for the majority of cancer associated mortality. The establishment of metastasis, mediated by disseminated tumor cells (DTCs), is a complex process that involves extensive interactions between cancer cells and the stroma in the primary tumor or the metastasis host organs. Albeit majority of DTCs die while being challenged by microenvironmental changes, small proportions of cells survive and develop overt metastasis (Oskarsson et al., 2014). While cancer cells may arrive at any random site via the circulation, each host organ differs in the composition of stromal cells, extracellular matrix components, cytokines and growth factors (Oskarsson et al., 2014). DTCs that thrived in the same organ often reside in "metastatic niches" that share key pro-survival properties. How the diverse "metastatic niches" affects tumor cell survival, proliferation and migration is an important subject of study.

Transcriptome profiling is a highly efficient and informative approach to assist our understanding about the molecular events within the metastatic niches. Recently, explosive advancements in next generation sequencing technologies enabled rapid and high-throughput transcriptome analysis at single cell resolution. To apply sequencing technologies to heterogeneous tissues such as embryos, organs, primary tumors and metastases, tools that permit transcriptome isolation from cells *in situ* is a prerequisite.

Translating Ribosome Affinity Purification (TRAP) is a method developed to identify specific molecular markers in neural subtypes (Doyle et al., 2008; Heiman et al., 2008). In the TRAP protocol, the cell type of interest is engineered with an EGFP-RPL10a (Ribosome Protein, Large Subunit 10A) fusion protein construct. Expressed EGFP-RPL10a fusion protein will be incorporated into polysome-mRNA translation complexes. After whole tissue homogenization, the EGFP epitope tag allows enrichment of cell type specific polysomes through immunoprecipitation (Doyle et al., 2008; Heiman et al., 2008). The optimized anti-EGFP antibodies are among the best-characterized monoclonal antibodies and offer highly specific immunoprecipitation. Following immunoprecipitation, polysome associated mRNA can be extracted *in vitro* and analyzed for gene expression.

When combined with next-generation sequencing technologies, TRAP-Sequencing (TRAP-Seq) can be utilized for high-throughput gene expression profiling of target cell types *in situ* (Figure 5-1). Compared with conventional approaches, TRAP-Seq exhibits two main advantages: (1) It bypasses the cell isolation process commonly achieved by flow cytometry or tissue micro-dissection. Flow cytometry analyses often require hours of sample processing, special handling of tissues, and post-sorting cell recovery and expansion. These procedures and manipulations are likely to introduce unpredictable perturbations to the transcriptome or translatome. The laborious process and high-demand for special reagents limits the quantity of samples that

could be processed simultaneously. In contrast, TRAP-Seq directly dissociates the tissue samples and fixes the translatome through cyclohexamide with minimized gene expression disturbances. By skipping the flow cytometry process, TRAP-Seq does not rely on pre-defined cell surface markers and flow cytometry compatible antibodies. Moreover, the ease of sample handling allows large number of samples to be processed in parallel. (2) TRAP-Seq enriches polysome-associated mRNAs that are actively translating into protein products. The loading of specific mRNAs to ribosome machinery is a tightly regulated process for translational control, and recent reports suggested that the amount of translating mRNA correlates with the protein products stronger than the total mRNA population in the cell (Ingolia, 2014; Ingolia et al., 2009; Ingolia et al., 2011).
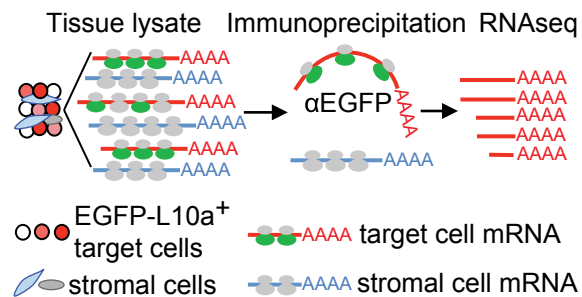
Figure 5-1. Scheme of the Translational Ribosome Affinity Purification Sequencing (TRAP-Seq) protocol and quality assessment of TRAP RNA. Cell type of interest (e.g. metastatic stem cells) is initially engineered with the expression of an EGFP-RPL10a (Ribosome protein large subunit 10A, L10a) fusion protein and implanted onto the host mouse models. Tumor tissue containing EGFP-RPL10a+ cells are then quickly harvested and processed through an immunoprecipitation protocol with anti-EGFP monoclonal antibodies. The highly specific anti-EGFP antibodies ensure only polysomes from the labeled cell population will be enriched. mRNA can be recovered from beads and prepared for RNA-Seq. Reprinted from Cell, 154(5), Zhang XH, Jin X, Malladi S, Zou Y, Wen YH, Brogi E, Smid M, Foekens JA, Massagué J., Selection of bone metastasis seeds by mesenchymal signals in the primary tumor stroma. 1060-73. Copyright 2013, with permission from Elsevier.

5.3 Characterization and optimization of TRAP

5.3.1   Background

Originally established in the brain tissues to identify cell-type specific markers, TRAP-Seq meets the need for tools that enable gene expression profiling of defined cell populations *in situ*. The application of this technology in tumor biology relies on its sensitivity, specificity, and the quality of mRNA products. (1) Sensitivity is defined by the minimum number of target cells required to immunoprecipitate sufficient mRNA material for downstream qRT-PCR or RNA-Seq analysis. TRAP was tested for isolating transcripts of cell-type restricted molecular markers that are often abundant. It remains to be determined whether TRAP-Seq could be applied to characterize low number of target cells such as drug-resistant cells, early metastasis initiating cells, and latent competent cancer cells. (2) The specificity of *in situ* gene expression characterization is defined by the signal-noise ratio. Most noise will be derived from non-target cell population in the tissue sample. Despite EGFP antibodies are highly efficient at on-target recognition, non-specific binding will inevitably occur when large amounts of host tissues are dissociated. Determining the specificity of TRAP-Seq in various tissues will allow better assessment of noise strengths. (3) Protecting polysome-associated mRNA from free RNase released during tissue dissociation is a major technical challenge in TRAP-Seq. Each tissue type exhibits distinctive spectrum and quantity of RNases. It has not been

tested whether this technique could be used to profile cancer cells within host organs besides the brain.

5.3.2   Results

To explore the sensitivity of TRAP with low number of target cells, I performed TRAP with 1000, 500, 250, 100 or 50 EGFP-L10a+ MDA-MB-231 human breast cancer cells and analyzed the expression levels of a selected list of genes including *B2M*, *CTGF*, *SERPINE1* by qRT-PCR (Figure 5-2). *B2M* is an abundantly expressed house-keeping gene, whereas *CTGF* and *SERPINE1* are target genes of TGF-β and are moderately expressed in this cell line (Xi et al., 2008). While total mRNA from low number of cells could be purified without difficulty, qRT-PCR detectable amount of mRNA could be isolated from as few as 50 EGFP-RPL10a+ cells via the TRAP protocol. Remarkably, the Ct numbers of all three tested genes exhibited a linear correlation with the logarithmic transformed cell number, suggesting that TRAP faithfully captured the abundances of target mRNA species. These results imply that TRAP is suitable for mRNA purification in low number of target cells in vitro.

To test the specificity of the TRAP protocol in vitro, I mixed 200 TGF-β treated EGFP-RPL10a+ MDA-MB-231 cells with $1x10^5$ transformed mouse mammary epithelial cells and processed through the TRAP protocol. qRT-PCR analysis was performed with TGF-β responsive genes *SERPINE1* and *SKIL*

using human specific primers. Using gene expression changes in total mRNA as references, TRAP effectively captured the induction of both *SERPINE1* and *SKIL* transcripts in as few as 200 target cells while mouse RNA remain undetectable by mouse specific primers (Figure 5-2). Thus TRAP is highly specific in distinguishing EGFP-RPL10a expressing cells from background, and it is highly sensitive in capturing gene expression changes.

To validate our qRT-PCR analysis with an unbiased approach, I performed microarray analysis with the TGF-β or mock treated total and TRAP mRNA. Within the 29 genes that were significantly de-regulated in the total mRNA by TGF-β treatment, 27 of which were recapitulated in the TRAP mRNA (Figure 5-2). This analysis reinforced the high specificity and fidelity of TRAP through genome-wide analysis.

Compared with the brain tissue, many other tissues such as the mammary fat pad and lung are more enriched in RNase. To test whether high levels of free RNase could cause RNA degradation during TRAP processing, I injected EGFP-RPL10a+ MDA-MB-231 cells to various organs orthotopically and analyzed the integrity of TRAP mRNA products. Indeed, the mRNA from lung, mammary fat pad, kidney and subcutaneous tissues showed more severe degradation than samples from the brain (data not shown). To overcome RNA degradation during sample processing, I tested a panel of more robust and broad-spectrum RNase inhibitors from commercial sources, and

optimized the tissue processing method. These optimizations led to improved TRAP RNA integrity even when whole mouse organs were homogenized for immunoprecipitation. The integrity of the mRNA from lung, mammary fat pad, brain, kidney and subcutaneous tissues purified via the optimized protocol is sufficient for majority of downstream analyses including next-generation sequencing (Figure 5-3).

5.3.3   Conclusions and Discussions

TRAP is highly sensitive *in vitro*, it could linearly detect as few as 50 EGFP-RPL10a+ cells *in vitro*, and could recapitulate TGF-β responses in as few as 200 EGFP-RPL10a+ cells mixed with 500 fold more EGFP-RPL10a- cells with robust signal-noise ratio. The protocol has been optimized to prevent the degradation of polysome-associated mRNA during immunoprecipitation in various organs including the lung, brain, subcutaneous tissues, kidney and mammary fat pad. Collectively, these results demonstrate that TRAP is a sensitive, specific and adaptable technique and it is suitable for translational profiling of target cell types *in situ*. Combining TRAP with next-generation sequencing technologies and bioinformatics analysis will provide a powerful high-throughput method for interrogating gene expression profiles *in situ* in complex tissues represented by embryos and tumors.
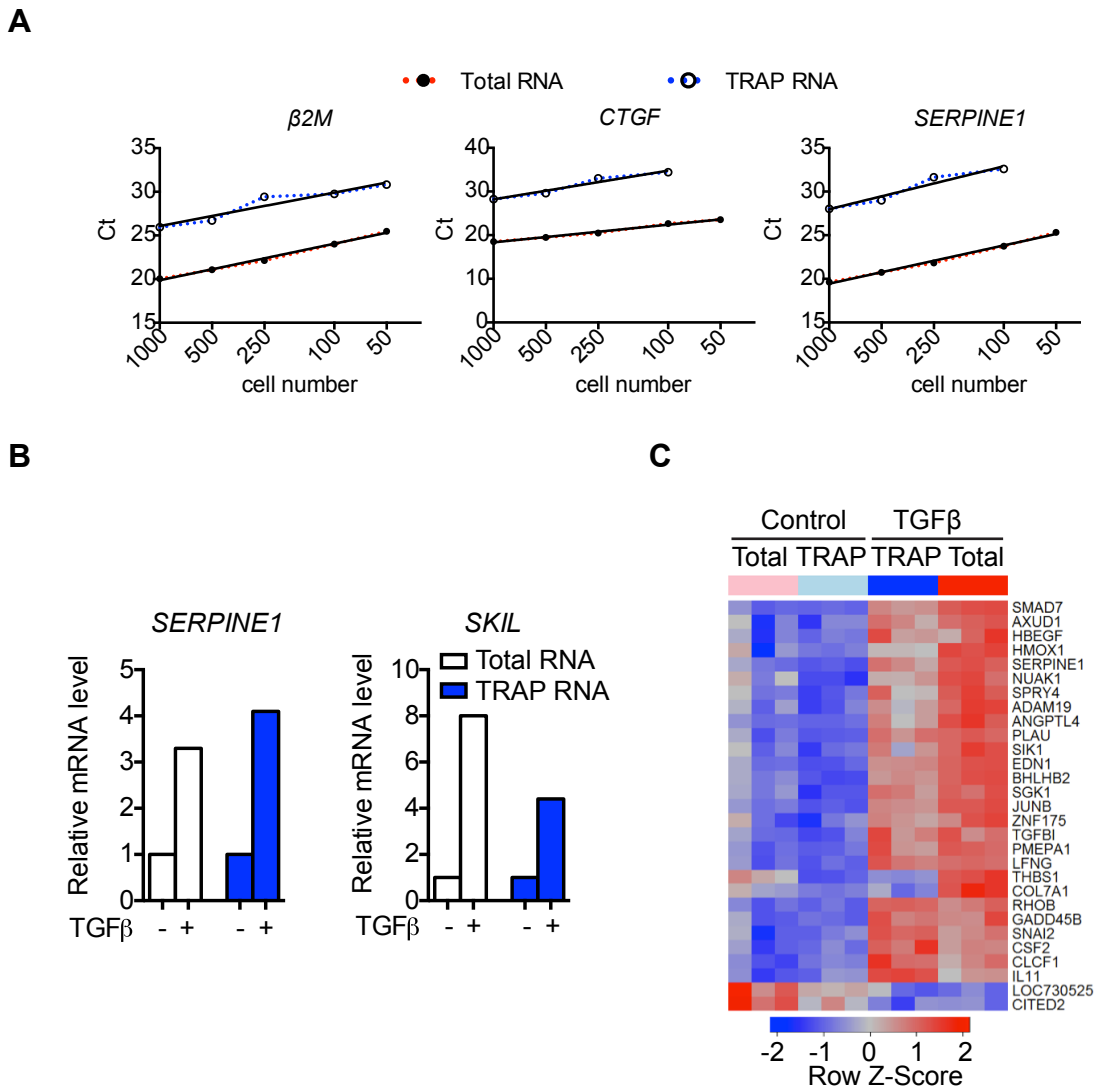
Figure 5-2.   Determination of sensitivity and specificity of TRAP technology. (**A**) Sensitivity of the TRAP technology. The amount of TRAP RNA, represented by β2M, is linear through 50-1000 cells. (**B**) qPCR validation of the TGF-β response genes. TRAP RNA from 200 TGF-β treated EGFP-L10a cells mixed with 10^5 mouse cells shows similar expression induction of TGF-β responsive genes (*SKIL* and *SERPINE1*) with total RNA. Mouse RNA is not detectable in TRAP RNA. (**C**) Heatmap of differentially expressed genes in response to TGF-β treatment. Both total RNA and TRAP RNA were analyzed by microarrays.
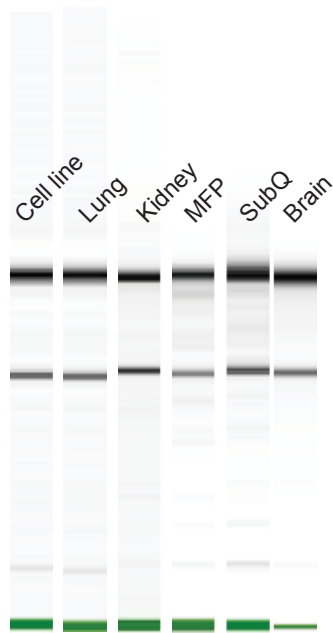
Figure 5-3.   Quality assessment of TRAP RNA acquired from tumor tissues in various organs and conditions. EGFP-RPL10a expressing MDA-MB-231 cells were injected orthotopically or by intracardiac injection to NOD/SCID mice and metastatic tumor tissues from each organ were harvested after 4-6 weeks for TRAP analysis. MFP: mammary fat pad injection; SubQ: subcutaneous injection. TRAP RNA is analyzed by Bioanalyzer 2000. High integrity of 28S and 18S ribosomal RNA indicates that the TRAP RNA is of high integrity.

5.4 Applications of TRAP-Seq in tumor microenvironment analysis

5.4.1 Mesenchymal stromal cells select breast cancer bone metastasis seeds

Background

The capacity to develop bone metastasis in breast cancer especially triple-negative breast cancer is associated with the Src response gene signature in the primary tumor. Src signaling enhances tumor cell responsiveness to PI3K/AKT cascade and promotes survival in the bone marrow microenvironment (Zhang et al., 2009). In HER2+ breast cancers, Src pathway could be directly activated by hyperactive HER2 (ERBB2) signaling. However, the signaling sources that activate the Src pathway in triple-negative breast cancer cells have not been characterized. Through molecular signature analysis, Zhang, Jin et al identified that mesenchymal stromal cells (MSCs) in breast cancer primary tumors promotes Src activation in cancer cells (Zhang et al., 2013). Co-implantation with MSCs did not affect primary tumor growth but elevated bone metastasis incidences through pre-selection of bone metastasis seeds with high Src activity (Zhang et al., 2013). Nonetheless, the mechanisms by which MSCs mediates Src pathway activation in cancer cells is unknown (Zhang et al., 2013).

Results

To investigate how MSCs impact the gene expression profile of triple-negative breast cancer cells in the primary tumor, I co-transplanted EGFP-RPL10a+ MDA-MB-231 cells with MSCs at 1:1 ratio into mouse mammary fat pad. EGFP-RPL10a+ MDA-MB-231 cells transplanted alone were used as control. Tumors were allowed to grow for three weeks prior to harvest. No overt metastasis was detected at this stage. Tumors were then freshly excised and homogenized for anti-EGFP immunoprecipitation (Figure 5-4). RNA-Seq analysis of the TRAP samples showed that MSC-supplemented tumors were significantly more enriched with a gene signature correlated with abundant cancer-associated fibroblasts (a cell type differentiated from mesenchymal cells) in human tumor samples (Zhang et al., 2013). Moreover, the gene expression profile of cancer cells from MSC-supplemented tumors resembled the gene expression signature of MDA-MB-231 cells treated with CXCL12 (Chemokine (C-X-C motif) Ligand 12) and IGF1 (Insulin-like Growth Factor 1) at optimal concentrations. PI3K/AKT signaling downstream of CXCL12/IGF1 was the third most enriched pathways in the MSC+ tumors. However, no detectable CXCL12 or IGF1 transcripts were purified by TRAP from the cancer cells. Collectively these results suggest that CXCL12 and IGF are secreted by the MSCs, and then act on the cancer cells to boost survival and metastasis.

Conclusions and Discussions

In this study, I dissected the molecular mechanisms underlying the pro-metastatic impact of mesenchymal stromal cells in breast cancer. TRAP-Seq analysis revealed that mesenchymal stromal cells in breast cancer primary tumor secret CXCL12 and IGF1, which activates Src pathway in breast cancer cells and promotes bone metastasis traits. Immunofluorescence analysis in the MSC-supplemented tumor sections validated the induction of high levels of IGF1 and CXCL12 in the tumor microenvironment (Zhang et al., 2013). This finding provides a molecular link between the mesenchymal stromal cells in the primary tumor with increased bone metastasis incidences.

A

B

Predicted regulators			
Upstream regulators	P-value	Predicted state	Z-score
IGF1R	1.06E-05	Activated	2.585
CXCL12	3.59E-05	Activated	2.255
AKT1	2.16E-03	Activated	2.951
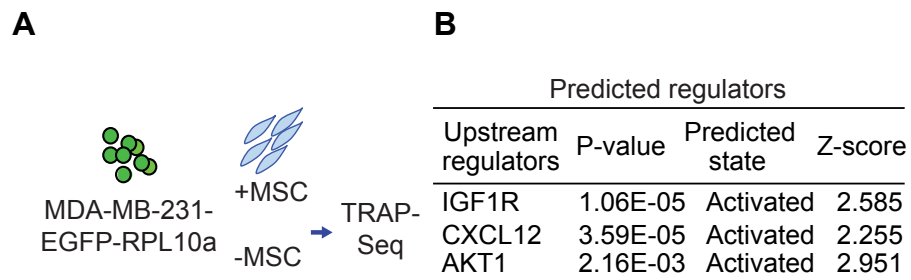
Figure 5-4. (**A**) Applying TRAP for breast cancer-mesenchymal stromal cell (MSC) interaction analysis. MDA-MB-231-EGFP-RPL10a cells were transplanted to the mammary fat pad of athymic mice with or without MSCs. (**B**) TRAP-Seq analysis predicted that MSC induces an IGF1/CXCL12 signaling cascade in the tumor cells. Reprinted from Cell, 154 (5), Zhang XH, Jin X, Malladi S, Zou Y, Wen YH, Brogi E, Smid M, Foekens JA, Massagué J., Selection of bone metastasis seeds by mesenchymal signals in the primary tumor stroma. 1060-73. Copyright 2013, with permission from Elsevier.

5.4.2 Therapy-induced tumor secretome promotes resistance to targeted therapy

Background

Majority of therapeutic agents against cancer induces de-bulking of the tumor mass. Often inevitably, drug-resistance develops along the course of drug treatment. How the clones of drug-resistant cells react to the dramatic changes in the regressing tumor microenvironment has been poorly delineated. Recent studies of my colleague Anna Obenauf, PhD, uncovered that therapy-induced secretome promotes the outgrowth, dissemination, and metastasis of drug-resistant cancer cell clones and supports the survival of drug-sensitive cancer cells (Obenauf et al., 2015). The pro-survival effects of therapy-induced secretome diminishes the efficacy of targeted therapeutic agents, and contribute to incomplete tumor regression and eventual tumor recurrence (Obenauf et al., 2015). Importantly, the molecular mechanisms underlying the actions of therapy-induced tumor secretome have not been characterized. A deeper understanding of this process could advise better usage of therapeutic agents with the goal of overcoming drug resistance.

Vemurafenib is a mutant BRAF^{V600E} inhibitor approved for the treatment of patients with late-stage metastatic melanoma. Though Vemurafenib induces marked tumor regression and improved survival, drug resistance often develops in later course of treatment with unknown mechanisms. In BRAF^{V600E} driven

human melanoma cancer cell lines such as A375, Vemurafenib resistant cells (A375^R) could be isolated through chronic drug exposure (Obenauf et al., 2015). When co-implanted with the sensitive cells (A375/A375^R, 99.95/0.05%) and treated by Vemurafenib, A375^R cells grow significantly faster than the vehicle treated group. This study suggests that the de-bulking of the sensitive tumor benefits the growth and survival of rare drug resistant population (Obenauf et al., 2015), a process mediated by a dynamic and reactive therapy-induced secretome in the regressing tumor (Obenauf et al., 2015).

Results

To determine the effect of the reactive secretome on the drug-resistant tumour subpopulation in a regressing tumour, I engineered human A375 Vemurafenib resistant cells (Obenauf et al., 2015) with EGFP-RPL10a transgene expression and co-implanted with large number of Vemurafenib-sensitive A375 cells into mice subcutaneously (Figure 5-5). Vemurafenib was administered for 5 consecutive days to induce tumor regression, followed by TRAP-Seq analysis on the drug-resistant cell population. Cytokine and growth factor array analysis suggest that the regressing tumor microenvironment contains many cytokines and growth factors that will stimulate signaling responses in the resident, rare, drug-resistant cells (Obenauf et al., 2015). Compared with drug-resistant cells that reside in a tumor-progressing microenvironment, the gene expression pattern of resistant cells in the regressing microenvironment was enriched for biological processes involved in cell viability, proliferation and cell movement

214

(Figure 5-5). These molecular features are in line with the in vivo phenotype of accelerated growth of drug-resistant cells in response to the regressing tumor microenvironment.

Pathway analysis of the regressing tumor microenvironment-specific genes revealed the activation of a number of pathways including the PI(3)K/AKT, BMP-SMAD and NFκB pathways (Figure 5-5). The hyperactivity of the PI(3)K/AKT cascade implied potential vulnerabilities of the cells to PI(3)K/mTOR inhibitors (Figure 5-5). Remarkably, the combined inhibition of MAPK and PI(3K)/AKT/mTOR pathways significantly blunted the outgrowth of Vemurafenib-resistant cells in the A375/A375^R tumors (Obenauf et al., 2015). The growth inhibition was specific for the enhanced proliferation in the regressing tumor and had no impact on the expansion of Vemurafenib-resistant tumors alone. These results support that the secretome-induced proliferation of drug-resistant cells could be exploited for therapeutic targeting.

Conclusions and Discussions

Development of targeted therapies that could inhibit cancer cell growth and de-bulk the tumor has been a common research field for many cancer types. However, resistance to targeted therapy often inevitably develops, limiting the effect of therapeutic agents. Targeted abrogation of oncogenic signaling significantly alters the secretome of drug-sensitive tumor cells, paradoxically establishing a regressing environment that favors the proliferation and

215

metastasis of therapy-resistant cells. In this study, *in situ* translatome analysis of drug-resistant melanoma cells responding to targeted therapy-induced tumor regression revealed hyper-activation of several pro-survival pathways including the PI3K/AKT pathway. These results provide a rationale for combining Vemurafenib and inhibitors of PI3K/AKT/mTOR pathway in the treatment of Vemurafenib sensitive melanoma patients (Obenauf et al., 2015). The concept of therapy-induced tumor secretome promoting resistance and tumor progression could be extrapolated to the clinical handling of other cancer types.

A

+Vemurafenib
(5 days)

Sensitive cells

EGFP-RPL10a
drug-resistant cells

Stromal cells
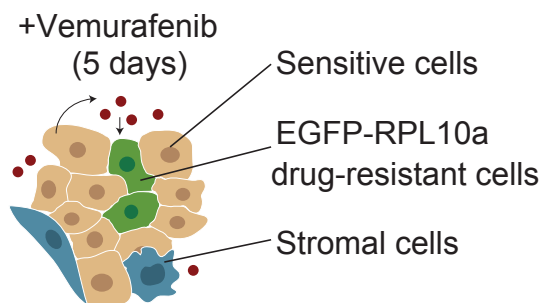
B

Regulator	Z score	P value
NFκB (complex)	2.745	1.20×10^{-3}
IGF1	2.736	1.46×10^{-5}
BMP2	2.629	8.98×10^{-4}
MTPN	2.621	7.92×10^{-6}
ESR1	2.581	5.86×10^{-4}

Target	Inhibitor	Z-Score	P-value
PI3K	LY294002	-2.287	2.87E-03
PKC	bisindolylmaleimide I	-2.200	1.26E-03
mTOR	sirolimus	-2.750	1.98E-02

Figure 5-5.   Applying TRAP-Seq to analyze transcriptome of rare resistant population in therapy-induced regressing tumor. (**A**) Vemurafenib resistant A375 (melanoma) cells were labeled and implanted subcutaneously to NOD/SCID mice. TRAP-Seq analysis for resistant cells in the presence or absence of Vemurafenib treatment was performed.   (**B**) Gene signature analysis predicted that IGF1- PI3K signaling is upregulated in resistant cells upon treatment. Obenauf AC, Zou Y, Ji AL, Vanharanta S, Shu W, Shi H, Kong X, Bosenberg MC, Wiesner T, Rosen N, Lo RS, Massagué J. (2015) Therapy-induced tumour secretomes promote resistance and tumour progression. Nature, 520, 368-72.

5.4.3   L1CAM mediates vascular co-option and brain metastasis through Hippo/Yap pathway

Background

Brain metastasis is the leading cause of lethality in many late stage cancer patients. Recent studies in the Massague group have identified the ability to co-opt with the vascular as a key property of brain metastatic cancer cells (Valiente et al., 2014). L1CAM, a cell adhesion molecule normally expressed by neurons and mediates axon guidance, plays a critical role in mediating cancer cell vascular co-option (Valiente et al., 2014; Schafer and Altevogt, 2010). Expression of L1CAM in cancers is associated with poor prognosis (Doberstein et al., 2011; Schroder et al., 2009). L1CAM expressing cancer cells efficiently attach to vascular endothelial cells and to each other (Valiente et al., 2014), whereas L1CAM-depletion in brain metastatic cells blocks cancer cell-vascular association, inhibits cancer cell proliferation, and reduces overall brain metastasis rate (Valiente et al., 2014).

Monoclonal antibodies targeting the extracellular domains of L1CAM hold great promise for diminishing brain metastasis development. However, the downstream signaling cascade that mediates L1CAM function remains poorly characterized, adding mystery to the clinical application of anti-L1CAM therapies. In collaboration with Manual Valiente, PhD, I utilized TRAP to investigate the L1CAM-dependent gene expression responses in *ex vivo* organotypic brain

slice cultures. This analysis unveiled an integral participation of the Hippo/Yap pathway in mediating L1CAM's pro-survival and pro-metastatic functions.

Results

To determine the L1CAM dependent gene expression responses, I transduced H2030 (human metastatic non-Small Cell Lung Cancer cell line)-BrM2-Ctrl^sh or *L1CAM*^sh cells with EGFP-RPL10a expression. Both cell lines were inoculated to organotypic brain slices and cultured for 48h to allow these cells co-opting with brain vasculature. Whole brain slices with cancer cells were harvested and processed for TRAP analysis. Ctrl^sh or L1CAM^sh cells maintained in culture dishes were used as a non-vascular co-opting control.

RNA-Seq analysis identified 110 genes that are differentially expressed between H2030-BrM2-Ctrl^sh and *L1CAM*^sh cells in vitro and 135 genes differentially expressed *ex vivo* (Figure 5-6). Efficient knockdown of L1CAM mRNA is confirmed by RNA-Seq. To explore whether the L1CAM dependent genes enriches for certain biological pathways, I performed signaling pathway response signature analysis. Analysis results indicated the Hippo/Yap and NFκB pathways as the most significantly inhibited pathways in response to L1CAM depletion (Figure 5-6, 5-7). Further investigation is required to test the efficacy of anti-Hippo/Yap and anti-NFκB approaches in reducing brain metastasis.

Conclusions and Discussions

Previous research in our lab has identified L1CAM as a mediator of vascular co-option during cancer cell brain colonization, yet the downstream effectors of L1CAM signaling in cancer cells have not been defined. In this study, I utilized TRAP-Seq to interrogate the L1CAM dependent translatome changes. This analysis led to the finding that the Hippo/Yap pathway and NFκB pathway might be mediating the intracellular function of L1CAM. Preliminary results from my colleague Ekram Emrah Er, PhD, showed that shRNA mediated Yap knockdown in brain metastatic cancer cells significantly reduced their capacity in forming large brain lesions in xenograft models, supporting the eligibility of pursuing anti-Hippo/Yap therapeutics. In sum, these results suggest novel mechanistical insights for the L1CAM signaling and vascular co-option, and offers alternative therapeutic opportunities for blocking brain metastasis development.

Collectively, TRAP-Seq is a powerful tool that allows the efficient and real-time profiling of translatome in target cells *in situ*. It assists the revealing of biological processes that are highly dynamic, microenvironment-dependent, and sensitive to tissue dissociation. My characterization and applications suggest that this tool is broadly applicable to the dissection of transcriptional programs of heterogenous tissues from embryos, organs and tumors, etc.
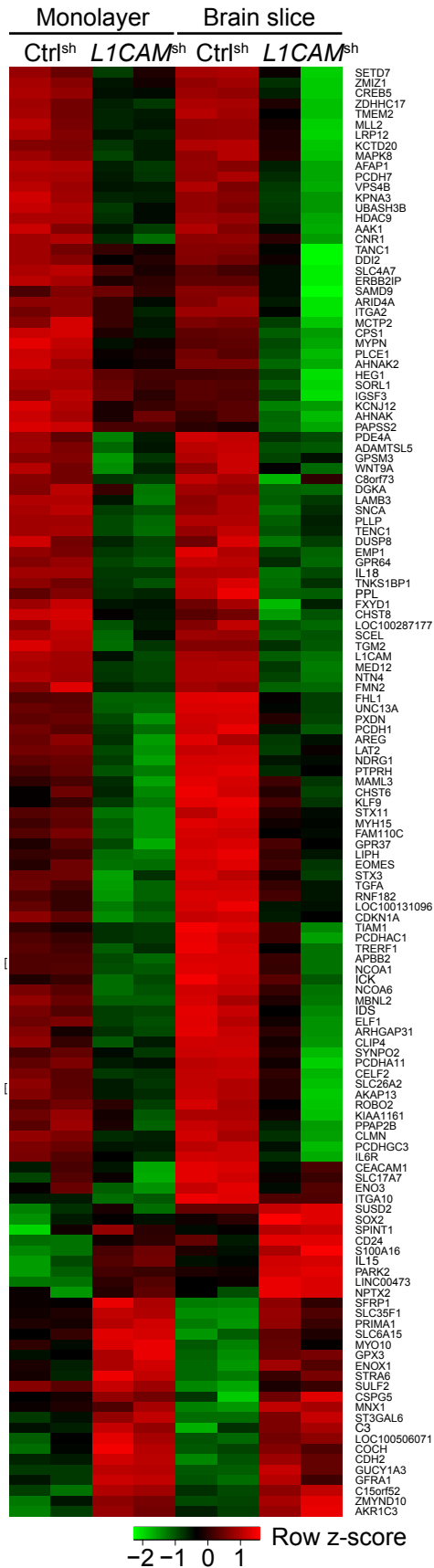
Figure 5-6. Heatmap representing the TRAP-Seq analysis of L1CAM-dependent genes in H2030-BrM2 cells cultured in 2-D monolayer condition or organotypic brain slices.

A

Brain slice TRAP

Ctrl^sh *L1CAM*^sh

	P-value
Hippo	0.02
NFkB	0.02
TGFb	0.13
SRC	ns
Akt	ns
Myc	ns
Ras	ns
TNC	ns

-1 0 1
Row z-score

B

NFkB pathway target genes

Hippo/Yap pathway target genes

CCL20
ITGA2
CXCL3
IL6
IL8
CXCL2
…

NES=1.43
FDR=0.036

C3   AREG
IL32   TGM2
TRAF1 CLMN
ITGAM TGFA
…   NT5DC3
       …
       UPP1
       HBEGF

NES=1.35
FDR=0.108

CDH2
SCD5
IL32
PLAU
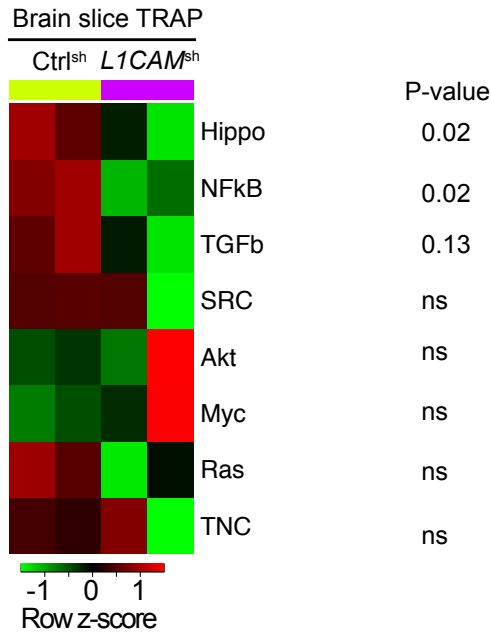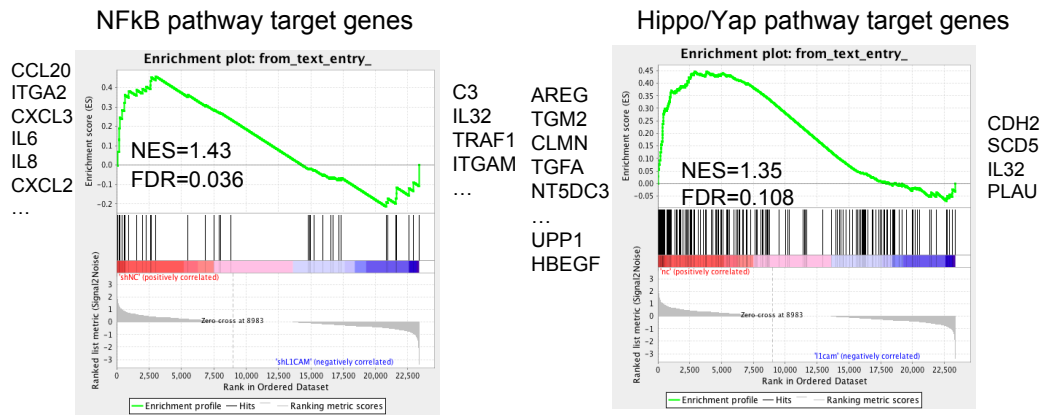
Figure 5-7. (**A**) Heatmap of enriched pathways in L1CAM dependent genes acquired by TRAP-Seq analysis. (**B**) Gene Set Enrichment Analysis highlighting the significant enrichment of NFkB and Hippo pathway in L1CAM dependent gene expression responses.

222

Chapter 6   Conclusions and Discussion

This dissertation describes three major aspects of progresses in the biology of embryonic and metastatic stem cells: first, it identifies that the p53 family members are essential regulators of mesendoderm specification in pluripotent stem cells and early mouse embryos (Chapter 3); second, it documents the regulatory functions of Nodal co-activators in mesendoderm differentiation, and underscores the essential roles of a major *Smad2* splicing isoform, Smad2X, in mediating cell type-specific functions of the TGFβ/Nodal/ Activin pathway (Chapter 4); third, it describes the novel mechanistical insights provided by transcriptional interrogation of metastatic stem cells *in situ* by TRAP-Seq (Chapter 5).

6.1 The p53 family redundantly regulates mesendoderm specification

Despite the numerous evidence on the involvement of p53 in stem cell differentiation, the conceptual importance of p53 in development has been challenged by the normal development of p53 null mice (Donehower et al., 1992; Jacks et al., 1994; Purdie et al., 1994; Tsukada et al., 1993). What is causing these discrepancies has been a long-term debate. Here I show that the p53 family is directly involved in differentiation regulation through Wnt3 expression and Wnt pathway activation. Knockout p53, p63, p73 together abrogates mesendoderm differentiation and embryo development in mice, whereas lack of single or double of the three p53 family members did not cause severe early embryonic defects potentially due to the functional compensation from the remaining family members. Thus, independently of their established role as tumor suppressors that guard genome integrity in mature cells, p53/p63/p73 serve a primordial role in ES cell differentiation and embryo development by driving a cooperation of Wnt and Nodal/Activin/TGF-β transcriptional inputs on mesendoderm identity gene enhancers. The activity of the p53 family members is controlled by Aurka in this context.

Wnt3 and its receptor *Fzd1* are among a small set of p53 target genes that are specifically activated in this context. Induction of Wnt signaling by p53 is critical for activation of mesendoderm differentiation genes. Globally, I show

that Wnt3-activated Tcf3 and Nodal-activated Smad2/3 transcription factors depend on each other for co-occupancy of target enhancer elements in master differentiation loci. My results reveal a selective interdependence between signal-activated Tcf and Smad transcription factors. Thus, the p53 family governs a regulatory network that integrates essential Wnt-Tcf and Nodal-Smads inputs for mesendoderm differentiation in the early embryo.

This study resolves the surprising and puzzling difference between the developmental phenotypes of p53-deficient *Xenopus* and mouse embryos by emphasizing the importance of functional compensation from other family members. Redundancy in fundamental developmental functions has conceivable advantages as it improves the robustness of normal developmental processes. Moreover it provides a mechanistic basis for understanding the role of p53 as a roadblock to induced pluripotency, which should shed light on the efficient induction and secure manipulations of pluripotent stem cells for regenerative medicine.

There are a number of remaining questions that might worth further investigation: (1) Besides the chimeric embryo approaches utilized in our studies, will p53/p63/p73 triple germline knockout in mouse cause embryonic lethality? To study this effect, the triple heterozygous mice should be crossed and a sufficient number (expecting 1/64 siblings are triple homozygous knockout if born at Mendellian ratio) of embryos should be analyzed for the survival rate and

developmental feature of each genotype. (2) How are the activities and stabilities of the p53 family members regulated in developing embryos? As revealed in my study, the protein expression of p63 and p73 is highly context-dependent. Itch, a Nedd4-like ubiquitin ligase, is required for p63 and p73 protein ubiquitination and degradation. YAP1, a transcriptional co-activator downstream of Hippo pathway, binds the PY motif of p73 and stabilizes the protein (Levy et al., 2007). What other p53 family regulators are involved? (3) What are the roles of other p53 target genes besides Wnt3 and Fzd1 in differentiation and development? (4) Are the differentiation-promoting functions of the p53 family members separable from their roles in maintaining genome-integrity? What are the mechanisms that coordinate both functions?

6.2 Transcriptional regulation of Nodal-driven mesendoderm specification through cofactors and Smad2X

In this study I determined the regulatory components of Nodal signaling, and their requirement for transcriptional activation of specific Nodal target genes. I demonstrated the importance of Smad2X in mediating Nodal-driven mesendoderm differentiation. Smad2X is the major protein product expressed from the *Smad2* gene. It is exclusively required for the transcriptional activation of master differentiation regulators downstream of Nodal signaling. In response to Nodal, Smad2X nucleates a transcriptional complex composed of effectors downstream of multiple signaling inputs. This unique function of Smad2X is potentially facilitated by its lack of DNA binding capacity, and by other functions of the highly-conserved exon3 region. Combining the findings described in Chapter 3 and 4, I illustrated a transcriptional regulatory network composed of the p53 family, the Wnt-Tcf3/β-Catenin pathway and the Nodal-Smad2X signaling cascade that drives mesendoderm specification.

My investigation implies that despite the functional redundancy, each R-Smad protein exerts distinct functions in mediating TGF-β family responses due to their differences in the structural and biochemical properties. To further demonstrate the unique role of Smad2X in mediating cell-type specific Nodal responses, isoform-specific ChIP-Seq analysis against Smad2X and Smad2/3 is required. This analysis should reveal the similarities and differences in Smad2X

and Smad2/3 binding sites in ES cells and EBs.

Smad2X and Smad2/3 are co-expressed by almost all major tissue types. Though they are structurally diverged by the presence or absence of the exon3 insert, they both mediate TGF-β signaling (Labbe et al., 1998; Yagi et al., 1999). Whether Smad2X and Smad2/3 are functionally interchangeable has been a long-term debate. In this study, I provide a potential explanation that the DNA binding ability of Smad2 is suppressed by exon3. Moreover, Smad2 primarily mediates tissue-specific gene responses of the TGF-β pathway. The relatively lower affinity towards a major target DNA sequence provides a selective advantage for Smad2 to form transcriptionally productive complexes together with lineage-specific transcription factors. Retaining its DNA binding specificity, Smad3 preferentially binds feedback regulatory target genes of TGF-β signaling without requiring assistance from co-factors. These characterizations potentially provide a structural basis for the pleiotropic functions of the TGF-β pathway in various cell types.

6.3 Characterization of metastatic stem cells through TRAP-Seq

Due to limited tools that permit gene expression profiling of DTCs within their microenvironment, our understanding about the impact of specific stromal components on metastasis development is still rudimentary. Here I adapted the TRAP-Seq technology for profiling the translatome of metastatic stem cells *in situ*. The quality and quantity of purified RNA, sensitivity and specificity of the technology is explored. I also extended the applicability of TRAP-Seq from the brain to more tissue types including the lung, kidney, mammary fat pad and subcutaneous tissues. These protocols allow broader application of this tool for multi-organ analyses. Moreover, I applied this technology to the interrogation of metastatic stem cells *in vitro*, *ex vivo* and *in situ*. These studies shed light on the survival and metastasis mechanisms of breast cancer cells, the drug-resistance mechanisms of melanoma, and the signaling cascade required for brain metastasis development, and unveiled substantial biology that could be exploited therapeutically for tumor inhibition and metastasis blockade.

TRAP-Seq can be applied to various contexts such as developmental processes, tumorigenesis, and therapy responses. It provides a unique opportunity to investigate the impact of microenvironmental factors on the behavior of target cell types. Combining the *in situ* gene expression data with systematic profiling of the genetic and epigenetic landscape of target cell types

in situ will provide pivotal insights about the physiology and the molecular mechanisms of biological processes of interest. Due to the intrinsic limitations of immunoprecipitation assays, I expect there are limitations for applying TRAP-Seq to sparse cell populations within bulky tissues. Further characterization is required to determine the sensitivity of TRAP-Seq in various *in vivo* settings. Recently, the Friedman group established an EF1a1 promoter-driven *Lox-Stop-Lox-EGFP-Rpl10a* transgenic mouse strain (Stanley et al., 2013). This mouse strain offers an opportunity to establish tissue-specific, TRAP compatible mouse strains through crossing it with cell-type specific Cre-expressing mouse lines. The establishment of lineage-specific TRAP mouse models will open broad avenues for investigating molecular mediators of cell-cell interactions in complex biological settings such as cancer immunotherapy, neurodegeneration and aging.

REFERENCES

1. Albano, R.M., Groome, N., and Smith, J.C. (1993). Activins are expressed in preimplantation mouse embryos and in ES and EC cells and are regulated on their differentiation. Development 117, 711-723.

2. Alexander, J.M., Hota, S.K., He, D., Thomas, S., Ho, L., Pennacchio, L.A., and Bruneau, B.G. (2015). Brg1 modulates enhancer activation in mesoderm lineage commitment. Development 142, 1418-1430.

3. Almog, N., and Rotter, V. (1997). Involvement of p53 in cell differentiation and development. Biochimica et biophysica acta 1333, F1-27.

4. Anastas, J.N., and Moon, R.T. (2013). WNT signalling pathways as therapeutic targets in cancer. Nat Rev Cancer 13, 11-26.

5. Anders, S., and Huber, W. (2010). Differential expression analysis for sequence count data. Genome biology 11, R106.

6. Arce, L., Yokoyama, N.N., and Waterman, M.L. (2006). Diversity of LEF/TCF action in development and disease. Oncogene 25, 7492-7504.

7. Arnold, S.J., Hofmann, U.K., Bikoff, E.K., and Robertson, E.J. (2008). Pivotal roles for eomesodermin during axis formation, epithelium-to-mesenchyme transition and endoderm specification in the mouse. Development 135, 501-511.

8. Arnold, S.J., and 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.

9. Ashcroft, G.S. (1999). Bidirectional regulation of macrophage function by TGF-beta. Microbes and infection / Institut Pasteur 1, 1275-1282.

10. Ashcroft, G.S., Yang, X., Glick, A.B., Weinstein, M., Letterio, J.L., Mizel, D.E., Anzano, M., Greenwell-Wild, T., Wahl, S.M., Deng, C., et al. (1999). Mice lacking Smad3 show accelerated wound healing and an impaired local inflammatory response. Nat Cell Biol 1, 260-266.

11. Bai, X., Kim, J., Yang, Z., Jurynec, M.J., Akie, T.E., Lee, J., LeBlanc, J., Sessa, A., Jiang, H., DiBiase, A., et al. (2010). TIF1gamma controls erythroid cell fate by regulating transcription elongation. Cell 142, 133-143.

12. Baker, J.C., and Harland, R.M. (1996). A novel mesoderm inducer, Madr2, functions in the activin signal transduction pathway. Genes & development 10, 1880-1889.

13. Barton, C.E., Tahinci, E., Barbieri, C.E., Johnson, K.N., Hanson, A.J., Jernigan, K.K., Chen, T.W., Lee, E., and Pietenpol, J.A. (2009). DeltaNp63 antagonizes p53 to regulate mesoderm induction in Xenopus laevis. Dev Biol 329, 130-139.

14. Bazzi, H., and Anderson, K.V. (2014). Acentriolar mitosis activates a p53-dependent apoptosis pathway in the mouse embryo. Proc Natl Acad Sci U S A 111, E1491-1500.

15. Beattie, G.M., Lopez, A.D., Bucay, N., Hinton, A., Firpo, M.T., King, C.C., and Hayek, A. (2005). Activin A maintains pluripotency of human embryonic stem cells in the absence of feeder layers. Stem cells (Dayton, Ohio) 23, 489-495.

16. Berghmans, S., Murphey, R.D., Wienholds, E., Neuberg, D., Kutok, J.L., Fletcher, C.D., Morris, J.P., Liu, T.X., Schulte-Merker, S., Kanki, J.P., et al. (2005). tp53 mutant zebrafish develop malignant peripheral nerve sheath tumors. Proc Natl Acad Sci U S A 102, 407-412.

17. Bernstein, B.E., Mikkelsen, T.S., Xie, X., Kamal, M., Huebert, D.J., Cuff, J., Fry, B., Meissner, A., Wernig, M., Plath, K., et al. (2006). A bivalent chromatin structure marks key developmental genes in embryonic stem cells. Cell 125, 315-326.

18. Bertero, A., Madrigal, P., Galli, A., Hubner, N.C., Moreno, I., Burks, D., Brown, S., Pedersen, R.A., Gaffney, D., Mendjan, S., et al. (2015). Activin/nodal signaling and NANOG orchestrate human embryonic stem cell fate decisions by controlling the H3K4me3 chromatin mark. Genes & development 29, 702-717.

19. Beyer, T.A., Weiss, A., Khomchuk, Y., Huang, K., Ogunjimi, A.A., Varelas, X., and Wrana, J.L. (2013). Switch enhancers interpret TGF-beta and Hippo signaling to control cell fate in human embryonic stem cells. Cell reports 5, 1611-1624.

20. Biechele, S., Cox, B.J., and Rossant, J. (2011). Porcupine homolog is required for canonical Wnt signaling and gastrulation in mouse embryos. Dev Biol 355, 275-285.

21. Bieging, K.T., Mello, S.S., and Attardi, L.D. (2014). Unravelling mechanisms of p53-mediated tumour suppression. Nat Rev Cancer 14, 359-370.

22. Bisgrove, B.W., Essner, J.J., and Yost, H.J. (1999). Regulation of midline development by antagonism of lefty and nodal signaling. Development 126, 3253-3262.

23. Brennan, J., Lu, C.C., Norris, D.P., Rodriguez, T.A., Beddington, R.S., and Robertson, E.J. (2001). Nodal signalling in the epiblast patterns the early mouse embryo. Nature 411, 965-969.

24. Brons, I.G., Smithers, L.E., Trotter, M.W., Rugg-Gunn, P., Sun, B., Chuva de Sousa Lopes, S.M., Howlett, S.K., Clarkson, A., Ahrlund-Richter, L., Pedersen, R.A., et al. (2007). Derivation of pluripotent epiblast stem cells from mammalian embryos. Nature 448, 191-195.

25. Buecker, C., Srinivasan, R., Wu, Z., Calo, E., Acampora, D., Faial, T., Simeone, A., Tan, M., Swigut, T., and Wysocka, J. (2014). Reorganization of enhancer patterns in transition from naive to primed pluripotency. Cell Stem Cell 14, 838-853.

26. Burkert, U., von Ruden, T., and Wagner, E.F. (1991). Early fetal hematopoietic development from in vitro differentiated embryonic stem cells. The New biologist 3, 698-708.

27. Cam, H., Griesmann, H., Beitzinger, M., Hofmann, L., Beinoravi-ciute-Kellner, R., Sauer, M., Huttinger-Kirchhof, N., Oswald, C., Friedl, P., Gattenlohner, S., et al. (2006). p53 family members in myogenic differentiation and rhabdomyosarcoma development. Cancer cell 10, 281-293.

28. Camus, A., Perea-Gomez, A., Moreau, A., and Collignon, J. (2006). Absence of Nodal signaling promotes precocious neural differentiation in the mouse embryo. Dev Biol 295, 743-755.

29. Chacko, B.M., Qin, B., Correia, J.J., Lam, S.S., de Caestecker, M.P., and Lin, K. (2001). The L3 loop and C-terminal phosphorylation jointly define Smad protein trimerization. Nature structural biology 8, 248-253.

30. Chacko, B.M., Qin, B.Y., Tiwari, A., Shi, G., Lam, S., Hayward, L.J., De Caestecker, M., and Lin, K. (2004). Structural basis of heteromeric smad protein assembly in TGF-beta signaling. Molecular cell 15, 813-823.

31. Chambers, I., Colby, D., Robertson, M., Nichols, J., Lee, S., Tweedie, S., and Smith, A. (2003). Functional expression cloning of Nanog, a pluripotency sustaining factor in embryonic stem cells. Cell 113, 643-655.

32. Chen, C., and Shen, M.M. (2004). Two modes by which Lefty proteins inhibit nodal signaling. Curr Biol 14, 618-624.

33. Chen, K.G., Mallon, B.S., McKay, R.D., and Robey, P.G. (2014). Human pluripotent stem cell culture: considerations for maintenance, expansion, and therapeutics. Cell Stem Cell 14, 13-26.

34. Chen, T., and Dent, S.Y. (2014). Chromatin modifiers and remodellers: regulators of cellular differentiation. Nature reviews Genetics 15, 93-106.

35. Chen, W., Jin, W., Hardegen, N., Lei, K.J., Li, L., Marinos, N., McGrady, G., and Wahl, S.M. (2003). Conversion of peripheral CD4+CD25- naive T cells to CD4+CD25+ regulatory T cells by TGF-beta induction of transcription factor Foxp3. The Journal of experimental medicine 198, 1875-1886.

36. Chen, X., Rubock, M.J., and Whitman, M. (1996). A transcriptional partner for MAD proteins in TGF-beta signalling. Nature 383, 691-696.

37. Chen, X., Weisberg, E., Fridmacher, V., Watanabe, M., Naco, G., and Whitman, M. (1997). Smad4 and FAST-1 in the assembly of activin-responsive factor. Nature 389, 85-89.

38. Cheng, A.M., Thisse, B., Thisse, C., and Wright, C.V. (2000). The lefty-related factor Xatv acts as a feedback inhibitor of nodal signaling in mesoderm induction and L-R axis development in xenopus. Development 127, 1049-1061.

39. Chiu, W.T., Charney Le, R., Blitz, I.L., Fish, M.B., Li, Y., Biesinger, J., Xie, X., and Cho, K.W. (2014). Genome-wide view of TGFbeta/Foxh1 regulation of the early mesendoderm program. Development 141, 4537-4547.

40. Ciruna, B.G., and Rossant, J. (1999). Expression of the T-box gene Eomesodermin during early mouse development. Mech Dev 81, 199-203.

41. Clevers, H., and Nusse, R. (2012). Wnt/beta-catenin signaling and disease. Cell 149, 1192-1205.

42. Cole, M.F., Johnstone, S.E., Newman, J.J., Kagey, M.H., and Young, R.A. (2008). Tcf3 is an integral component of the core regulatory circuitry of embryonic stem cells. Genes & development 22, 746-755.

43. Collignon, J., Varlet, I., and Robertson, E.J. (1996). Relationship between asymmetric nodal expression and the direction of embryonic turning. Nature 381, 155-158.

44. Cong, L., Ran, F.A., Cox, D., Lin, S., Barretto, R., Habib, N., Hsu, P.D., Wu, X., Jiang, W., Marraffini, L.A., et al. (2013). Multiplex genome engineering using CRISPR/Cas systems. Science 339, 819-823.

45. Conlon, F.L., Barth, K.S., and Robertson, E.J. (1991). A novel retrovirally induced embryonic lethal mutation in the mouse: assessment of the developmental fate of embryonic stem cells homozygous for the 413.d proviral integration. Development 111, 969-981.

46. Conlon, F.L., Lyons, K.M., Takaesu, N., Barth, K.S., Kispert, A., Herrmann, B., and Robertson, E.J. (1994). A primary requirement for nodal in the formation and maintenance of the primitive streak in the mouse. Development 120, 1919-1928.

47. Cordenonsi, M., Dupont, S., Maretto, S., Insinga, A., Imbriano, C., and Piccolo, S. (2003). Links between tumor suppressors: p53 is required for TGF-beta gene responses by cooperating with Smads. Cell 113, 301-314.

48. Costello, I., Nowotschin, S., Sun, X., Mould, A.W., Hadjantonakis, A.K., Bikoff, E.K., and 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.

49. Creyghton, M.P., Cheng, A.W., Welstead, G.G., Kooistra, T., Carey, B.W., Steine, E.J., Hanna, J., Lodato, M.A., Frampton, G.M., Sharp, P.A., et al. (2010). Histone H3K27ac separates active from poised enhancers and predicts developmental state. Proc Natl Acad Sci U S A 107, 21931-21936.

50. Dai, C., and Gu, W. (2010). p53 post-translational modification: deregulated in tumorigenesis. Trends in molecular medicine 16, 528-536.

51. Dalton, S. (2013). Signaling networks in human pluripotent stem cells. Current opinion in cell biology 25, 241-246.

52. Danilova, N., Sakamoto, K.M., and Lin, S. (2008). p53 family in development. Mech Dev 125, 919-931.

53. Das, D., Randall, R.A., and Hill, C.S. (2009). An N-terminally truncated Smad2 protein can partially compensate for loss of full-length Smad2. Biochem J 417, 205-212.

54. David, C.J., Huang, Y.H., Chen, M., Su, J., Zou, Y., Bardeesy, N., Iacobuzio-Donahue, C.A., and Massague, J. (2016). TGF-beta Tumor Suppression through a Lethal EMT. Cell 164, 1015-1030.

55. Davidson, K.C., Mason, E.A., and Pera, M.F. (2015). The pluripotent state in mouse and human. Development 142, 3090-3099.

56. de Caestecker, M.P., Yahata, T., Wang, D., Parks, W.T., Huang, S., Hill, C.S., Shioda, T., Roberts, A.B., and Lechleider, R.J. (2000). The Smad4 activation domain (SAD) is a proline-rich, p300-dependent transcriptional activation domain. The Journal of biological chemistry 275, 2115-2122.

57. De Los Angeles, A., Ferrari, F., Xi, R., Fujiwara, Y., Benvenisty, N., Deng, H., Hochedlinger, K., Jaenisch, R., Lee, S., Leitch, H.G., et al. (2015). Hallmarks of pluripotency. Nature 525, 469-478.

58. de Sousa Lopes, S.M., Carvalho, R.L., van den Driesche, S., Goumans, M.J., ten Dijke, P., and Mummery, C.L. (2003). Distribution of phosphorylated Smad2 identifies target tissues of TGF beta ligands in mouse development. Gene expression patterns : GEP 3, 355-360.

59. Dennler, S., Huet, S., and Gauthier, J.M. (1999). A short amino-acid sequence in MH1 domain is responsible for functional differences between Smad2 and Smad3. Oncogene 18, 1643-1648.

60. Dennler, S., Itoh, S., Vivien, D., ten Dijke, P., Huet, S., and Gauthier, J.M. (1998). Direct binding of Smad3 and Smad4 to critical TGF beta-inducible elements in the promoter of human plasminogen activator inhibitor-type 1 gene. The EMBO journal 17, 3091-3100.

61. Dixon, J.R., Jung, I., Selvaraj, S., Shen, Y., Antosiewicz-Bourget, J.E., Lee, A.Y., Ye, Z., Kim, A., Rajagopal, N., Xie, W., et al. (2015). Chromatin architecture reorganization during stem cell differentiation. Nature 518, 331-336.

62. Doberstein, K., Wieland, A., Lee, S.B., Blaheta, R.A., Wedel, S., Moch, H., Schraml, P., Pfeilschifter, J., Kristiansen, G., and Gutwein, P. (2011). L1-CAM expression in ccRCC correlates with shorter patients survival times and confers chemoresistance in renal cell carcinoma cells. Carcinogenesis 32, 262-270.

63. Dobin, A., Davis, C.A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M., and Gingeras, T.R. (2013). STAR: ultrafast universal RNA-seq aligner. Bioinformatics 29, 15-21.

64. Donehower, L.A., Harvey, M., Slagle, B.L., McArthur, M.J., Montgomery, C.A., Jr., Butel, J.S., and Bradley, A. (1992). Mice deficient for p53 are developmentally normal but susceptible to spontaneous tumours. Nature 356, 215-221.

65. Dotsch, V., Bernassola, F., Coutandin, D., Candi, E., and Melino, G. (2010). p63 and p73, the ancestors of p53. Cold Spring Harb Perspect Biol 2, a004887.

66. Dow, L.E., O'Rourke, K.P., Simon, J., Tschaharganeh, D.F., van Es, J.H., Clevers, H., and Lowe, S.W. (2015). Apc Restoration Promotes Cellular Differentiation and Reestablishes Crypt Homeostasis in Colorectal Cancer. Cell 161, 1539-1552.

67. Doyle, J.P., Dougherty, J.D., Heiman, M., Schmidt, E.F., Stevens, T.R., Ma, G., Bupp, S., Shrestha, P., Shah, R.D., Doughty, M.L., et al. (2008). Application of a translational profiling approach for the comparative analysis of CNS cell types. Cell 135, 749-762.

68. Dunn, N.R., Koonce, C.H., Anderson, D.C., Islam, A., Bikoff, E.K., and Robertson, E.J. (2005). Mice exclusively expressing the short isoform of Smad2 develop normally and are viable and fertile. Genes & development 19, 152-163.

69. Dunn, S.J., Martello, G., Yordanov, B., Emmott, S., and Smith, A.G. (2014). Defining an essential transcription factor program for naive pluripotency. Science 344, 1156-1160.

70. el-Deiry, W.S., Tokino, T., Velculescu, V.E., Levy, D.B., Parsons, R., Trent, J.M., Lin, D., Mercer, W.E., Kinzler, K.W., and Vogelstein, B. (1993). WAF1, a potential mediator of p53 tumor suppression. Cell 75, 817-825.

71. Erster, S., Palacios, G., Rosenquist, T., Chang, C., and Moll, U.M. (2006). Deregulated expression of DeltaNp73alpha causes early embryonic lethality. Cell death and differentiation 13, 170-173.

72. Estaras, C., Benner, C., and Jones, K.A. (2015). SMADs and YAP Compete to Control Elongation of beta-Catenin:LEF-1-Recruited RNAPII during hESC Differentiation. Molecular cell.

73. Evans, M.J., and Kaufman, M.H. (1981). Establishment in culture of pluripotential cells from mouse embryos. Nature 292, 154-156.

74. Faucourt, M., Houliston, E., Besnardeau, L., Kimelman, D., and Lepage, T. (2001). The pitx2 homeobox protein is required early for endoderm formation and nodal signaling. Dev Biol 229, 287-306.

75. Feldman, B., Gates, M.A., Egan, E.S., Dougan, S.T., Rennebeck, G., Sirotkin, H.I., Schier, A.F., and Talbot, W.S. (1998). Zebrafish organizer development and germ-layer formation require nodal-related signals. Nature 395, 181-185.

76. Ferrari, S., Marin, O., Pagano, M.A., Meggio, F., Hess, D., El-Shemerly, M., Krystyniak, A., and Pinna, L.A. (2005). Aurora-A site specificity: a study with synthetic peptide substrates. Biochem J 390, 293-302.

77. Flores, E.R., Tsai, K.Y., Crowley, D., Sengupta, S., Yang, A., McKeon, F., and Jacks, T. (2002). p63 and p73 are required for p53-dependent apoptosis in response to DNA damage. Nature 416, 560-564.

78. Fu, J., Jiang, M., Mirando, A.J., Yu, H.M., and Hsu, W. (2009). Reciprocal regulation of Wnt and Gpr177/mouse Wntless is required for embryonic axis formation. Proc Natl Acad Sci U S A 106, 18598-18603.

79. Funa, N.S., Schachter, K.A., Lerdrup, M., Ekberg, J., Hess, K., Dietrich, N., Honore, C., Hansen, K., and Semb, H. (2015). beta-Catenin Regulates Primitive Streak Induction through Collaborative Interactions with SMAD2/SMAD3 and OCT4. Cell Stem Cell.

80. Germain, S., Howell, M., Esslemont, G.M., and Hill, C.S. (2000). Homeodomain and winged-helix transcription factors recruit activated Smads to distinct promoter elements via a common Smad interaction motif. Genes & development 14, 435-451.

81. Glinka, A., Delius, H., Blumenstock, C., and Niehrs, C. (1996). Combinatorial signalling by Xwnt-11 and Xnr3 in the organizer epithelium. Mech Dev 60, 221-231.

82. Gritsman, K., Zhang, J., Cheng, S., Heckscher, E., Talbot, W.S., and Schier, A.F. (1999). The EGF-CFC protein one-eyed pinhead is essential for nodal signaling. Cell 97, 121-132.

83. Gu, Z., Nomura, M., Simpson, B.B., Lei, H., Feijen, A., van den Eijnden-van Raaij, J., Donahoe, P.K., and Li, E. (1998). The type I activin receptor ActRIB is required for egg cylinder organization and gastrulation in the mouse. Genes & development 12, 844-857.

84. Gu, Z., Reynolds, E.M., Song, J., Lei, H., Feijen, A., Yu, L., He, W., MacLaughlin, D.T., van den Eijnden-van Raaij, J., Donahoe, P.K., et al. (1999). The type I serine/threonine kinase receptor ActRIA (ALK2) is required for gastrulation of the mouse embryo. Development 126, 2551-2561.

85. Guo, X., Ramirez, A., Waddell, D.S., Li, Z., Liu, X., and Wang, X.F. (2008). Axin and GSK3- control Smad3 protein stability and modulate TGF-signaling. Genes & development 22, 106-120.

86. Guo, X., Waddell, D.S., Wang, W., Wang, Z., Liberati, N.T., Yong, S., Liu, X., and Wang, X.F. (2008). Ligand-dependent ubiquitination of Smad3 is regulated by casein kinase 1 gamma 2, an inhibitor of TGF-beta signaling. Oncogene 27, 7235-7247.

87. Hart, A.H., Hartley, L., Sourris, K., Stadler, E.S., Li, R., Stanley, E.G., Tam, P.P., Elefanty, A.G., and Robb, L. (2002). Mixl1 is required for axial mesendoderm morphogenesis and patterning in the murine embryo. Development 129, 3597-3608.

88. Hata, A., Lo, R.S., Wotton, D., Lagna, G., and Massague, J. (1997). Mutations increasing autoinhibition inactivate tumour suppressors Smad2 and Smad4. Nature 388, 82-87.

89. Hayashi, H., Abdollah, S., Qiu, Y., Cai, J., Xu, Y.Y., Grinnell, B.W., Richardson, M.A., Topper, J.N., Gimbrone, M.A., Jr., Wrana, J.L., et al. (1997). The MAD-related protein Smad7 associates with the TGFbeta receptor and functions as an antagonist of TGFbeta signaling. Cell 89, 1165-1173.

90. He, T.C., Sparks, A.B., Rago, C., Hermeking, H., Zawel, L., da Costa, L.T., Morin, P.J., Vogelstein, B., and Kinzler, K.W. (1998). Identification of c-MYC as a target of the APC pathway. Science 281, 1509-1512.

91. He, W., Dorn, D.C., Erdjument-Bromage, H., Tempst, P., Moore, M.A., and Massague, J. (2006). Hematopoiesis controlled by distinct TIF1gamma and Smad4 branches of the TGFbeta pathway. Cell 125, 929-941.

92. Heikkinen, P.T., Nummela, M., Leivonen, S.K., Westermarck, J., Hill, C.S., Kahari, V.M., and Jaakkola, P.M. (2010). Hypoxia-activated Smad3-specific dephosphorylation by PP2A. The Journal of biological chemistry 285, 3740-3749.

93. Heiman, M., Schaefer, A., Gong, S., Peterson, J.D., Day, M., Ramsey, K.E., Suarez-Farinas, M., Schwarz, C., Stephan, D.A., Surmeier, D.J., et al. (2008). A translational profiling approach for the molecular characterization of CNS cell types. Cell 135, 738-748.

94. Heinz, S., Benner, C., Spann, N., Bertolino, E., Lin, Y.C., Laslo, P., Cheng, J.X., Murre, C., Singh, H., and Glass, C.K. (2010). Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. Molecular cell 38, 576-589.

95. Heyer, J., Escalante-Alcalde, D., Lia, M., Boettinger, E., Edelmann, W., Stewart, C.L., and Kucherlapati, R. (1999). Postgastrulation Smad2-deficient embryos show defects in embryo turning and anterior morphogenesis. Proc Natl Acad Sci U S A 96, 12595-12600.

96. Ho, R., Papp, B., Hoffman, J.A., Merrill, B.J., and Plath, K. (2013). Stage-specific regulation of reprogramming to induced pluripotent stem cells by Wnt signaling and T cell factor proteins. Cell reports 3, 2113-2126.

97. Hong, H., Takahashi, K., Ichisaka, T., Aoi, T., Kanagawa, O., Nakagawa, M., Okita, K., and Yamanaka, S. (2009). Suppression of induced pluripotent stem cell generation by the p53-p21 pathway. Nature 460, 1132-1135.

98. Hoodless, P.A., Pye, M., Chazaud, C., Labbe, E., Attisano, L., Rossant, J., and Wrana, J.L. (2001). FoxH1 (Fast) functions to specify the anterior primitive streak in the mouse. Genes & development 15, 1257-1271.

99. Hoppler, S., and Kavanagh, C.L. (2007). Wnt signalling: variety at the core. Journal of cell science 120, 385-393.

100. Hsieh, J.C., Lee, L., Zhang, L., Wefer, S., Brown, K., DeRossi, C., Wines, M.E., Rosenquist, T., and Holdener, B.C. (2003). Mesd encodes an LRP5/6 chaperone essential for specification of mouse embryonic polarity. Cell 112, 355-367.

101. Hsu, P.D., Scott, D.A., Weinstein, J.A., Ran, F.A., Konermann, S., Agarwala, V., Li, Y., Fine, E.J., Wu, X., Shalem, O., et al. (2013). DNA targeting specificity of RNA-guided Cas9 nucleases. Nature biotechnology 31, 827-832.

102. Hu, W., Feng, Z., Teresky, A.K., and Levine, A.J. (2007). p53 regulates maternal reproduction through LIF. Nature 450, 721-724.

103. Huelsken, J., Vogel, R., Brinkmann, V., Erdmann, B., Birchmeier, C., and Birchmeier, W. (2000). Requirement for beta-catenin in anterior-posterior axis formation in mice. The Journal of cell biology 148, 567-578.

104. Huttinger-Kirchhof, N., Cam, H., Griesmann, H., Hofmann, L., Beitzinger, M., and Stiewe, T. (2006). The p53 family inhibitor DeltaNp73 interferes with multiple developmental programs. Cell death and differentiation 13, 174-177.

105. Ichida, J.K., Blanchard, J., Lam, K., Son, E.Y., Chung, J.E., Egli, D., Loh, K.M., Carter, A.C., Di Giorgio, F.P., Koszka, K., et al. (2009). A small-molecule inhibitor of tgf-Beta signaling replaces sox2 in reprogramming by inducing nanog. Cell Stem Cell 5, 491-503.

106. Ingolia, N.T. (2014). Ribosome profiling: new views of translation, from single codons to genome scale. Nature reviews Genetics 15, 205-213.

107. Ingolia, N.T., Ghaemmaghami, S., Newman, J.R., and Weissman, J.S. (2009). Genome-wide analysis in vivo of translation with nucleotide resolution using ribosome profiling. Science 324, 218-223.

108. Ingolia, N.T., Lareau, L.F., and Weissman, J.S. (2011). Ribosome profiling of mouse embryonic stem cells reveals the complexity and dynamics of mammalian proteomes. Cell 147, 789-802.

109. Iratni, R., Yan, Y.T., Chen, C., Ding, J., Zhang, Y., Price, S.M., Reinberg, D., and Shen, M.M. (2002). Inhibition of excess nodal signaling during mouse gastrulation by the transcriptional corepressor DRAP1. Science 298, 1996-1999.

110. Jacks, T., Remington, L., Williams, B.O., Schmitt, E.M., Halachmi, S., Bronson, R.T., and Weinberg, R.A. (1994). Tumor spectrum analysis in p53-mutant mice. Curr Biol 4, 1-7.

111. James, D., Levine, A.J., Besser, D., and Hemmati-Brivanlou, A. (2005). TGFbeta/activin/nodal signaling is necessary for the maintenance of pluripotency in human embryonic stem cells. Development 132, 1273-1282.

112. Jayaraman, L., and Massague, J. (2000). Distinct oligomeric states of SMAD proteins in the transforming growth factor-beta pathway. The Journal of biological chemistry 275, 40710-40717.

113. Jho, E.H., Zhang, T., Domon, C., Joo, C.K., Freund, J.N., and Costantini, F. (2002). Wnt/beta-catenin/Tcf signaling induces the transcription of Axin2, a negative regulator of the signaling pathway. Mol Cell Biol 22, 1172-1183.

114. Jiang, H., Shukla, A., Wang, X., Chen, W.Y., Bernstein, B.E., and Roeder, R.G. (2011). Role for Dpy-30 in ES cell-fate specification by regulation of H3K4 methylation within bivalent domains. Cell 144, 513-525.

115. Jones, S.N., Roe, A.E., Donehower, L.A., and Bradley, A. (1995). Rescue of embryonic lethality in Mdm2-deficient mice by absence of p53. Nature 378, 206-208.

116. Kang, H.J., Feng, Z., Sun, Y., Atwal, G., Murphy, M.E., Rebbeck, T.R., Rosenwaks, Z., Levine, A.J., and Hu, W. (2009). Single-nucleotide polymorphisms in the p53 pathway regulate fertility in humans. Proc Natl Acad Sci U S A 106, 9761-9766.

117. Katayama, H., Sasai, K., Kawai, H., Yuan, Z.M., Bondaruk, J., Suzuki, F., Fujii, S., Arlinghaus, R.B., Czerniak, B.A., and Sen, S. (2004). Phosphorylation by aurora kinase A induces Mdm2-mediated destabilization and inhibition of p53. Nature genetics 36, 55-62.

118. Kawamura, T., Suzuki, J., Wang, Y.V., Menendez, S., Morera, L.B., Raya, A., Wahl, G.M., and Izpisua Belmonte, J.C. (2009). Linking the p53 tumour suppressor pathway to somatic cell reprogramming. Nature 460, 1140-1144.

119. Kawase, T., Ohki, R., Shibata, T., Tsutsumi, S., Kamimura, N., Inazawa, J., Ohta, T., Ichikawa, H., Aburatani, H., Tashiro, F., et al. (2009). PH domain-only protein PHLDA3 is a p53-regulated repressor of Akt. Cell 136, 535-550.

120. Kelly, O.G., Pinson, K.I., and Skarnes, W.C. (2004). The Wnt co-receptors Lrp5 and Lrp6 are essential for gastrulation in mice. Development 131, 2803-2815.

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

122. Kim, S.W., Yoon, S.J., Chuong, E., Oyolu, C., Wills, A.E., Gupta, R., and Baker, J. (2011). Chromatin and transcriptional signatures for Nodal signaling during endoderm formation in hESCs. Dev Biol 357, 492-504.

123. Kimelman, D., and Griffin, K.J. (2000). Vertebrate mesendoderm induction and patterning. Current opinion in genetics & development 10, 350-356.

124. Kimura, T., Tomooka, M., Yamano, N., Murayama, K., Matoba, S., Umehara, H., Kanai, Y., and Nakano, T. (2008). AKT signaling promotes derivation of embryonic germ cells from primordial germ cells. Development 135, 869-879.

125. Korinek, V., Barker, N., Willert, K., Molenaar, M., Roose, J., Wagenaar, G., Markman, M., Lamers, W., Destree, O., and Clevers, H. (1998). Two members of the Tcf family implicated in Wnt/beta-catenin signaling during embryogenesis in the mouse. Mol Cell Biol 18, 1248-1256.

126. Kunath, T., Saba-El-Leil, M.K., Almousailleakh, M., Wray, J., Meloche, S., and Smith, A. (2007). FGF stimulation of the Erk1/2 signalling cascade triggers transition of pluripotent embryonic stem cells from self-renewal to lineage commitment. Development 134, 2895-2902.

127. Kurokawa, M., Mitani, K., Irie, K., Matsuyama, T., Takahashi, T., Chiba, S., Yazaki, Y., Matsumoto, K., and Hirai, H. (1998). The oncoprotein Evi-1 represses TGF-beta signalling by inhibiting Smad3. Nature 394, 92-96.

128. Labbe, E., Letamendia, A., and Attisano, L. (2000). Association of Smads with lymphoid enhancer binding factor 1/T cell-specific factor mediates cooperative signaling by the transforming growth factor-beta and wnt pathways. Proc Natl Acad Sci U S A 97, 8358-8363.

129. Labbe, E., Silvestri, C., Hoodless, P.A., Wrana, J.L., and Attisano, L. (1998). Smad2 and Smad3 positively and negatively regulate TGF beta-dependent transcription through the forkhead DNA-binding protein FAST2. Molecular cell 2, 109-120.

130. Lane, D., and Levine, A. (2010). p53 Research: the past thirty years and the next thirty years. Cold Spring Harb Perspect Biol 2, a000893.

131. Langheinrich, U., Hennen, E., Stott, G., and Vacun, G. (2002). Zebrafish as a model organism for the identification and characterization of drugs and genes affecting p53 signaling. Curr Biol 12, 2023-2028.

132. Langmead, B., and Salzberg, S.L. (2012). Fast gapped-read alignment with Bowtie 2. Nature methods 9, 357-359.

133. Lawson, K.A., Meneses, J.J., and Pedersen, R.A. (1991). Clonal analysis of epiblast fate during germ layer formation in the mouse embryo. Development 113, 891-911.

134. Lee, D.F., Su, J., Ang, Y.S., Carvajal-Vergara, X., Mulero-Navarro, S., Pereira, C.F., Gingold, J., Wang, H.L., Zhao, R., Sevilla, A., et al. (2012). Regulation of embryonic and induced pluripotency by aurora kinase-p53 signaling. Cell Stem Cell 11, 179-194.

135. Lee, H., and Kimelman, D. (2002). A dominant-negative form of p63 is required for epidermal proliferation in zebrafish. Developmental cell 2, 607-616.

136. Lee, K.H., Li, M., Michalowski, A.M., Zhang, X., Liao, H., Chen, L., Xu, Y., Wu, X., and Huang, J. (2010). A genomewide study identifies the Wnt signaling pathway as a major target of p53 in murine embryonic stem cells. Proc Natl Acad Sci U S A 107, 69-74.

137. Lee, M.H., Reynisdottir, I., and Massague, J. (1995). Cloning of p57KIP2, a cyclin-dependent kinase inhibitor with unique domain structure and tissue distribution. Genes & development 9, 639-649.

138. Leeb, M., Dietmann, S., Paramor, M., Niwa, H., and Smith, A. (2014). Genetic exploration of the exit from self-renewal using haploid embryonic stem cells. Cell Stem Cell 14, 385-393.

139. Lemaire, P., Darras, S., Caillol, D., and Kodjabachian, L. (1998). A role for the vegetally expressed Xenopus gene Mix.1 in endoderm formation and in the restriction of mesoderm to the marginal zone. Development 125, 2371-2380.

140. Levy, D., Adamovich, Y., Reuven, N., and Shaul, Y. (2007). The Yes-associated protein 1 stabilizes p73 by preventing Itch-mediated ubiquitination of p73. Cell death and differentiation 14, 743-751.

141. Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R., and Genome Project Data Processing, S. (2009). The Sequence Alignment/Map format and SAMtools. Bioinformatics 25, 2078-2079.

142. Li, M., He, Y., Dubois, W., Wu, X., Shi, J., and Huang, J. (2012). Distinct regulatory mechanisms and functions for p53-activated and p53-repressed DNA damage response genes in embryonic stem cells. Molecular cell 46, 30-42.

143. Lian, I., Kim, J., Okazawa, H., Zhao, J., Zhao, B., Yu, J., Chinnaiyan, A., Israel, M.A., Goldstein, L.S., Abujarour, R., et al. (2010). The role of YAP transcription coactivator in regulating stem cell self-renewal and differentiation. Genes & development 24, 1106-1118.

144. Lin, S.W., Lee, M.T., Ke, F.C., Lee, P.P., Huang, C.J., Ip, M.M., Chen, L., and Hwang, J.J. (2000). TGFbeta1 stimulates the secretion of matrix metalloproteinase 2 (MMP2) and the invasive behavior in human ovarian cancer cells, which is suppressed by MMP inhibitor BB3103. Clinical & experimental metastasis 18, 493-499.

145. Lin, T., Chao, C., Saito, S., Mazur, S.J., Murphy, M.E., Appella, E., and Xu, Y. (2005). p53 induces differentiation of mouse embryonic stem cells by suppressing Nanog expression. Nat Cell Biol 7, 165-171.

146. Lindsley, R.C., Gill, J.G., Kyba, M., Murphy, T.L., and Murphy, K.M. (2006). Canonical Wnt signaling is required for development of embryonic stem cell-derived mesoderm. Development 133, 3787-3796.

147. Ling, N., Ying, S.Y., Ueno, N., Shimasaki, S., Esch, F., Hotta, M., and Guillemin, R. (1986). Pituitary FSH is released by a heterodimer of the beta-subunits from the two forms of inhibin. Nature 321, 779-782.

148. Liu, B., Dou, C.L., Prabhu, L., and Lai, E. (1999). FAST-2 is a mammalian winged-helix protein which mediates transforming growth factor beta signals. Mol Cell Biol 19, 424-430.

149. Liu, F., Pouponnot, C., and Massague, J. (1997). Dual role of the Smad4/DPC4 tumor suppressor in TGFbeta-inducible transcriptional complexes. Genes & development 11, 3157-3167.

150. Liu, P., Wakamiya, M., Shea, M.J., Albrecht, U., Behringer, R.R., and Bradley, A. (1999). Requirement for Wnt3 in vertebrate axis formation. Nature genetics 22, 361-365.

151. Liu, Q., Kaneko, S., Yang, L., Feldman, R.I., Nicosia, S.V., Chen, J., and Cheng, J.Q. (2004). Aurora-A abrogation of p53 DNA binding and transactivation activity by phosphorylation of serine 215. The Journal of biological chemistry 279, 52175-52182.

152. Love, M.I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome biology 15, 550.

153. Lustig, B., Jerchow, B., Sachs, M., Weiler, S., Pietsch, T., Karsten, U., van de Wetering, M., Clevers, H., Schlag, P.M., Birchmeier, W., et al. (2002). Negative feedback loop of Wnt signaling through upregulation of conductin/axin2 in colorectal and liver tumors. Mol Cell Biol 22, 1184-1193.

154. Lutzker, S.G., and Levine, A.J. (1996). A functionally inactive p53 protein in teratocarcinoma cells is activated by either DNA damage or cellular differentiation. Nat Med 2, 804-810.

155. MacDonald, B.T., Tamai, K., and He, X. (2009). Wnt/beta-catenin signaling: components, mechanisms, and diseases. Developmental cell 17, 9-26.

156. Machanick, P., and Bailey, T.L. (2011). MEME-ChIP: motif analysis of large DNA datasets. Bioinformatics 27, 1696-1697.

157. Macias, M.J., Martin-Malpartida, P., and Massague, J. (2015). Structural determinants of Smad function in TGF-beta signaling. Trends in biochemical sciences 40, 296-308.

158. Macias-Silva, M., Li, W., Leu, J.I., Crissey, M.A., and Taub, R. (2002). Up-regulated transcriptional repressors SnoN and Ski bind Smad proteins to antagonize transforming growth factor-beta signals during liver regeneration. The Journal of biological chemistry 277, 28483-28490.

159. Maherali, N., and Hochedlinger, K. (2009). Tgfbeta signal inhibition cooperates in the induction of iPSCs and replaces Sox2 and cMyc. Curr Biol 19, 1718-1723.

160. Malhotra, N., Robertson, E., and Kang, J. (2010). SMAD2 is essential for TGF beta-mediated Th17 cell generation. The Journal of biological chemistry 285, 29044-29048.

161. Mali, P., Yang, L., Esvelt, K.M., Aach, J., Guell, M., DiCarlo, J.E., Norville, J.E., and Church, G.M. (2013). RNA-guided human genome engineering via Cas9. Science 339, 823-826.

162. Marcel, V., Dichtel-Danjoy, M.L., Sagne, C., Hafsi, H., Ma, D., Ortiz-Cuaran, S., Olivier, M., Hall, J., Mollereau, B., Hainaut, P., et al. (2011). Biological functions of p53 isoforms through evolution: lessons from animal and cellular models. Cell death and differentiation 18, 1815-1824.

163. Marion, R.M., Strati, K., Li, H., Murga, M., Blanco, R., Ortega, S., Fernandez-Capetillo, O., Serrano, M., and Blasco, M.A. (2009). A p53-mediated DNA damage response limits reprogramming to ensure iPS cell genomic integrity. Nature 460, 1149-1153.

164. Marson, A., Foreman, R., Chevalier, B., Bilodeau, S., Kahn, M., Young, R.A., and Jaenisch, R. (2008). Wnt signaling promotes reprogramming of somatic cells to pluripotency. Cell Stem Cell 3, 132-135.

165. Martin, G.R. (1981). Isolation of a pluripotent cell line from early mouse embryos cultured in medium conditioned by teratocarcinoma stem cells. Proc Natl Acad Sci U S A 78, 7634-7638.

166. Massague, J. (2012). TGFbeta signalling in context. Nat Rev Mol Cell Biol 13, 616-630.

167. Massague, J., and Obenauf, A.C. (2016). Metastatic colonization by circulating tumour cells. Nature 529, 298-306.

168. Massague, J., Seoane, J., and Wotton, D. (2005). Smad transcription factors. Genes & development 19, 2783-2810.

169. Matzuk, M.M., Kumar, T.R., and Bradley, A. (1995). Different phenotypes for mice deficient in either activins or activin receptor type II. Nature 374, 356-360.

170. Matzuk, M.M., Kumar, T.R., Vassalli, A., Bickenbach, J.R., Roop, D.R., Jaenisch, R., and Bradley, A. (1995). Functional analysis of activins during mammalian development. Nature 374, 354-356.

171. McDowell, N., and Gurdon, J.B. (1999). Activin as a morphogen in Xenopus mesoderm induction. Seminars in cell & developmental biology 10, 311-317.

172. McMahon, A.P., and Bradley, A. (1990). The Wnt-1 (int-1) proto-oncogene is required for development of a large region of the mouse brain. Cell 62, 1073-1085.

173. McMahon, A.P., Joyner, A.L., Bradley, A., and McMahon, J.A. (1992). The midbrain-hindbrain phenotype of Wnt-1-/Wnt-1- mice results from stepwise deletion of engrailed-expressing cells by 9.5 days postcoitum. Cell 69, 581-595.

174. Menendez, D., Inga, A., and Resnick, M.A. (2009). The expanding universe of p53 targets. Nat Rev Cancer 9, 724-737.

175. Meno, C., Gritsman, K., Ohishi, S., Ohfuji, Y., Heckscher, E., Mochida, K., Shimono, A., Kondoh, H., Talbot, W.S., Robertson, E.J., et al. (1999). Mouse Lefty2 and zebrafish antivin are feedback inhibitors of nodal signaling during vertebrate gastrulation. Molecular cell 4, 287-298.

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

177. Mesnard, D., Guzman-Ayala, M., and Constam, D.B. (2006). Nodal specifies embryonic visceral endoderm and sustains pluripotent cells in the epiblast before overt axial patterning. Development 133, 2497-2505.

178. Migeotte, I., Grego-Bessa, J., and Anderson, K.V. (2011). Rac1 mediates morphogenetic responses to intercellular signals in the gastrulating mouse embryo. Development 138, 3011-3020.

179. Mills, A.A., Zheng, B., Wang, X.J., Vogel, H., Roop, D.R., and Bradley, A. (1999). p63 is a p53 homologue required for limb and epidermal morphogenesis. Nature 398, 708-713.

180. Mizuide, M., Hara, T., Furuya, T., Takeda, M., Kusanagi, K., Inada, Y., Mori, M., Imamura, T., Miyazawa, K., and Miyazono, K. (2003). Two short segments of Smad3 are important for specific interaction of Smad3 with c-Ski and SnoN. The Journal of biological chemistry 278, 531-536.

181. Momand, J., Zambetti, G.P., Olson, D.C., George, D., and Levine, A.J. (1992). The mdm-2 oncogene product forms a complex with the p53 protein and inhibits p53-mediated transactivation. Cell 69, 1237-1245.

182. Monaghan, A.P., Kaestner, K.H., Grau, E., and Schutz, G. (1993). Postimplantation expression patterns indicate a role for the mouse forkhead/HNF-3 alpha, beta and gamma genes in determination of the definitive endoderm, chordamesoderm and neuroectoderm. Development 119, 567-578.

183. Montes de Oca Luna, R., Wagner, D.S., and Lozano, G. (1995). Rescue of early embryonic lethality in mdm2-deficient mice by deletion of p53. Nature 378, 203-206.

184. Morsut, L., Yan, K.P., Enzo, E., Aragona, M., Soligo, S.M., Wendling, O., Mark, M., Khetchoumian, K., Bressan, G., Chambon, P., et al. (2010). Negative control of Smad activity by ectodermin/Tif1gamma patterns the mammalian embryo. Development 137, 2571-2578.

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

186. Nakao, A., Afrakhte, M., Moren, A., Nakayama, T., Christian, J.L., Heuchel, R., Itoh, S., Kawabata, M., Heldin, N.E., Heldin, C.H., et al. (1997). Identification of Smad7, a TGFbeta-inducible antagonist of TGF-beta signalling. Nature 389, 631-635.

187. Nawshad, A., Medici, D., Liu, C.C., and Hay, E.D. (2007). TGFbeta3 inhibits E-cadherin gene expression in palate medial-edge epithelial cells through a Smad2-Smad4-LEF1 transcription complex. Journal of cell science 120, 1646-1653.

188. Negoescu, A., Lafeuillade, B., Pellerin, S., Chambaz, E.M., and Feige, J.J. (1995). Transforming growth factors beta stimulate both thrombospondin-1 and CISP/thrombospondin-2 synthesis by bovine adrenocortical cells. Experimental cell research 217, 404-409.

189. Nguyen, D.X., Bos, P.D., and Massague, J. (2009). Metastasis: from dissemination to organ-specific colonization. Nat Rev Cancer 9, 274-284.

190. Nichols, J., and Smith, A. (2009). Naive and primed pluripotent states. Cell Stem Cell 4, 487-492.

191. Niida, A., Hiroko, T., Kasai, M., Furukawa, Y., Nakamura, Y., Suzuki, Y., Sugano, S., and Akiyama, T. (2004). DKK1, a negative regulator of Wnt signaling, is a target of the beta-catenin/TCF pathway. Oncogene 23, 8520-8526.

192. Nishikawa, S.I., Nishikawa, S., Hirashima, M., Matsuyoshi, N., and Kodama, H. (1998). Progressive lineage analysis by cell sorting and culture identifies FLK1+VE-cadherin+ cells at a diverging point of endothelial and hemopoietic lineages. Development 125, 1747-1757.

193. Nishioka, N., Inoue, K., Adachi, K., Kiyonari, H., Ota, M., Ralston, A., Yabuta, N., Hirahara, S., Stephenson, R.O., Ogonuki, N., et al. (2009). The Hippo signaling pathway components Lats and Yap pattern Tead4 activity to distinguish mouse trophectoderm from inner cell mass. Developmental cell 16, 398-410.

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

195. Niwa, H., Ogawa, K., Shimosato, D., and Adachi, K. (2009). A parallel circuit of LIF signalling pathways maintains pluripotency of mouse ES cells. Nature 460, 118-122.

196. Noggle, S.A., James, D., and Brivanlou, A.H. (2005). A molecular basis for human embryonic stem cell pluripotency. Stem cell reviews 1, 111-118.

197. Nomura, M., and Li, E. (1998). Smad2 role in mesoderm formation, left-right patterning and craniofacial development. Nature 393, 786-790.

198. Norris, D.P., and Robertson, E.J. (1999). Asymmetric and node-specific nodal expression patterns are controlled by two distinct cis-acting regulatory elements. Genes & development 13, 1575-1588.

199. Nowotschin, S., Costello, I., Piliszek, A., Kwon, G.S., Mao, C.A., Klein, W.H., Robertson, E.J., and Hadjantonakis, A.K. (2013). The T-box transcription factor Eomesodermin is essential for AVE induction in the mouse embryo. Genes & development 27, 997-1002.

200. Nusse, R., and Varmus, H.E. (1982). Many tumors induced by the mouse mammary tumor virus contain a provirus integrated in the same region of the host genome. Cell 31, 99-109.

201. Obenauf, A.C., and Massague, J. (2015). Surviving at a distance: organ specific metastasis. Trends in cancer 1, 76-91.

202. Obenauf, A.C., Zou, Y., Ji, A.L., Vanharanta, S., Shu, W., Shi, H., Kong, X., Bosenberg, M.C., Wiesner, T., Rosen, N., et al. (2015). Therapy-induced tumour secretomes promote resistance and tumour progression. Nature 520, 368-372.

203. Ogawa, K., Saito, A., Matsui, H., Suzuki, H., Ohtsuka, S., Shimosato, D., Morishita, Y., Watabe, T., Niwa, H., and Miyazono, K. (2007). Activin-Nodal signaling is involved in propagation of mouse embryonic stem cells. Journal of cell science 120, 55-65.

204. Oh, S.P., and Li, E. (1997). The signaling pathway mediated by the type IIB activin receptor controls axial patterning and lateral asymmetry in the mouse. Genes & development 11, 1812-1826.

205. Ohuchi, H., Yoshioka, H., Tanaka, A., Kawakami, Y., Nohno, T., and Noji, S. (1994). Involvement of androgen-induced growth factor (FGF-8) gene in mouse embryogenesis and morphogenesis. Biochemical and biophysical research communications 204, 882-888.

206. Okamoto, K., and Beach, D. (1994). Cyclin G is a transcriptional target of the p53 tumor suppressor protein. The EMBO journal 13, 4816-4822.

207. Olson, E.N., Sternberg, E., Hu, J.S., Spizz, G., and Wilcox, C. (1986). Regulation of myogenic differentiation by type beta transforming growth factor. The Journal of cell biology 103, 1799-1805.

208. Osada, S.I., and Wright, C.V. (1999). Xenopus nodal-related signaling is essential for mesendodermal patterning during early embryogenesis. Development 126, 3229-3240.

209. Oshimori, N., and Fuchs, E. (2012). The harmonies played by TGF-beta in stem cell biology. Cell Stem Cell 11, 751-764.

210. Oskarsson, T., Batlle, E., and Massague, J. (2014). Metastatic stem cells: sources, niches, and vital pathways. Cell Stem Cell 14, 306-321.

211. Pan, H., Dung, H.N., Hsu, H.M., Hsiao, K.M., and Chen, L.Y. (2003). Cloning and developmental expression of p73 cDNA in zebrafish. Biochemical and biophysical research communications 307, 395-400.

212. Parant, J., Chavez-Reyes, A., Little, N.A., Yan, W., Reinke, V., Jochemsen, A.G., and Lozano, G. (2001). Rescue of embryonic lethality in Mdm4-null mice by loss of Trp53 suggests a nonoverlapping pathway with MDM2 to regulate p53. Nature genetics 29, 92-95.

213. Paria, B.C., Jones, K.L., Flanders, K.C., and Dey, S.K. (1992). Localization and binding of transforming growth factor-beta isoforms in mouse preimplantation embryos and in delayed and activated blastocysts. Dev Biol 151, 91-104.

214. Pauklin, S., and Vallier, L. (2013). The cell-cycle state of stem cells determines cell fate propensity. Cell 155, 135-147.

215. Pauklin, S., and Vallier, L. (2015). Activin/Nodal signalling in stem cells. Development 142, 607-619.

216. Pavletich, N.P., Chambers, K.A., and Pabo, C.O. (1993). The DNA-binding domain of p53 contains the four conserved regions and the major mutation hot spots. Genes & development 7, 2556-2564.

217. Perea-Gomez, A., Vella, F.D., Shawlot, W., Oulad-Abdelghani, M., Chazaud, C., Meno, C., Pfister, V., Chen, L., Robertson, E., Hamada, H., et al. (2002). Nodal antagonists in the anterior visceral endoderm prevent the formation of multiple primitive streaks. Developmental cell 3, 745-756.

218. Pereira, L., Yi, F., and Merrill, B.J. (2006). Repression of Nanog gene transcription by Tcf3 limits embryonic stem cell self-renewal. Mol Cell Biol 26, 7479-7491.

219. Piedra, M.E., Icardo, J.M., Albajar, M., Rodriguez-Rey, J.C., and Ros, M.A. (1998). Pitx2 participates in the late phase of the pathway controlling left-right asymmetry. Cell 94, 319-324.

220. Piek, E., Ju, W.J., Heyer, J., Escalante-Alcalde, D., Stewart, C.L., Weinstein, M., Deng, C., Kucherlapati, R., Bottinger, E.P., and Roberts, A.B. (2001). Functional characterization of transforming growth factor beta signaling in Smad2- and Smad3-deficient fibroblasts. The Journal of biological chemistry 276, 19945-19953.

221. Pierreux, C.E., Nicolas, F.J., and Hill, C.S. (2000). Transforming growth factor beta-independent shuttling of Smad4 between the cytoplasm and nucleus. Mol Cell Biol 20, 9041-9054.

222. Polyak, K., Kato, J.Y., Solomon, M.J., Sherr, C.J., Massague, J., Roberts, J.M., and Koff, A. (1994). p27Kip1, a cyclin-Cdk inhibitor, links transforming growth factor-beta and contact inhibition to cell cycle arrest. Genes & development 8, 9-22.

223. Polyak, K., Lee, M.H., Erdjument-Bromage, H., Koff, A., Roberts, J.M., Tempst, P., and Massague, J. (1994). Cloning of p27Kip1, a cyclin-dependent kinase inhibitor and a potential mediator of extracellular antimitogenic signals. Cell 78, 59-66.

224. Pouponnot, C., Jayaraman, L., and Massague, J. (1998). Physical and functional interaction of SMADs and p300/CBP. The Journal of biological chemistry 273, 22865-22868.

225. Purdie, C.A., Harrison, D.J., Peter, A., Dobbie, L., White, S., Howie, S.E., Salter, D.M., Bird, C.C., Wyllie, A.H., Hooper, M.L., et al. (1994). Tumour incidence, spectrum and ploidy in mice with a large deletion in the p53 gene. Oncogene 9, 603-609.

226. Qin, H., Hejna, M., Liu, Y., Percharde, M., Wossidlo, M., Blouin, L., Durruthy-Durruthy, J., Wong, P., Qi, Z., Yu, J., et al. (2016). YAP Induces Human Naive Pluripotency. Cell reports.

227. Qin, H., Yu, T., Qing, T., Liu, Y., Zhao, Y., Cai, J., Li, J., Song, Z., Qu, X., Zhou, P., et al. (2007). Regulation of apoptosis and differentiation by p53 in human embryonic stem cells. The Journal of biological chemistry 282, 5842-5852.

228. Quinlan, A.R., and Hall, I.M. (2010). BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics 26, 841-842.

229. Ran, F.A., Hsu, P.D., Wright, J., Agarwala, V., Scott, D.A., and Zhang, F. (2013). Genome engineering using the CRISPR-Cas9 system. Nature protocols 8, 2281-2308.

230. Randall, R.A., Germain, S., Inman, G.J., Bates, P.A., and Hill, C.S. (2002). Different Smad2 partners bind a common hydrophobic pocket in Smad2 via a defined proline-rich motif. The EMBO journal 21, 145-156.

231. Reid, C.D., Zhang, Y., Sheets, M.D., and Kessler, D.S. (2012). Transcriptional integration of Wnt and Nodal pathways in establishment of the Spemann organizer. Dev Biol 368, 231-241.

232. Rentzsch, F., Kramer, C., and Hammerschmidt, M. (2003). Specific and conserved roles of TAp73 during zebrafish development. Gene 323, 19-30.

233. Riley, T., Sontag, E., Chen, P., and Levine, A. (2008). Transcriptional control of human p53-regulated genes. Nat Rev Mol Cell Biol 9, 402-412.

234. Rivera-Perez, J.A., and Magnuson, T. (2005). Primitive streak formation in mice is preceded by localized activation of Brachyury and Wnt3. Dev Biol 288, 363-371.

235. Rivera-Perez, J.A., Wakamiya, M., and Behringer, R.R. (1999). Goosecoid acts cell autonomously in mesenchyme-derived tissues during craniofacial development. Development 126, 3811-3821.

236. Robinson, J.T., Thorvaldsdottir, H., Winckler, W., Guttman, M., Lander, E.S., Getz, G., and Mesirov, J.P. (2011). Integrative genomics viewer. Nature biotechnology 29, 24-26.

237. Rodaway, A., and Patient, R. (2001). Mesendoderm. an ancient germ layer? Cell 105, 169-172.

238. Rodaway, A., Takeda, H., Koshida, S., Broadbent, J., Price, B., Smith, J.C., Patient, R., and Holder, N. (1999). Induction of the mesendoderm in the zebrafish germ ring by yolk cell-derived TGF-beta family signals and discrimination of mesoderm and endoderm by FGF. Development 126, 3067-3078.

239. Ruscetti, F., Varesio, L., Ochoa, A., and Ortaldo, J. (1993). Pleiotropic effects of transforming growth factor-beta on cells of the immune system. Annals of the New York Academy of Sciences 685, 488-500.

240. Russ, A.P., Wattler, S., Colledge, W.H., Aparicio, S.A., Carlton, M.B., Pearce, J.J., Barton, S.C., Surani, M.A., Ryan, K., Nehls, M.C., et al. (2000). Eomesodermin is required for mouse trophoblast development and mesoderm formation. Nature 404, 95-99.

241. Ruthenburg, A.J., Li, H., Patel, D.J., and Allis, C.D. (2007). Multivalent engagement of chromatin modifications by linked binding modules. Nat Rev Mol Cell Biol 8, 983-994.

242. Ryan, A.K., Blumberg, B., Rodriguez-Esteban, C., Yonei-Tamura, S., Tamura, K., Tsukui, T., de la Pena, J., Sabbagh, W., Greenwald, J., Choe, S., et al. (1998). Pitx2 determines left-right asymmetry of internal organs in vertebrates. Nature 394, 545-551.

243. Ryan, K., Garrett, N., Bourillot, P., Stennard, F., and Gurdon, J.B. (2000). The Xenopus eomesodermin promoter and its concentration-dependent response to activin. Mech Dev 94, 133-146.

244. Ryan, K., Garrett, N., Mitchell, A., and Gurdon, J.B. (1996). Eomesodermin, a key early gene in Xenopus mesoderm differentiation. Cell 87, 989-1000.

245. Sabapathy, K., Klemm, M., Jaenisch, R., and Wagner, E.F. (1997). Regulation of ES cell differentiation by functional and conformational modulation of p53. The EMBO journal 16, 6217-6229.

246. Saijoh, Y., Adachi, H., Sakuma, R., Yeo, C.Y., Yashiro, K., Watanabe, M., Hashiguchi, H., Mochida, K., Ohishi, S., Kawabata, M., et al. (2000). Left-right asymmetric expression of lefty2 and nodal is induced by a signaling pathway that includes the transcription factor FAST2. Molecular cell 5, 35-47.

247. Santos-Rosa, H., Schneider, R., Bannister, A.J., Sherriff, J., Bernstein, B.E., Emre, N.C., Schreiber, S.L., Mellor, J., and Kouzarides, T. (2002). Active genes are tri-methylated at K4 of histone H3. Nature 419, 407-411.

248.   Sasai, N., Yakura, R., Kamiya, D., Nakazawa, Y., and Sasai, Y. (2008). Ectodermal factor restricts mesoderm differentiation by inhibiting p53. Cell 133, 878-890.

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

250.   Scandura, J.M., Boccuni, P., Massague, J., and Nimer, S.D. (2004). Transforming growth factor beta-induced cell cycle arrest of human hematopoietic cells requires p57KIP2 up-regulation. Proc Natl Acad Sci U S A 101, 15231-15236.

251.   Schafer, M.K., and Altevogt, P. (2010). L1CAM malfunction in the nervous system and human carcinomas. Cellular and molecular life sciences : CMLS 67, 2425-2437.

252.   Schmid, P., Lorenz, A., Hameister, H., and Montenarh, M. (1991). Expression of p53 during mouse embryogenesis. Development 113, 857-865.

253.   Schrode, N., Saiz, N., Di Talia, S., and Hadjantonakis, A.K. (2014). GATA6 levels modulate primitive endoderm cell fate choice and timing in the mouse blastocyst. Developmental cell 29, 454-467.

254.   Schroder, C., Schumacher, U., Fogel, M., Feuerhake, F., Muller, V., Wirtz, R.M., Altevogt, P., Krenkel, S., Janicke, F., and Milde-Langosch, K. (2009). Expression and prognostic value of L1-CAM in breast cancer. Oncology reports 22, 1109-1117.

255.   Shi, Y., Wang, Y.F., Jayaraman, L., Yang, H., Massague, J., and Pavletich, N.P. (1998). Crystal structure of a Smad MH1 domain bound to DNA: insights on DNA binding in TGF-beta signaling. Cell 94, 585-594.

256.   Shigeta, M., Ohtsuka, S., Nishikawa-Torikai, S., Yamane, M., Fujii, S., Murakami, K., and Niwa, H. (2013). Maintenance of pluripotency in mouse ES cells without Trp53. Scientific reports 3, 2944.

257.   Shimizu, K., and Gurdon, J.B. (1999). A quantitative analysis of signal transduction from activin receptor to nucleus and its relevance to morphogen gradient interpretation. Proc Natl Acad Sci U S A 96, 6791-6796.

258.   Shiratori, H., Sakuma, R., Watanabe, M., Hashiguchi, H., Mochida, K., Sakai, Y., Nishino, J., Saijoh, Y., Whitman, M., and Hamada, H. (2001). Two-step regulation of left-right asymmetric expression of Pitx2: initiation by nodal signaling and maintenance by Nkx2. Molecular cell 7, 137-149.

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

260.   Sirard, C., de la Pompa, J.L., Elia, A., Itie, A., Mirtsos, C., Cheung, A., Hahn, S., Wakeham, A., Schwartz, L., Kern, S.E., et al. (1998). The tumor suppressor gene Smad4/Dpc4 is required for gastrulation and later for anterior development of the mouse embryo. Genes & development 12, 107-119.

261.   Slagle, C.E., Aoki, T., and Burdine, R.D. (2011). Nodal-dependent mesendoderm specification requires the combinatorial activities of FoxH1 and Eomesodermin. PLoS genetics 7, e1002072.

262.   Smith, J.R., Vallier, L., Lupo, G., Alexander, M., Harris, W.A., and Pedersen, R.A. (2008). Inhibition of Activin/Nodal signaling promotes specification of human embryonic stem cells into neuroectoderm. Dev Biol 313, 107-117.

263.   Solnica-Krezel, L., and Sepich, D.S. (2012). Gastrulation: making and shaping germ layers. Annual review of cell and developmental biology 28, 687-717.

264.   Song, J., Oh, S.P., Schrewe, H., Nomura, M., Lei, H., Okano, M., Gridley, T., and Li, E. (1999). The type II activin receptors are essential for egg cylinder growth, gastrulation, and rostral head development in mice. Dev Biol 213, 157-169.

265.   Stanley, S., Domingos, A.I., Kelly, L., Garfield, A., Damanpour, S., Heisler, L., and Friedman, J. (2013). Profiling of Glucose-Sensing Neurons Reveals that GHRH Neurons Are Activated by Hypoglycemia. Cell metabolism 18, 596-607.

266.   Stiewe, T. (2007). The p53 family in differentiation and tumorigenesis. Nat Rev Cancer 7, 165-168.

267. Stroschein, S.L., Wang, W., Zhou, S., Zhou, Q., and Luo, K. (1999). Negative feedback regulation of TGF-beta signaling by the SnoN oncoprotein. Science 286, 771-774.

268. Sturzbecher, H.W., Brain, R., Addison, C., Rudge, K., Remm, M., Grimaldi, M., Keenan, E., and Jenkins, J.R. (1992). A C-terminal alpha-helix plus basic region motif is the major structural determinant of p53 tetramerization. Oncogene 7, 1513-1523.

269. Sun, X., Meyers, E.N., Lewandoski, M., and Martin, G.R. (1999). Targeted disruption of Fgf8 causes failure of cell migration in the gastrulating mouse embryo. Genes & development 13, 1834-1846.

270. Takahashi, K., Tanabe, K., Ohnuki, M., Narita, M., Ichisaka, T., Tomoda, K., and Yamanaka, S. (2007). Induction of pluripotent stem cells from adult human fibroblasts by defined factors. Cell 131, 861-872.

271. Takahashi, K., Tanabe, K., Ohnuki, M., Narita, M., Sasaki, A., Yamamoto, M., Nakamura, M., Sutou, K., Osafune, K., and Yamanaka, S. (2014). Induction of pluripotency in human somatic cells via a transient state resembling primitive streak-like mesendoderm. Nature communications 5, 3678.

272. Takahashi, K., and Yamanaka, S. (2006). Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. Cell 126, 663-676.

273. Takebayashi-Suzuki, K., Funami, J., Tokumori, D., Saito, A., Watabe, T., Miyazono, K., Kanda, A., and Suzuki, A. (2003). Interplay between the tumor suppressor p53 and TGF beta signaling shapes embryonic body axes in Xenopus. Development 130, 3929-3939.

274. Takenoshita, S., Mogi, A., Nagashima, M., Yang, K., Yagi, K., Hanyu, A., Nagamachi, Y., Miyazono, K., and Hagiwara, K. (1998). Characterization of the MADH2/Smad2 gene, a human Mad homolog responsible for the transforming growth factor-beta and activin signal transduction pathway. Genomics 48, 1-11.

275. Talos, F., Nemajerova, A., Flores, E.R., Petrenko, O., and Moll, U.M. (2007). p73 suppresses polyploidy and aneuploidy in the absence of functional p53. Molecular cell 27, 647-659.

276. Tam, P.P., and Behringer, R.R. (1997). Mouse gastrulation: the formation of a mammalian body plan. Mech Dev 68, 3-25.

277. Tao, Q., Yokota, C., Puck, H., Kofron, M., Birsoy, B., Yan, D., Asashima, M., Wylie, C.C., Lin, X., and Heasman, J. (2005). Maternal wnt11 activates the canonical wnt signaling pathway required for axis formation in Xenopus embryos. Cell 120, 857-871.

278. Teo, A.K., Arnold, S.J., Trotter, M.W., Brown, S., Ang, L.T., Chng, Z., Robertson, E.J., Dunn, N.R., and Vallier, L. (2011). Pluripotency factors regulate definitive endoderm specification through eomesodermin. Genes & development 25, 238-250.

279. The Cancer Genome Atlast. Comprehensive molecular characterization of human colon and rectal cancer. Nature 487, 330-337.

280. Thomas, K.R., and Capecchi, M.R. (1990). Targeted disruption of the murine int-1 proto-oncogene resulting in severe abnormalities in midbrain and cerebellar development. Nature 346, 847-850.

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

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

283. Tremblay, K.D., Hoodless, P.A., Bikoff, E.K., and Robertson, E.J. (2000). Formation of the definitive endoderm in mouse is a Smad2-dependent process. Development 127, 3079-3090.

284. Tsukada, T., Tomooka, Y., Takai, S., Ueda, Y., Nishikawa, S., Yagi, T., Tokunaga, T., Takeda, N., Suda, Y., Abe, S., et al. (1993). Enhanced proliferative potential in culture of cells from p53-deficient mice. Oncogene 8, 3313-3322.

285. Tufegdzic Vidakovic, A., Rueda, O.M., Vervoort, S.J., Sati Batra, A., Goldgraben, M.A., Uribe-Lewis, S., Greenwood, W., Coffer, P.J., Bruna, A., and Caldas, C. (2015). Context-Specific Effects of TGF-beta/SMAD3 in Cancer Are Modulated by the Epigenome. Cell reports 13, 2480-2490.

286. Utikal, J., Polo, J.M., Stadtfeld, M., Maherali, N., Kulalert, W., Walsh, R.M., Khalil, A., Rheinwald, J.G., and Hochedlinger, K. (2009). Immortalization eliminates a roadblock during cellular reprogramming into iPS cells. Nature 460, 1145-1148.

287.   Vale, W., Rivier, J., Vaughan, J., McClintock, R., Corrigan, A., Woo, W., Karr, D., and Spiess, J. (1986). Purification and characterization of an FSH releasing protein from porcine ovarian follicular fluid. Nature 321, 776-779.

288.   Valiente, M., Obenauf, A.C., Jin, X., Chen, Q., Zhang, X.H., Lee, D.J., Chaft, J.E., Kris, M.G., Huse, J.T., Brogi, E., et al. (2014). Serpins promote cancer cell survival and vascular co-option in brain metastasis. Cell 156, 1002-1016.

289.   Vallier, L., Mendjan, S., Brown, S., Chng, Z., Teo, A., Smithers, L.E., Trotter, M.W., Cho, C.H., Martinez, A., Rugg-Gunn, P., et al. (2009). Activin/Nodal signalling maintains pluripotency by controlling Nanog expression. Development 136, 1339-1349.

290.   Vallier, L., Reynolds, D., and Pedersen, R.A. (2004). Nodal inhibits differentiation of human embryonic stem cells along the neuroectodermal default pathway. Dev Biol 275, 403-421.

291.   Varlet, I., Collignon, J., and Robertson, E.J. (1997). nodal expression in the primitive endoderm is required for specification of the anterior axis during mouse gastrulation. Development 124, 1033-1044.

292.   Vincent, S.D., Dunn, N.R., Hayashi, S., Norris, D.P., and Robertson, E.J. (2003). Cell fate decisions within the mouse organizer are governed by graded Nodal signals. Genes & development 17, 1646-1662.

293.   Viotti, M., Nowotschin, S., and Hadjantonakis, A.K. (2014). SOX17 links gut endoderm morphogenesis and germ layer segregation. Nat Cell Biol 16, 1146-1156.

294.   von Both, I., Silvestri, C., Erdemir, T., Lickert, H., Walls, J.R., Henkelman, R.M., Rossant, J., Harvey, R.P., Attisano, L., and Wrana, J.L. (2004). Foxh1 is essential for development of the anterior heart field. Developmental cell 7, 331-345.

295.   Vousden, K.H., and Prives, C. (2009). Blinded by the Light: The Growing Complexity of p53. Cell 137, 413-431.

296.   Waldrip, W.R., Bikoff, E.K., Hoodless, P.A., Wrana, J.L., and Robertson, E.J. (1998). Smad2 signaling in extraembryonic tissues determines anterior-posterior polarity of the early mouse embryo. Cell 92, 797-808.

297.	Wallingford, J.B., Seufert, D.W., Virta, V.C., and Vize, P.D. (1997). p53 activity is essential for normal development in Xenopus. Curr Biol 7, 747-757.

298.	Wang, L., and Chen, Y.G. (2015). Signaling Control of Differentiation of Embryonic Stem Cells toward Mesendoderm. Journal of molecular biology.

299.	Wardle, F.C., and Smith, J.C. (2006). Transcriptional regulation of mesendoderm formation in Xenopus. Seminars in cell & developmental biology 17, 99-109.

300.	Watanabe, M., Masuyama, N., Fukuda, M., and Nishida, E. (2000). Regulation of intracellular dynamics of Smad4 by its leucine-rich nuclear export signal. EMBO reports 1, 176-182.

301.	Watanabe, S., Umehara, H., Murayama, K., Okabe, M., Kimura, T., and Nakano, T. (2006). Activation of Akt signaling is sufficient to maintain pluripotency in mouse and primate embryonic stem cells. Oncogene 25, 2697-2707.

302.	Weinberger, L., Ayyash, M., Novershtern, N., and Hanna, J.H. (2016). Dynamic stem cell states: naive to primed pluripotency in rodents and humans. Nat Rev Mol Cell Biol 17, 155-169.

303.	Weinstein, D.C., Ruiz i Altaba, A., Chen, W.S., Hoodless, P., Prezioso, V.R., Jessell, T.M., and Darnell, J.E., Jr. (1994). The winged-helix transcription factor HNF-3 beta is required for notochord development in the mouse embryo. Cell 78, 575-588.

304.	Weinstein, M., Yang, X., Li, C., Xu, X., Gotay, J., and Deng, C.X. (1998). Failure of egg cylinder elongation and mesoderm induction in mouse embryos lacking the tumor suppressor smad2. Proc Natl Acad Sci U S A 95, 9378-9383.

305.	Weisberg, E., Winnier, G.E., Chen, X., Farnsworth, C.L., Hogan, B.L., and Whitman, M. (1998). A mouse homologue of FAST-1 transduces TGF beta superfamily signals and is expressed during early embryogenesis. Mech Dev 79, 17-27.

306.	Whitman, M. (1998). Smads and early developmental signaling by the TGFbeta superfamily. Genes & development 12, 2445-2462.

307. Whyte, W.A., Bilodeau, S., Orlando, D.A., Hoke, H.A., Frampton, G.M., Foster, C.T., Cowley, S.M., and Young, R.A. (2012). Enhancer decommissioning by LSD1 during embryonic stem cell differentiation. Nature 482, 221-225.

308. Wikner, N.E., Persichitte, K.A., Baskin, J.B., Nielsen, L.D., and Clark, R.A. (1988). Transforming growth factor-beta stimulates the expression of fibronectin by human keratinocytes. The Journal of investigative dermatology 91, 207-212.

309. Wilkinson, D.G., Bhatt, S., and Herrmann, B.G. (1990). Expression pattern of the mouse T gene and its role in mesoderm formation. Nature 343, 657-659.

310. Williams, R.L., Hilton, D.J., Pease, S., Willson, T.A., Stewart, C.L., Gearing, D.P., Wagner, E.F., Metcalf, D., Nicola, N.A., and Gough, N.M. (1988). Myeloid leukaemia inhibitory factor maintains the developmental potential of embryonic stem cells. Nature 336, 684-687.

311. Wray, J., Kalkan, T., Gomez-Lopez, S., Eckardt, D., Cook, A., Kemler, R., and Smith, A. (2011). Inhibition of glycogen synthase kinase-3 alleviates Tcf3 repression of the pluripotency network and increases embryonic stem cell resistance to differentiation. Nat Cell Biol 13, 838-845.

312. Xi, Q., He, W., Zhang, X.H., Le, H.V., and Massague, J. (2008). Genome-wide impact of the BRG1 SWI/SNF chromatin remodeler on the transforming growth factor beta transcriptional program. The Journal of biological chemistry 283, 1146-1155.

313. Xi, Q., Wang, Z., Zaromytidou, A.I., Zhang, X.H., Chow-Tsang, L.F., Liu, J.X., Kim, H., Barlas, A., Manova-Todorova, K., Kaartinen, V., et al. (2011). A poised chromatin platform for TGF-beta access to master regulators. Cell 147, 1511-1524.

314. Xu, J., Lamouille, S., and Derynck, R. (2009). TGF-beta-induced epithelial to mesenchymal transition. Cell research 19, 156-172.

315. Xu, L., Chen, Y.G., and Massague, J. (2000). The nuclear import function of Smad2 is masked by SARA and unmasked by TGFbeta-dependent phosphorylation. Nat Cell Biol 2, 559-562.

316. Xu, L., Kang, Y., Col, S., and Massague, J. (2002). Smad2 nucleocytoplasmic shuttling by nucleoporins CAN/Nup214 and Nup153 feeds TGFbeta signaling complexes in the cytoplasm and nucleus. Molecular cell 10, 271-282.

317. Yagi, K., Goto, D., Hamamoto, T., Takenoshita, S., Kato, M., and Miyazono, K. (1999). Alternatively spliced variant of Smad2 lacking exon 3. Comparison with wild-type Smad2 and Smad3. The Journal of biological chemistry 274, 703-709.

318. Yamamoto, M., Meno, C., Sakai, Y., Shiratori, H., Mochida, K., Ikawa, Y., Saijoh, Y., and Hamada, H. (2001). The transcription factor FoxH1 (FAST) mediates Nodal signaling during anterior-posterior patterning and node formation in the mouse. Genes & development 15, 1242-1256.

319. Yamamoto, M., Mine, N., Mochida, K., Sakai, Y., Saijoh, Y., Meno, C., and Hamada, H. (2003). Nodal signaling induces the midline barrier by activating Nodal expression in the lateral plate. Development 130, 1795-1804.

320. Yamamoto, M., Saijoh, Y., Perea-Gomez, A., Shawlot, W., Behringer, R.R., Ang, S.L., Hamada, H., and Meno, C. (2004). Nodal antagonists regulate formation of the anteroposterior axis of the mouse embryo. Nature 428, 387-392.

321. Yan, D., Wiesmann, M., Rohan, M., Chan, V., Jefferson, A.B., Guo, L., Sakamoto, D., Caothien, R.H., Fuller, J.H., Reinhard, C., et al. (2001). Elevated expression of axin2 and hnkd mRNA provides evidence that Wnt/beta -catenin signaling is activated in human colon tumors. Proc Natl Acad Sci U S A 98, 14973-14978.

322. Yanagisawa, K.O., Fujimoto, H., and Urushihara, H. (1981). Effects of the brachyury (T) mutation on morphogenetic movement in the mouse embryo. Dev Biol 87, 242-248.

323. Yang, A., Schweitzer, R., Sun, D., Kaghad, M., Walker, N., Bronson, R.T., Tabin, C., Sharpe, A., Caput, D., Crum, C., et al. (1999). p63 is essential for regenerative proliferation in limb, craniofacial and epithelial development. Nature 398, 714-718.

324. Yang, A., Walker, N., Bronson, R., Kaghad, M., Oosterwegel, M., Bonnin, J., Vagner, C., Bonnet, H., Dikkes, P., Sharpe, A., et al. (2000). p73-deficient mice have neurological, pheromonal and inflammatory defects but lack spontaneous tumours. Nature 404, 99-103.

325. Yang, J., and Weinberg, R.A. (2008). Epithelial-mesenchymal transition: at the crossroads of development and tumor metastasis. Developmental cell 14, 818-829.

326. Yang, X., Letterio, J.J., Lechleider, R.J., Chen, L., Hayman, R., Gu, H., Roberts, A.B., and Deng, C. (1999). Targeted disruption of SMAD3 results in impaired mucosal immunity and diminished T cell responsiveness to TGF-beta. The EMBO journal 18, 1280-1291.

327. Yang, X., Li, C., Xu, X., and Deng, C. (1998). The tumor suppressor SMAD4/DPC4 is essential for epiblast proliferation and mesoderm induction in mice. Proc Natl Acad Sci U S A 95, 3667-3672.

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

329. Ying, Q.L., Nichols, J., Chambers, I., and Smith, A. (2003). BMP induction of Id proteins suppresses differentiation and sustains embryonic stem cell self-renewal in collaboration with STAT3. Cell 115, 281-292.

330. Ying, Q.L., Wray, J., Nichols, J., Batlle-Morera, L., Doble, B., Woodgett, J., Cohen, P., and Smith, A. (2008). The ground state of embryonic stem cell self-renewal. Nature 453, 519-523.

331. Yoon, J.H., Sudo, K., Kuroda, M., Kato, M., Lee, I.K., Han, J.S., Nakae, S., Imamura, T., Kim, J., Ju, J.H., et al. (2015). Phosphorylation status determines the opposing functions of Smad2/Smad3 as STAT3 cofactors in TH17 differentiation. Nature communications 6, 7600.

332. Yoon, S.J., Wills, A.E., Chuong, E., Gupta, R., and Baker, J.C. (2011). HEB and E2A function as SMAD/FOXH1 cofactors. Genes & development 25, 1654-1661.

333. Yoon, Y., Huang, T., Tortelote, G.G., Wakamiya, M., Hadjantonakis, A.K., Behringer, R.R., and Rivera-Perez, J.A. (2015). Extra-embryonic Wnt3 regulates the establishment of the primitive streak in mice. Dev Biol 403, 80-88.

334. Yoshioka, H., Meno, C., Koshiba, K., Sugihara, M., Itoh, H., Ishimaru, Y., Inoue, T., Ohuchi, H., Semina, E.V., Murray, J.C., et al. (1998). Pitx2, a bicoid-type homeobox gene, is involved in a lefty-signaling pathway in determination of left-right asymmetry. Cell 94, 299-305.

335. Young, R.A. (2011). Control of the embryonic stem cell state. Cell 144, 940-954.

336. Zawel, L., Dai, J.L., Buckhaults, P., Zhou, S., Kinzler, K.W., Vogelstein, B., and Kern, S.E. (1998). Human Smad3 and Smad4 are sequence-specific transcription activators. Molecular cell 1, 611-617.

337. Zhang, X.H., Jin, X., Malladi, S., Zou, Y., Wen, Y.H., Brogi, E., Smid, M., Foekens, J.A., and Massague, J. (2013). Selection of bone metastasis seeds by mesenchymal signals in the primary tumor stroma. Cell 154, 1060-1073.

338. Zhang, X.H., Wang, Q., Gerald, W., Hudis, C.A., Norton, L., Smid, M., Foekens, J.A., and Massague, J. (2009). Latent bone metastasis in breast cancer tied to Src-dependent survival signals. Cancer cell 16, 67-78.

339. Zhao, Y., Yin, X., Qin, H., Zhu, F., Liu, H., Yang, W., Zhang, Q., Xiang, C., Hou, P., Song, Z., et al. (2008). Two supporting factors greatly improve the efficiency of human iPSC generation. Cell Stem Cell 3, 475-479.

340. Zhong, C., Yin, Q., Xie, Z., Bai, M., Dong, R., Tang, W., Xing, Y.H., Zhang, H., Yang, S., Chen, L.L., et al. (2015). CRISPR-Cas9-Mediated Genetic Screening in Mice with Haploid Embryonic Stem Cells Carrying a Guide RNA Library. Cell Stem Cell 17, 221-232.

341. Zhou, C., Pourmal, S., and Pavletich, N.P. (2015). Dna2 nuclease-helicase structure, mechanism and regulation by Rpa. eLife 4.

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

343. Zhu, C.C., Yamada, G., Nakamura, S., Terashi, T., Schweickert, A., and Blum, M. (1998). Malformation of trachea and pelvic region in goosecoid mutant mice. Developmental dynamics : an official publication of the American Association of Anatomists 211, 374-381.

344. Zilfou, J.T., and Lowe, S.W. (2009). Tumor suppressive functions of p53. Cold Spring Harb Perspect Biol 1, a001883.